

PARTHENOS

Pooling Activities, Resources and Tools
for Heritage E-research Networking,
Optimization and Synergies

REPORT ON USER REQUIREMENTS

CLARIN (overall coordination)

MIBACT-ICCU

PIN

KNAW

CNR-OVI (earlier SISMEL)

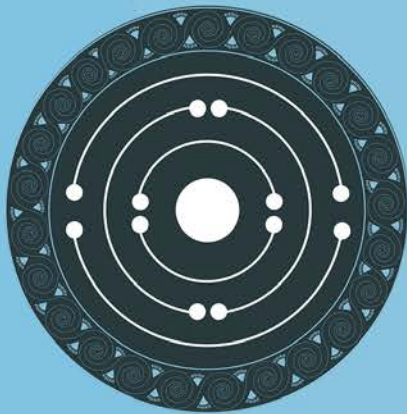
FHP

With contributions from all PARTHENOS partners

20 October 2016



PARTHENOS is a Horizon 2020 project funded by the European Commission. The views and opinions expressed in this publication are the sole responsibility of the author and do not necessarily reflect the views of the European Commission.



PARTHENOS

Pooling Activities, Resources and Tools
for Heritage E-research Networking,
Optimization and Synergies

REPORT ON USER REQUIREMENTS

CLARIN (overall coordination)

MIBACT-ICCU

PIN

KNAW

CNR-OVI (earlier SISMELE)

FHP

With contributions from all PARTHENOS partners

20 October 2016



PARTHENOS is a Horizon 2020 project funded by the European Commission. The views and opinions expressed in this publication are the sole responsibility of the author and do not necessarily reflect the views of the European Commission.

HORIZON 2020 - INFRADEV-4-2014/2015:

Grant Agreement No. 654119

PARTHENOS

Pooling Activities, Resources and Tools for Heritage E-research Networking,
Optimization and Synergies

REPORT ON USER REQUIREMENTS

Deliverable Number D2.1

Dissemination Level Public

Delivery date 31 January 2016 (first version)

20 October 2016 (final version)

Status *Final version*

Main Author(s)

Sebastian Drude (CLARIN)

Sara di Giorgio (MIBACT-ICCU)

Paola Ronzino (PIN)

Petra Links, Annelies van Nispen, Karolien Verbrugge (KNAW / NIOD)

Emiliano Degl'Innocenti (CNR-OVI, formerly SISMEL)

Juliane Stiller, Jenny Oltersdorf, Claus Spiecker (FHP)

With contributions from all PARTHENOS partners

| | |
|---------------------|--|
| Project Acronym | PARTHENOS |
| Project Full title | Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies |
| Grant Agreement nr. | 654119 |

Deliverable/Document Information

| | |
|--------------------------------------|--|
| Deliverable nr./title | D2.1, Report on User Requirements |
| Document file name | PARTHENOS-D21-user-requirements-report |
| Author(s) | <p>Sebastian Drude (CLARIN, overall coordination)</p> <p>Sara di Giorgio (MIBACT-ICCU)</p> <p>Paola Ronzino (PIN)</p> <p>Petra Links, Annelies van Nispen (KNAW / NIOD), Karolien Vergrugge (KNAW / NIOD, responsible for the final edition of chapter 2)</p> <p>Emiliano Degl’Innocenti (CNR-OVI, formerly SISMEL)</p> <p>Jenny Oltersdorf (FHP)</p> <p>Juliane Stiller (FHP), replaced by Jenny Oltersdorf and Claus Spiecker (responsible for the final edition of chapters 4 and 5 and sections in chapter 1) and Paola Ronzino)</p> <p><i>With contributions from all PARTHENOS partners</i></p> |
| Dissemination level/ distribution | Public |

Document History

| Version/date | Changes/approval | Author/Approved by |
|-------------------|--|--|
| V0.1 20 Nov 2015 | First complete drafts of all chapters except Chapter 0 | Sara di Giorgio, Petra Links, Annelies van Nispen, Emiliano Degl'Innocenti, Jenny Oltersdorf, Juliane Stiller, and collaborators |
| V0.5 11 Jan 2016 | Revised complete drafts of all chapters including Chapter 0 | Sebastian Drude, Sara di Giorgio, Paola Ronzino, Petra Links, Annelies van Nispen, Emiliano Degl'Innocenti, Jenny Oltersdorf, and collaborators |
| V0.6 18 Jan 2016 | Second revised complete drafts of all chapters after internal peer-review | Sebastian Drude, Sara di Giorgio, Paola Ronzino, Petra Links, Annelies van Nispen, Emiliano Degl'Innocenti, Jenny Oltersdorf, Juliane Stiller, and collaborators |
| V0.8 25 Jan 2016 | Pre-final complete revised drafts with English corrections by Mark Hedges, Sheena Bassett, and Vicky Garnett | Sebastian Drude, Sara di Giorgio, Paola Ronzino, Petra Links, Annelies van Nispen, Emiliano Degl'Innocenti, Jenny Oltersdorf, and collaborators |
| V1.0 31 Jan 2016 | First complete revised version with coherent formatting and bibliographic references added | As for V0.8, above, and also Sebastian Drude and Sheena Bassett (editing) Vanessa Hanneschläger (references) |
| V1.5 31 July 2016 | First versions of improved individual chapters | As above for V0.8, now with Claus Spiecker (responsible |

| | | |
|-------------------|--|---|
| | | for chapter 5), and with Karolien Verbrugge replacing Petra Links and Annelies van Nispen (chapter 2) |
| V1.7 31 Aug 2016 | Improved individual chapters after internal peer-review | As for V1.5 above |
| V1.9 25 Sept 2016 | Final versions of individual chapters after revision by native English speakers Mark Hedges, Sheena Bassett, and Vicky Garnett | As for V1.5 above |
| V2.0 20 Oct 2016 | Final revised version, combined, formatted and final edition by Sebastian Drude (CLARIN) | As for V1.5 above |

Contents

| | |
|--|-----------|
| Executive Summary | 1 |
| Abbreviations | 5 |
| 0. Overview and Methodology | 9 |
| 0.0. Introduction; embedding | 9 |
| 0.1. Structure of this report | 10 |
| 0.2. User communities in PARTHENOS | 11 |
| 0.2.1. Overview..... | 11 |
| 0.2.2. History in a broad sense..... | 12 |
| 0.2.3. Language Related Studies..... | 13 |
| 0.2.4. Cultural Heritage, applied disciplines, and Archaeology..... | 15 |
| 0.2.5. Social Sciences in a broader sense..... | 16 |
| 0.3. Methods of identifying user requirements | 17 |
| 0.4. Methods of presenting user requirements | 19 |
| 1. Requirements concerning data policies | 22 |
| 1.0. Introduction | 22 |
| 1.0.0. Overview..... | 22 |
| 1.0.1. Gathering the requirements..... | 23 |
| 1.1. Definition of Policy Requirements Concerning the Research Data Lifecycle | 23 |
| 1.1.1. Introduction: Models of the Research Data Lifecycle..... | 23 |
| 1.1.2. Overview of the Analysis of the Research Data Lifecycle..... | 27 |
| 1.1.3. Results – Definition of Policy Requirements Concerning the Data Lifecycle..... | 32 |
| 1.2. Definition of Policy Requirements on Quality Assessment of Digital Repositories and Quality Assurance of Data and Metadata Items | 57 |
| 1.2.1. Introduction..... | 57 |
| 1.2.2. Overview of the Analysis of Quality Assessment of Digital Repositories..... | 58 |
| 1.2.3. Results of the quality assessment..... | 59 |
| 1.2.4. Overview of the Analysis of Data and Metadata Assessment..... | 63 |
| 1.2.5. Results – Metadata and Data Quality Assessment..... | 66 |
| 1.3. Definition of Policy Requirements Concerning IPR, Open Data and Open Access | 73 |
| 1.3.1. Introduction – Overview of IPR..... | 73 |
| 1.3.2. Overview of the analysis of IPR, Open Data and Open Access requirements..... | 74 |
| 1.3.3. Results – the IPR requirements..... | 76 |
| 1.3.4. Overview of Open Data..... | 83 |
| 1.3.5. Results – the Open Data requirements..... | 84 |

| | | |
|-----------|---|------------|
| 1.3.6. | Overview of the analysis of Open Access requirements..... | 86 |
| 1.3.7. | Results – Open Access..... | 88 |
| 1.3.8. | Narrative Use Case..... | 96 |
| 2. | Use Cases and Requirements on Standardization | 100 |
| 2.0. | Introduction | 100 |
| 2.1. | Use cases..... | 100 |
| 2.1.1. | History | 101 |
| 2.1.2. | Language Related Studies..... | 113 |
| 2.1.3. | Archaeology, Heritage and Applied Disciplines | 125 |
| 2.1.4. | Social Sciences | 148 |
| 2.2. | Requirements | 153 |
| 3. | Interoperability, services and tools requirements | 159 |
| 3.0. | Objectives | 159 |
| 3.1. | Method | 159 |
| 3.2. | A working definition of interoperability | 161 |
| 3.3. | PARTHENOS reference model..... | 161 |
| 3.4. | Use cases modelling and requirements extraction..... | 162 |
| 3.5. | From the requirements to the architectural design..... | 165 |
| 3.6. | Requirements for interoperability: Use Cases | 169 |
| 3.6.1. | Use cases from Archaeology, Heritage and applied disciplines | 169 |
| 3.6.2. | Use cases from Language-related studies | 178 |
| 3.6.3. | Use cases from Studies of the Past | 183 |
| 3.7. | Requirements for interoperability: Mapped requirements..... | 194 |
| 3.7.1. | Mapped requirements from Studies of the Past | 194 |
| 3.7.2. | Mapped requirements from Social Sciences | 200 |
| 3.8. | Requirements for interoperability: General requirements..... | 218 |
| 3.8.1. | Requirements from Archaeology, Heritage and applied disciplines | 218 |
| 3.8.2. | Requirements from Language-related Studies..... | 225 |
| 3.8.3. | Requirements from Studies of the Past..... | 236 |
| 3.9. | Conclusions..... | 242 |
| 4. | Definition of education & training requirements..... | 243 |
| 4.0. | Introduction | 243 |
| 4.1. | Method | 243 |
| 4.1.1. | Text analysis..... | 244 |
| 4.1.2. | Survey | 245 |
| 4.1.3. | Desktop Research | 246 |

| | | |
|-------------------|--|------------|
| 4.1.4. | User study | 246 |
| 4.1.5. | European Summer University in Digital Humanities, Leipzig | 247 |
| 4.2. | Document analysis..... | 247 |
| 4.3. | Experiences of the PARTHENOS community..... | 262 |
| 4.4. | User study about digital approaches at University of Copenhagen | 266 |
| 4.5. | Inventory of existing platforms for training and education | 267 |
| 4.6. | ESU - European Summer University in Digital Humanities (ESU)..... | 270 |
| 4.7. | Conclusions..... | 275 |
| 5. | Communication requirements..... | 277 |
| 5.0. | Introduction | 277 |
| 5.1. | Method | 277 |
| 5.2. | Scholarly communication | 279 |
| 5.3. | Requirements for dissemination | 282 |
| 5.3.1. | Target groups | 282 |
| 5.3.2. | Dissemination activities..... | 283 |
| 5.3.3. | Success criteria | 283 |
| 5.4. | Next steps | 284 |
| References | | 285 |

Executive Summary

This document is the final, updated, version of deliverable D2.1 of the PARTHENOS project, which addresses user requirements and needs. It contains a comprehensive report based on a review of literature produced by previous relevant projects, supplemented with additional direct input from PARTHENOS partners. The document is structured in chapters, as follows:

An introduction (Chapter 0) that characterizes the four main user communities on which PARTHENOS is focusing, and describes the methodology that has been followed in identifying requirements. The targeted user communities are: (i) History (in a broad sense); (ii) Language-related Studies; (iii) Archaeology, Heritage & Applied Disciplines¹; and, to a lesser degree, (iv) Social Sciences (in a broad sense). The methodology consisted of gathering and organizing relevant reports using Zotero and D4Science, extracting use cases and user requirements from them, and presenting these using in general the Simplified Language approach proposed by Cockburn (2000). The chapter also shows the direct relationship between the subsequent chapters and the other Work Packages in PARTHENOS.

Chapter 1: Requirements concerning data policies. Concerning data the research communities involved in this analysis has shown the need for better transparency of available data and for improvements to data accessibility. Data and metadata quality are also relevant concerns for researchers and for data managers. More than 40% of the requirements collected from the different communities constituting the PARTHENOS consortium report that one of the major concerns regards data preservation; this holds especially for the archaeological community.

The selection and promotion of high-quality deposit services is, instead, very important for researchers in the Social Sciences and Humanities, who also require the development of clear guidelines and procedures for management, archiving, and sharing of data.

For the Language related studies community, metadata harmonization is an important concern, intended as the challenge of verifying the structural and syntactic interoperability between the resources. Completeness reflecting the functional purpose of

¹ This term covers archivists, museum experts, preservation specialists, people working on digital curation and editions, and so forth.

metadata is required, including the resource type, its relation to the local collection and the metadata guidelines.

Regarding IPR, Open Data and Open Access research communities find desirable to have a framework of licences that standardizes and harmonizes rights for allowing data re-use. The provision of a Licensing Framework, within guideline for common policies implementation, would bring clarity to a complex area, and make transparent the relationship between end users and the institutions that provide data.

A need expressed by the research communities, in the IPR field, is the means to manage restricted access to protected resources by users. From this point of view, a better solution is represented by the AAI (Authentication and Authorization Infrastructure). Thanks to this system, for safeguarding privacy and data protection, it is possible to define different user levels and allow limited access to the resources that don't have a level of public dissemination.

Chapter 2: Standardisation requirements. This section deals with the requirements of standardisation expressed by the research communities involved in the project. Twenty use case are at the heart of this chapter. Each of them highlights a research community that doesn't use standards yet, or is in an early stage of doing so, or that has difficulties with implementing standards. Being developed by the research communities itself the use cases reflect common issues and shared needs in achieving a greater level of standardisation in order to provide access and to preserve data through time and space.

Chapter 3: Interoperability, services and tools requirements. The use cases in Chapter 3 are documenting requirements expressed by a vast number of disciplines in the PARTHENOS community, leveraging on the documentation made available by different partners and networks i.e.: ARIADNE (PIN, MIBACT-ICCU) for Archaeology, Heritage and Applied Disciplines, CENDARI (TCD, SISMELE) and EHRI (KNAW-DANS) for History, CLARIN for Language related studies, Huma-Num (CNRS) for Social Sciences, etc. Despite the different approach and methodological focus, we found a number of general-level requirements, shared across several use cases and disciplines, expressing the same needs e.g.: data quality, availability, accessibility and enrichment, as well as other specific needs (i.e.: visual media documents enrichment, integration of authority lists, gazetteers and reference tools and/or resources) driven by particular disciplinary concerns. Other requirements both from the backend (i.e.: like storage and preservation) and the frontend (i.e.: tools for collaborative work and data analysis) perspective

were gathered. The same holds for tools, where we found a similar situation with a shared set of priorities at the general level (i.e. search and information display tools) as well as some detailed, domain driven requirements (i.e. tools to prepare digital editions etc.).

Finally, a set of not (only) technical requirements, such as the sustainability of tools and datasets were expressed by the PARTHENOS community: we plan to consider them as action points for other WPs (namely WP3), and insert them in the agenda for the development of mid and long-term actions and policies.

Chapter 4: Education and training needs. This chapter concerns education and training, describing the current provision and the needs identified by the communities, and indicates priorities, common areas and emerging issues. It is based on the outcomes of Task 2.4. Main findings: The topics of already offered training courses mostly derive from surveys conducted within research projects. Thus training needs are mainly focused on concrete infrastructure or tools developed in the projects. A systematic improvement of training and education services on a more generalized, meta-level does not happen in the surveyed communities currently. In terms of implementation of training and education modules the feedback provided by the communities revealed a preference for face-to-face meetings. The combination of workshops, summer schools or Skype conferences with moderated distance-learning modules like webinars seem to be the most common and promising way of implementation. The experiences have shown that a human moderator / a contact person to ask questions to is one characteristic for a successful training module. Online tutorials or written documentations without a point of contact are classified as of minor effectiveness.

Chapter 5: Communication needs. This chapter concerns the communication needs identified by the various communities, and indicates priorities, common areas and emerging issues. It is based on the outcomes of Task 2.5.

Main findings: Evaluation criteria derived from the analyzed journals and repositories range from the domain and covered topics of the journals to regional and international coverage, languages, formats and outputs accepted to the ability to be quantitatively analyzed.

The dissemination reports revealed a group of five most evident activities. These are firstly, dissemination activities via project's website. Secondly, partners' institutional websites are used for the dissemination of information. Thirdly, newsletter and fourthly press releases are common means when it comes to dissemination strategies. Finally,

networking and consulting at conferences in various phases of the projects was also mentioned as one of the most important activities regarding the dissemination of project results.

Abbreviations

| | |
|------------|--|
| AA | Academy of Athens (Greece, PARTHENOS partner) |
| AAI | Authentication and Authorization Infrastructure |
| ADHO | The Alliance of Digital Humanities Organizations |
| AGORA | Scholarly Open Access Research in European Philosophy |
| APEX | Archives Portal Europe Foundation |
| ARIADNE | Advanced Research Infrastructure for Archaeological Dataset Networking in Europe |
| ATHENA | Access to Cultural Heritage Networks across Europe |
| AthenaPlus | Access to Cultural Heritage Networks for Europeana |
| BBAW | Berlin-Brandenburg Academy of Sciences and Humanities (Germany, see CLARIN) |
| CARMEN | The Worldwide Medieval Network |
| CENDARI | Collaborative European Digital Archival Research Infrastructure |
| CHARISMA | Cultural Heritage Advanced Research Infrastructures: Synergy for a Multidisciplinary Approach to Conservation/Restoration |
| CHI | Cultural Heritage Institution |
| CoHI | Content or Collection Holding Institution |
| CLARIN | Common Language and Technology Research Infrastructure (ERIC, Europe, PARTHENOS Partner, represents several institutions in PARTHENOS) |
| CMDI | Component MetaData Infrastructure |
| CNR | Consiglio Nazionale delle Ricerche (Italy, PARTHENOS partner; represents several institutions in PARTHENOS) |
| CNRS | Centre National de la Recherche Scientifique (France, PARTHENOS partner; represents several institutions in PARTHENOS) |
| COST | European Cooperation in Science and Technology |
| CSIC | Agencia Estatal Consejo Superior de Investigaciones Cientificas (Spain, PARTHENOS partner) |
| DANS | Data Archiving and Networked Services (see KNAW) |
| DARIAH-DE | Digital Research Infrastructure for the Arts and Humanities - Germany |

| | |
|-----------------|---|
| DARIAH EU | Digital Research Infrastructure for the Arts and Humanities (ERIC, Europe, PARTHENOS Partner, represents several institutions in PARTHENOS) |
| DARIAH-IT | Digital Research Infrastructure for the Arts and Humanities - Italy |
| DASISH | Digital Services Infrastructure for Social Sciences and Humanities |
| Data-PASS | Data Preservation Alliance for the Social Sciences |
| DC | Dublin Core |
| DCH-RP | Digital Cultural Heritage Roadmap for Preservation |
| DDI | Data Documentation Initiative |
| DH | Digital Humanities |
| DigCurV | Digital Curator Vocational Education Europe |
| DiRT | Digital Research Tool |
| DM | Digital Medievalist |
| DM2E | Digitised Manuscripts to Europeana |
| DSA | Data Seal of Approval |
| DYAS | Greek Research Infrastructure Network for the Humanities (=DARIAH GR) |
| E.C.C.O | European Confederation of Conservator-Restorers |
| ECLAP | European Collected Library of Artistic Performance |
| EHRI | European Holocaust Research Infrastructure |
| ERIC | European Research Infrastructure Consortium |
| ESFRI | European Strategy Forum on Research Infrastructures |
| EUDAT | Research Data Services, Expertise & Technology Solutions |
| Europeana Cloud | Unlocking Europe |
| ESU | European Summer University in Digital Humanities, Leipzig |
| FHP | Fachhochschule Potsdam (Germany, PARTHENOS partner, replaced UGOE) |
| FLaReNet | Fostering Language Resources Network |
| FORTH | Foundation for Research and Technology Hellas (Greece, PARTHENOS partner) |
| HHS | U.S. Dept. of Health and Human Services |
| HSS | Humanities and Social Sciences (synonymous with SSH) |
| Huma-Num | La TGIR Des Humanités Numériques (see CNRS) |

| | |
|-------------------------|---|
| ICCU | Istituto Centrale per il Catalogo Unico delle biblioteche italiane e per le informazioni bibliografiche (Italy, PARTHENOS partner) |
| INDIGO DataCloud | Towards a Sustainable European PaaS-Based Cloud Solution for E-Science |
| INRIA | Institut National De Recherche En Informatique Et En Automatique (PARTHENOS partner, France) |
| IPERION CH | Integrated Platform for the European Research Infrastructure ON Culture Heritage |
| IPR | Intellectual Property Rights |
| ISCH COST Action IS1005 | Medieval Europe - Medieval Cultures and Technological Resources (Medioevo Europeo) |
| ISIDORE | Portal for Digital Humanities by French National Research Center |
| ISO | International Standard Organization |
| ISTI | Istituto di Scienza e Tecnologie dell'Informazione (see CNR) |
| Jisc | Joint Information Systems Committee |
| KCL | King's College London (UK, PARTHENOS partner) |
| KNAW | Koninklijke Nederlandse Akademie van Wetenschappen (the Netherlands, PARTHENOS partner, represents two institutions: DANS and NIOD) |
| LR | Language Resource(s) |
| LREC | Conference on Language Resources and Evaluation |
| LRT | Language Resource(s) (and) Technology |
| MESO DARIAH WG. | n.d. 'Medievalist Sources (DARIAH Working Group) |
| META-NET | META-NET - META Multilingual Europe Technology Alliance |
| META-SHARE | META-SHARE - a Project of META-NET |
| MiBACT | Ministero dei Beni e delle Attività Culturali e del Turismo (Italy, see ICCU) |
| NeDiMAH | Network for Digital Methods in the Arts and Humanities |
| NIOD | Institute for War, Holocaust and Genocide studies (see KNAW) |
| NISO | National Information Standards Organization |
| NLP | Natural Language Processing |
| OAI-PMH | Open Archives Initiative Protocol for Metadata Harvesting |
| OEAW | Österreichische Akademie der Wissenschaften (Austria, PARTHENOS partner) |

| | |
|-----------|--|
| PARTHENOS | Pooling Activities, Resources and Tools for Heritage E-Research Networking, Optimization and Synergies |
| PERICLES | Promoting and Enhancing Reuse of Information throughout the Content Lifecycle Taking Account of Evolving Semantics |
| PIN SCRL | PIN SOC.CON.S. a r.l. (PIN is not an abbreviation) - Servizi Didattici e Scientifici per l'Universita' di Firenze (Italy, PARTHENOS partner) |
| PMH | Protocol for Metadata Harvesting |
| Q&A | Questions and Answers |
| RI | Research Infrastructure |
| SDM | Scholarly Domain Model |
| SISMEL | Societa Internazionale per lo Studio del Medioevo Latino (Italy, PARTHENOS partner) |
| SSH | Social Sciences and Humanities |
| ST | Sub-Task (within PARTHENOS; with number, e.g. ST2.1.1) |
| T | Task (within PARTHENOS; with number, e.g. T2.1) |
| TCD | Trinity College Dublin (Ireland, PARTHENOS partner) |
| TextGrid | Virtuelle Forschungsumgebung für die Geisteswissenschaften (virtual research environment for the humanities) |
| UGOE | Georg-August-Universität Göttingen (Germany, former PARTHENOS partner, replaced by FHP) |
| VRE | Virtual Research Environment |
| WP | Work Package (in particular within PARTHENOS; often with number, e.g. WP2) |

0. Overview and Methodology

Main author: Sebastian Drude (CLARIN)

0.0. Introduction; embedding

This document is the report on user requirements, deliverable D2.1 in the PARTHENOS project. It has been compiled as part of PARTHENOS Work Package 2 (WP2, for short) on “community involvement and requirements”, with input from members of all PARTHENOS partners.

The PARTHENOS project as a whole works on forming a cluster of infrastructures and similar initiatives that support research in the humanities in a broad sense, including language related studies, history, and archaeology, cultural heritage and related fields, and even certain social sciences (see later in this chapter for a more detailed characterization of the target user communities). In particular, PARTHENOS builds bridges between e-infrastructures, that is, infrastructures based on digital data and tools, usually ‘on-line’ (via the internet). It does so by (1) harmonizing and providing common solutions for policies throughout different phases of the data lifecycle; (2) identifying and supporting relevant standards, (3) establishing interoperability and a common semantic framework, (4) pooling, developing and adapting common tools for data-oriented services, and (5) joint training & education activities, and (6) coordinating networking and communication activities. All these activities are covered in PARTHENOS by dedicated Work Packages (WPs 3-8), which correspond closely to the Tasks within Work Package 2, as illustrated in the following table.

| Task in PARTHENOS Work Package 2 | Corresponding PARTHENOS Work Packages |
|--|--|
| T2.1: Definition of users’ requirements about data policies | WP3: Common policies and implementation strategies |
| T2.2: Definition of standardization requirements | WP4: Standardization |
| T2.3: Definition of interoperability & related services requirements | WP5: Interoperability and semantics WP6: Services and tools |

| | |
|---|---|
| T2.4: Def. of education & training requirements | WP7: Skills, Professional Development and Advancement |
| T2.5: Def. of communication requirements | WP8: Communication, dissemination and outreach (esp. T8.2 & T8.3) |

Accordingly, this report feeds into these other PARTHENOS Work Packages; much of the work on WPs 3–6 during the remainder of the project will be based on the results presented in this report, and the report will provide an important background to the work of WPs 7–8). The report is structured accordingly: each of the main chapters of this document has been developed by a single WP2 Task.²

0.1. Structure of this report

In the remainder of this introductory chapter we will first describe the “users” whose requirements are addressed in this document, and then explain how we proceeded to identify and present these requirements.

Chapter 1 addresses the topic of data policies, from various aspects that correspond to the three Sub-Tasks (ST) within T2.1, and which in turn are mirrored by three Tasks within WP3:

ST2.1.1: Definition of policy requirements concerning the data lifecycle (cf. T3.1)

ST2.1.2: Definition of policy requirements on quality assessment of digital repositories and quality assurance of data and metadata items (cf. T3.2)

ST2.1.3: Definition of policy requirements on IPR, Open Data and Open Access (cf. T3.3)

Chapter 2 is dedicated to standardization requirements. The approach taken in T2.2, in close coordination with WP4, differs somewhat from that taken by the other tasks. There are at best only a few, rather generic, *requirements on* standards (e.g., that they are clearly formulated and allow for being supported by relevant tools); standards in turn try to provide solutions for requirements such as (e.g.) interoperability, and there are accordingly *requirements that concern* standards. Hence, Chapter 2 contains rather a collection of use cases in which research practice could be considerably improved, either by implementing

² The term “Task”, with a capital T, is used here in a technical sense as a unit of project organization that is sub-ordinate to a Work Package. They are actually rather sub-work-packages or topic areas of responsibility.

standards, by enhancing or tailoring them, or sometimes just by applying existing but not known or considered standards.

Chapter 3 presents user requirements for common tools and interoperability, which will feed into shaping technical solutions and tools and their functions, to be developed in Work Packages 5 and 6. The use cases in Chapter 3 document requirements expressed by a large number of disciplines in the PARTHENOS community. Despite different approaches and methodologies, we found a number of general requirements. Other requirements, from both the backend and the frontend perspective, were gathered. The same holds for tools, for which we found a similar situation, in which there were shared priorities at the general level as well as some domain-driven requirements. Finally, a set of non-technical (or at least not only technical) requirements, such as the sustainability of tools and datasets, were expressed by the PARTHENOS community: we plan to consider these as action points for other WPs (specifically, WP3), and insert them into the agenda for the development of mid- and long-term actions and policies.

Chapter 4 contains requirements regarding education and training in digital methods at different stages of research careers (early, transitional, established), and also presents a collection of known syllabuses and curricula.

Finally, **Chapter 5** surveys communication requirements; co-organized scientific workshops and international conferences; and joint press releases/interviews and other publicity on themes of common interest.

0.2. User communities in PARTHENOS

0.2.1. Overview

PARTHENOS aims to serve the humanities in a broad sense, and may also be relevant for some social sciences and other neighbouring disciplines. Although sometimes treated as a single community (e.g. in the context of European research infrastructure consortia), this is actually a very broad and heterogeneous group, and PARTHENOS as a cluster of research infrastructure initiatives within this broad domain cannot serve all of them equally, but will have to prioritize. To identify the core user communities that PARTHENOS will focus on, we took a bottom-up approach, starting with the partners in PARTHENOS and the user groups they cater for or represent, either directly or indirectly, through their involvement in collaborative projects (see section 0.3, page 18 for a list of relevant projects).

Our survey of user communities relevant for PARTHENOS resulted in a list of disciplines, each of which was represented by between one and five partners (disregarding the too generic “humanities” or “Digital humanities”). We avoided going into the complexities of an ontology for scientific disciplines (although it seemed probable that some communities could be considered as subsets of others). Nevertheless, we organized these communities into the following larger groups:

- 1) **History** (in a broad sense: including Medieval Studies, Recent History, Art History, Epigraphy, etc.)³
- 2) **Language-related Studies** (including Literature, Linguistics, Philology, Language Technology, etc.)
- 3) **Archaeology, Heritage & Applied Disciplines** (including Cultural Heritage, Archives, Libraries, Museums, Preservation / Conservation experts, Digital curation / edition / publishing, etc.)
- 4) **Social Sciences** (in a broad sense: Sociology, Political Science, Geography, Anthropology, Cultural Studies etc.)

Of these, the first three were represented by a similar number (more than 20) of PARTHENOS partners. The social sciences were much less strongly represented (altogether eight partners). History, Language Studies and Heritage and Applied Disciplines can thus be considered the highest priority for PARTHENOS. The details can be found in [this online spreadsheet](#).

In what follows we give a short characterization of each of these broad groups.

0.2.2. History in a broad sense

Author: Emiliano Degl’Innocenti (CNR-OVI, formerly SISMEI)

The history group in PARTHENOS encompasses a vast set of disciplines and sub-communities. Some of them are driven by chronological borders and periodizations (i.e.: medieval studies and contemporary history), some others are mostly focussed on particular aspects or characteristics of the sources (i.e.: external/physical aspects for epigraphy, palaeography, codicology vs. internal aspects for art history, philology, etc.), and thus presenting different methods and research habits.

³ Initially, we tentatively included Archaeology with the historical disciplines in a comprehensive group “Studies of the Past”; but later it became clear that, in particular in terms of user requirements, archaeologists are better included in the Heritage & Applied group, which includes conservators and others who work mainly with physical objects. Although Archaeology is a large community within PARTHENOS, the re-grouping does not substantially alter the quantitative result of similarly strong representation of the major three communities reported below.

This richness and articulation is represented in the history group of PARTHENOS through the expertise brought by different partners and networks, in particular: CENDARI (TCD, SISMELE, KCL) and EHRI (KNAW-DANS, KCL), although other networks (CLARIN, DARIAH, Huma-Num etc.) also serve history in a broad sense.

For the above reasons the history group presents several similarities – but also relevant diversities – with the other communities represented in PARTHENOS. To name to just a few general examples, similarities could be found with Archaeology and Language related studies – due to shared methods and sources, while differences are with social sciences, where the notion of fact is characterized as “repeatable and measurable” while in history it has been characterized as “individual and unique” (Abbagnano 1959).

This situation is reflected also in the main findings: despite the different approach and methodological focus, we’ve found a number of general requirements that are shared across several use cases and disciplines, expressing the same needs e.g.: data quality, availability, accessibility and enrichment. Specific needs, such as visual media documents enrichment, integration of authority lists, gazetteers and reference tools and/or resources – are present at the level of sub-communities and disciplines, in correspondence with particular characteristics of the methods and/or sources involved. Other requirements were expressed both from the backend – such as storage and preservation – and the frontend – such as tools for collaborative work and data analysis. From the point of view of the tools – again – we found a shared set of priorities, i.e.: general (and advanced) search and information display, as well as some domain-driven requirements, e.g.: tools for preparing digital editions.

Finally, long-term issues such as the sustainability of tools and datasets were expressed by the history community: since those requirements do not only involve technical components, we plan to consider them as action points for other WPs (specifically, WP3), and insert them into the agenda for the development of mid- and long-term actions and policies.

0.2.3. Language Related Studies

Author: Sebastian Drude (CLARIN)

It is hard to delineate the borders of this group, because research in many disciplines is based on, or makes use of, language materials – including history, which often works with text, the exemplary language material. It is for this reason that the resources and tools de-

veloped by this group are relevant for many other user communities.⁴ However, for our purposes we consider a discipline to belong to this group if it addresses those *aspects and properties* of its objects of study that are language-related. Thus, the historian or sociologist (not part of this group) may be interested in the content and impact of, say, a novel, whereas the literary scholar (belonging to this group) will be interested in the way that language is used within a novel.

Thus delineated, this group shows still an enormous internal heterogeneity. Even within the group's core discipline, so to speak, Linguistics, there are very diverse research goals and methods; for example, the analytical levels of (i) sound, (ii) the inner structure of words and (iii) sentences, and (iv) their respective meanings, each constitute different subfields with different research workflows. Many linguists work with texts or single sentences, which constitute the prototypical datatype of this group, while others build lexical databases, and others compile treebanks that represent syntactic structure. Some analyse and annotate single sentences manually; others are interested in statistical analyses of large corpora. Linguistic typologists construct databases comparing forms or abstract features between different languages, and documentary linguists compile corpora of annotated multimedia recordings of natural speech. Psycholinguists perform experiments and measure reaction times or follow eye movements; recently, even Functional Magnetic Resonance Imaging (fMRI) and genetic data are being studied by linguists collaborating with other disciplines such as neurology and genetics. There are other branches of linguistics that are shared with other sciences, such as natural language processing (NLP), computational linguistics and language technologies, which are shared with informatics, and again use and produce data of quite different types, such as parsers, named entity recognizers, grammar systems, and speech recognition or synthesis technologies. Other disciplines outside linguistics, such as philologies of different languages or literary studies, have still different workflows, although in some cases they may make use of tools and datasets developed by the above.

Given this heterogeneity, it is difficult to provide summaries of user requirements that are valid even for a sizable part of this group, beyond very generic requirements concerning data management etc. Still, several use cases contain aspects that may be relevant also for other studies, and certain components re-occur in several workflows. Some tools de-

⁴ This is why CLARIN, which is one of the core participating infrastructures, and which focuses on language resources, is considered a research infrastructure for the humanities and social sciences at large (insofar as they make use of language resources).

veloped by computational linguists or NLP scholars are relevant for many other disciplines, in particular named entity recognition or different kinds of content extraction.

0.2.4. Cultural Heritage, applied disciplines, and Archaeology

Author: Paola Ronzino (PIN)

Museums, galleries, libraries and archives in Europe constitute a large and dynamic sector making an extraordinary cultural, educational, social and economic impact. They make their collections available on-line in accordance with common standards and using common services, and by doing so contribute to several European policy objectives for research and innovation.

These organizations manage and make available a vast quantity of digital data, including digital reproductions of books, paintings, museum objects, archival records, periodicals and millions of hours of film and video. Moreover, they actively cooperate with research centres on the development of innovative technologies for the conservation of cultural heritage artworks, including paintings, sculptures, metalwork, ceramics, manuscripts, printed books, archaeological objects, and others. Research on artwork materials and the development of related applications aiming at the conservation of cultural heritage may open a larger perspective on heritage conservation activities in Europe.

In this context, particularly relevant is Archaeology, an extensive and multi-disciplinary field that spans several domains of the humanities, natural sciences, cultural heritage research and public administration, and involves commercial services as well as academic scholarship. Archaeological research infrastructures form a very heterogeneous and fragmented landscape. The heterogeneity of the research methodologies of the archaeological community, together with the heterogeneity of the information technologies that are currently in use by researchers, are fundamental challenges that need to be addressed. Many archaeologists, like researchers in other disciplines, are not yet prepared to make data openly available outside a research project or organization. To address this issue, the ARIADNE project contributes to the emergence of a culture of open sharing of archaeological data, trusted data archives (where missing at present), and mobilization of data resources that are interoperable and re-useable. The project addresses the fragmentation of archaeological datasets in Europe, and aims to foster the (re-)use of data through the interoperability of digital archives and the implementation of an e-infrastructure that meets the needs of a large segment of the archaeological community. The infrastructure will support a culture of

sharing and the collaborative use of archaeological data across disciplinary, organizational and national boundaries.

EU initiatives such as DCH-RP (2012-2014) have facilitated cooperation between museums, galleries, libraries and archives, e-Infrastructure providers and research centres on the creation of a reference architecture for a more integrated and interoperable digital infrastructure for cultural heritage and digital humanities. The combined effort and commitment of such high-level partnerships resulted in studies that identified a common language and common vision for innovative solutions for data management, curation and access based on the potential of e-Infrastructure.

New innovative services are required to improve trans-national access, (re-)use, manipulation and long-term preservation of data, although commonalities and current solutions should be identified more clearly across the different domains. Several complex matters still need to be resolved, in particular concerning privacy management, data storage and security, IPR licence policies, interoperability of the platforms on various operating systems, data retrieval systems, and multilingualism and semantics.

0.2.5. Social Sciences in a broader sense

Authors: Mark Hedges (KCL), Adeline Joffres (Huma-Num), Emilie Kraaikamp (DANS)

The digital culture user community represented by KCL investigates the role, consequences and meaning of digital technologies within contemporary culture, addressing such topics as social media, gaming, digital memory, the digital economy, and politics. Research is both qualitative and quantitative, including for example text mining and other analytical methods, as well as critical and theoretical approaches.

Various disciplines are represented by Huma-Num, a large French research facility of the CNRS, which aims to help researchers in the humanities and social sciences apply digital technologies and the Semantic Web to process, enrich, and preserve their data. Represented communities are – besides the fields of history (medieval and contemporary), archaeology, linguistics, and literature, covered above – also geography, ethnology, and political science (mainly political sociology), as well as architecture and musicology. The requirements vary according to the different fields, but we can summarize it in broad terms in this way: data processing of native and non-native digital data (encoding, computation, data bases, and data migration), scanning, corpora creation and sharing, and archiving (long-term preservation, use of interoperable formats).

The social sciences user community represented by DANS, in turn, consists of both survey researchers and qualitative researchers. The qualitative researchers are mainly involved in conducting oral history research in the sub-disciplines sociology, psychology, and contemporary history. The survey researchers work in a broad range of sub-disciplines: sociology, political science, applied sciences, behavioural and educational sciences, communication sciences, psychology, and social geography. Additionally, this user community includes psychologists conducting statistical analyses of experimental studies.

0.3. Methods of identifying user requirements

Main author: Sebastian Drude (CLARIN), with contributions by other authors

The task of this document is to “define” user requirements.⁵ The first step in describing something is of course to know it, so in our case, to *identify* the user requirements that are out there, even if the users themselves are not in all cases aware of them. That is no easy task, and much research has been dedicated to the question of how this can best be achieved. Most often, research into user requirements employs surveys with questionnaires and/or interviews, but these need to be designed carefully, tested for usefulness, adjusted, and then sent out (or personally brought) to (usually) large target groups, often with low response rates.⁶ Given the timeframe in which WP2 was operating to produce this first deliverable, such an approach was not feasible. Luckily, it was also to a great extent not necessary: PARTHENOS was able to make use of earlier studies of this kind, many of them produced by previous or still ongoing projects in which PARTHENOS partners were or are involved.

Our general approach was, then, to:

- 1) Collect existing reports and similar documents that (may) contain user requirements,
- 2) Distribute them among the Tasks according to their main relevance,
- 3) Extract, within individual Tasks, relevant user requirement information from them, and
- 4) Present the user requirements so identified in the most coherent form possible.

Of course, the actual workflow and results in individual Tasks may have differed to a larger or smaller degree (see in particular above in section 0.1 the comments on Chapter 2, but

⁵ There are several possible meanings of “to define”; we use it here meaning “to describe something clearly, or to show all relevant aspects of something”, not in the sense of establishing the sense of a technical term. So defining user requirements is an empirical, not a policy-making enterprise.

⁶ Examples of this kind of report on user requirements include the ARIADNE Deliverable D2.1 and ARIADNE’s report on “Use Requirements”. (Selhofer and Geser 2014; Wright et al. 2014).

also Chapters 4 and 5 are somewhat different), but in general the approach proved useful. Where necessary, the user requirement information so collected has been complemented by other sources, such as individual interviews or other dedicated actions, which have been included in this final version.

WP2 used Zotero to collect the references to all possibly relevant documents; this proved useful, among other things, for producing the bibliography to this report. The WP2 Zotero Library is accessible via this link: https://www.zotero.org/groups/parthenos_wp2/items.

On the other hand, and in accordance with the general approach decided in PARTHENOS, WP2 used the D4Science Virtual Research Environment for collecting the actual reports and documents themselves. Specifically, we used the VRE folder at PARTHENOS > Work Package Activities > WP 2 - User requirements > User Requirement Documents. As some of the reports are internal and may not be freely distributed, we will unfortunately not be able to provide access to this folder beyond the PARTHENOS community. This is regrettable, as it may decrease the usability of this document for readers outside PARTHENOS. However, no research environment will be successful in gaining confidence and adherence if it does not itself respect privacy and other restrictions. In any case, the references at the end of the document should suffice for obtaining most of the material online, and, where necessary, for requesting access to the non-publically available documents from their respective holders.

By requesting input from all partners within PARTHENOS (all contribute to at least one, many to almost all tasks in WP2), WP2 was able to cover a large number of past and current projects with potentially relevant reports. Here is the complete list of projects that we took into consideration:⁷

AGORA, Apex, ARIADNE, Athena, CARMEN Worldwide Medievalists Network, CENDARI, Charisma / Iperion-CH, CLARIN, COST Action IS1005 (Medieval Europe), DARIAH, DARIAH-DE, DARIAH-GR, DARIAH-IT, DASISH, DCH-RP, DigCur, Digital Medievalist, DM2E, EHRI, EUDAT, Europeana Cloud, Flarenet, INDIGO-DataCloud, Isidore Platform, MESO DARIAH WG, Metashare, National Information Standards Organization (NISO) , Nedimah, PERICLES, TextGrid.

These projects were distributed among the PARTHENOS partners that showed most affinity with them; in most cases this was the partner who brought the respective project up as relevant, usually because the partner was involved in some way in the project. The partners were then asked to identify the relevant Tasks for which useful user requirements

⁷ For links and long names see the References at the end of this document.

could be extracted from the project's reports, and then the extraction work started, carried out either by the PARTHENOS partners participating in the respective Task, by other task members, or by the task leader.

Here is an overview of the projects covered in this review of documents, together with the distribution across partners and Tasks: https://docs.google.com/spreadsheets/d/1JrG1A2SUZEBiULXj9uPoqu66VCGHNRj_gkOqQLJraM/edit.

0.4. Methods of presenting user requirements

Main author: Sebastian Drude (CLARIN), with contributions by other authors

WP2 agreed to rely on real use cases as far as possible, and many of these were extracted from the reports and other documents examined in the review. WP2 aimed to present these use cases in a consistent way, and opted to apply the methodology proposed by Alistair Cockburn in his book *Writing Effective Use Cases* (2000).⁸

According to this approach, a good use case description should have certain elements.

A detailed overview of the elements included in the use cases section of this document is presented in the following table:

| Field Name | Explanation |
|---------------|--|
| Use case | The name of the use case. Could include or just be an ID-number. |
| User Story | A narrative description of the use case |
| Goal | A descriptive statement of the goal |
| Scope | The scope of the requirement |
| Preconditions | What is necessary for the realization of the goal |

⁸ WP2 gratefully recognizes that they, and others in PARTHENOS, were informed about this and other methods in a dedicated PARTHENOS webinar 'How to write use cases' by Edi Marchetti (ISTI/CNR), in September 2015. See <http://www.parthenos-project.eu/webinar-how-to-write-use-cases/>.

| | |
|-----------------------|---|
| Success End Condition | What is necessary for the realization of the goal |
| Failed End Condition | The state of the system if the goal is not achieved |
| Primary Actor | Who/what is the primary actor of the goal |
| Trigger | The event that causes the use case to be initiated |
| Extensions | Possible extensions to the basic flow of the use case, extending it |
| Frequency | An estimate of how often a particular use case will be exercised |
| Main Success Scenario | The basic flow for a use case in which nothing goes wrong |

As for the most part WP2 did not interact directly with users, but rather used a review of existing documents as its information source (see above), we were not able to ensure that information on each of these elements was available in all cases, even where it would make sense to provide it. This was an unavoidable consequence of relying mainly on secondary sources, and is another reason for undertaking complementary work in the future, which may lead to additions to this report. Furthermore, not all of these elements make sense in each use case, and it would of course be a mistake to artificially fill in information, which may be irrelevant, duplicated, or even guessed or invented, simply to provide content for each of them.

On the other hand, the different topics covered by the respective Tasks often required a different or adapted approach (we mentioned above the special procedure followed by T2.2). Therefore, some chapters will at their respective beginnings briefly summarize the specific approach that they takes to presenting the user requirement information.

As use cases play such an important role as basis for the actual requirements, which often only make sense (or at least can only be fully understood) in the context of a certain usage

scenario, this document may present use cases in several places. As mentioned previously, Task 2.2 relied on use cases more heavily than most other Tasks, and Chapter 2, Section 2.1 is therefore a primary place to go for someone consulting this document in search of use case descriptions.

Use case descriptions can be found in the following places in this document:

- Chapter 1, Subsections 1.1.3, 1.2.3, 1.2.5, 1.3.3, 1.3.5, 1.3.7, and 1.3.8.
- Chapter 2, Section 2.1.
- Chapter 3, Sections 3.6, 3.7, and 3.8.

1. Requirements concerning data policies

Overall coordination: Sara di Giorgio, with support from Antonio Davide Madonna and Marzia Piccininno (all MIBACT-ICCU)

1.0. Introduction

Authors: Sara di Giorgio, with support from Antonio Davide Madonna and Marzia Piccininno (all MIBACT-ICCU)

1.0.0. Overview

The main objective of T2.1 is to provide evidence about user requirements concerning all aspects of data policies, by collecting needs and experiences with the handling of data from the relevant stakeholders. This information will support the PARTHENOS project in taking informed decisions and will help to define guidelines for requirements concerning data policies.

The requirements were gathered from the different research communities identified within the project (see page (26), above), and reflect their specific needs; the results are briefly summarized in dedicated paragraphs, shown according to a simplified Cockburn schema (described above). A narrative use case was added to provide an example of the requirements that were gathered.

In particular, the mandate of sub-tasks 2.1.1 and 2.1.2 was to provide evidence on user requirements and related issues, notably through collecting feedback from the PARTHENOS user communities and through an analysis of prior work done within ESFRI and other integrating activities. This chapter describes user requirements as regards data production, storage, management, curation and long-term preservation (ST2.1.1), as well as requirements concerning the quality assessment of digital repositories, individual data items and individual metadata items, as expressed by the research communities involved in the project (ST2.1.2).

ST2.1.3 aimed to gather requirements about IPR (Intellectual Property Rights), Open Data and Open Access, both those expressed by the research communities involved in the project and others emerging from related national and European regulations. Its analysis is also based on prior work done within ESFRI and other integrating activities, and on the current panorama of access policies in EU countries. It describes the expressed policy re-

quirements and needs as regards IPR management, Open Data and Open Access policies.

This chapter reports the outcomes of the analysis carried out within sub-tasks 2.1.1, 2.1.2 and 2.1.3. This information will be used by WP3 as a roadmap for the definition and implementation of data lifecycle policies and related guidelines.

1.0.1. Gathering the requirements

The starting point of this task was to identify, collect and review literature that has explored the data requirements of researchers including their needs for support from data managers. The specific aim was to identify what is currently known about the needs relevant to data policies, as expressed by the different research communities. The review was guided by the gathering of a number of specific information sources, some of which were of a more generic nature, while others were focused more specifically on the identification of user needs that lead to the development of specific use cases. Documents and reports from different communities/projects (twenty for ST2.1.1, seventeen for ST2.1.2 and twelve for ST2.1.3) have been collected in the Zotero and D4Science platforms and have been analysed in order to extract requirements. The document analysis process was conducted by the T2.1 members according to their available resources, and can be found in this [online document](#).

The results of this review are summarized in the following sections. To provide easier access to the findings, details are presented clustered in columns to make them comparable. Although the data available for each community was of different level of detail, it is possible to give a useful overview of the needs of all communities.

1.1. Definition of Policy Requirements Concerning the Research Data Lifecycle

Main authors: Paola Ronzino (PIN), formerly Juliane Stiller (FHP), final edition: Claus Spiecker (FHP)

1.1.1. Introduction: Models of the Research Data Lifecycle

Research data has, in most cases, a longer lifespan than the research project that creates it. This holds especially when researchers continue to work on data after the funding has ended, or when data is reused by other researchers. Data becomes particularly valuable when it is well organised, documented, preserved and shared, because this facilitates the

advance of scientific inquiry and increases opportunities for learning and innovation (UK Data Archive 2016).

Various models of the data lifecycle have been implemented across the different research fields⁹. DARIAH-DE, for example, has worked on an elaborate data lifecycle model that includes the used data, the processing of the data through various research activities, and the resulting outcomes (Puhl et al. 2015), see Figure 1.

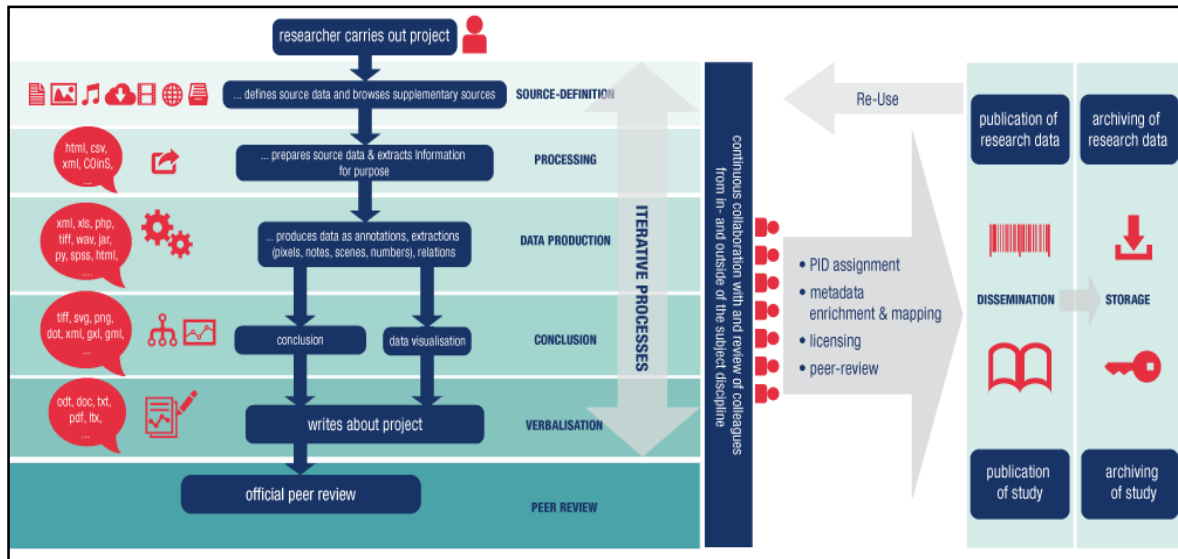


Figure 1 The DARIAH-DE data lifecycle model

The Digital Curation Centre (DCC 2016) has developed a model that helps institutions to plan their activities around data acquisition, access management, and long-term preservation.

On the other hand, the Data Documentation Initiative (DDI) – Lifecycle (Arofan 2011), the newer branch of the DDI family, focuses on the description of metadata as it is created and used throughout the data production lifecycle. Particularly, DDI is considered as supporting a metadata-driven survey design. In the case of a survey, for example, data produced from initial conceptualization to publication results in a huge amount of metadata. This metadata can be recorded in the DDI format and reused when the data collection, processing, tabulation, and reporting starts. As Arofan states, the DDI metadata is both documentary and also “machine-actionable”, meaning that it can be used to drive processes and to support additional steps in the data lifecycle (Figure 2).

⁹ See for instance CEOS / WGISS / DSIG (2011).

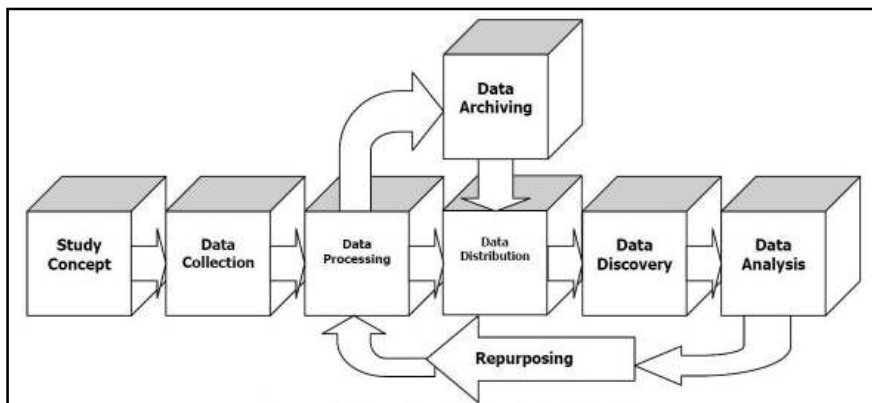


Figure 2 *The lifecycle model underlying the DDI Lifecycle specification*

Information used in historical research undergoes a production and consumption process in several distinct stages, each one of which is related to a specific transformation. These stages, which represent a major activity in the historical research process are referred to as a lifecycle (Boonstra, Breure, and Doorn 2004). The Lifecycle Historical Information Model is shown in Figure 3.



Figure 3 *The lifecycle of historical information*

This lifecycle model consists of six stages: creation, enrichment, editing, retrieval, analysis, and presentation.

Three practical aspects are in addition grouped in the middle of the lifecycle. These concern processes or issues that are central to computing in the humanities and are related to the six above-mentioned stages:

- Durability: this concerns the long-term deployment of the historical information produced;
- Usability: this concerns the ease of use, efficiency and satisfaction experienced by the intended audience when they use the information;

- Modelling: this refers to the more general modelling of research processes and historical information systems.

Moving on to the social sciences and humanities, the model developed by the UK Data Archive (UK Data Archive 2016) identifies various phases of the research data lifecycle, which are described below:

- 1) Creating data: this phase includes “design research”, “plan data management (format, storage)”, “plan consent for sharing”, “locate existing data (experiment, observe, measure, simulate)” and “capture and create metadata”;
- 2) Processing data: this phase includes “enter data, digitize, transcribe and translate”, “check, validate, clean data”, “anonymize data where necessary”, “describe data” and “manage and store data”;
- 3) Analysing data: this phase includes “interpret data”, “derive data”, “produce research outputs”, “author publications”, and “prepare data for preservation”;
- 4) Preserving data: this phase includes “migrate data to best format”, “migrate data to suitable medium”, “back-up and store data”, “create metadata and documentation” and “archive data”;
- 5) Giving access to data: this phase includes “distribute data”, “share data”, “control access”, “establish copyrights”, and “promote data”;
- 6) Re-using data: this phase includes “follow-up research”, “new research”, “undertake research reviews”, “scrutinise findings”, and “teach and learn”.



Figure 4 The UKDA research data lifecycle model

After an analysis of the various data lifecycle models, together with WP3 members, we decided to adopt the UKDA Research Data Lifecycle model as the backbone for aligning the

user requirements collected within the PARTHENOS community. This decision was taken because of the completeness and clarity of the various steps offered by the model, which would help us to identify a shared framework for the quality assessment of data and metadata, to identify common requirements and, finally, to produce guidelines defining common good practice for the research areas engaged in the project. These results will be achieved by WP3, with the contribution of WP2, subsequently to the assessment and the harmonization of the existing policies in use by the different disciplines.

1.1.2. Overview of the Analysis of the Research Data Lifecycle

The analysis of user requirements as documented in this section addresses important user needs, e.g. of scientists and data providers, with regard to all phases of the research data lifecycle. The research communities (see the introductory chapter) involved in this analysis require available data to be easy to find and accessible. Data and metadata quality are also relevant concerns for researchers and in particular for data managers. The analysis shows that more than 40% of the requirements collected from the different communities constituting the PARTHENOS consortium report that one of the major concerns regards data preservation; this holds especially for the archaeological community.

Although the various disciplines differ in the way their research is carried out, they have various factors in common when data storage and access are involved. The literature (Feijen 2011, Selhofer and Geser 2014, Van den Eynden et al. 2009) shows that data storage, data access, retrieval of stored data, and preservation of data for reuse are becoming major concerns for researchers. Problems are mainly caused by technical barriers, such as the use of obsolete software, or by non-technical barriers, such as the fear of competition, lack of trust, lack of incentives, and lack of control. It is not clear currently how these problems could be solved; nevertheless, all stakeholders, including funding agencies, data producers, data consumers, and data centres, agree that something needs to be done to improve the current situation.

One of the conclusions of the analysis carried out by Feijen (2011) on “what researchers need with respect to storing and accessing research data”, is that there is an important difference between data storage and access during a research project phase and data management after the publication of the research results. Researchers, indeed, have expressed a clear need for support in day-to-day storage, because they do not have the skills, awareness, or knowledge to improve their daily data management activities. On the

other hand, they see data preservation as a problem that falls somewhat outside their immediate scope of interest.

In order to be successful, services supporting researchers in the management of their digital data, have to meet some requirements. These requirements are summarised by Feijen (2011) as follows:

- Tools and services must be compatible researchers' workflows, which are often discipline-specific (and sometimes even project-specific).
- Tools and services must be easy to use.
- Researchers must be in control of what happens to their data, who has access to it, and under which conditions.
- Researchers expect tools and services to support their day-to-day work within the research project, and long-term/public requirements must be subordinate to that interest.

The survey additionally shows that investing time and effort in data management during the research phase is good practice, as it improves data preservation once the research phase has ended and data is published. Most researchers show reluctance to accept automatic responsibility for preserving their data after publication. Local storage is seen to offer them more control over their data than remote storage in a data centre. On the other hand, they admit that remote storage will probably alleviate some of the workload of data management. In any case, what counts more for researchers is that they wish to remain in control of their data when it is transferred to another party.

Another concern for researchers is data and metadata quality. Due to the lack of awareness about the importance of metadata for data sharing, most researchers often do not produce metadata for datasets they generate in projects. In order to allow data sharing, the effort needed to produce metadata needs to be covered somehow or other.

In particular, the community of archaeologists, represented by the ARIADNE project, criticizes the lack of transparency of available data and the difficulties in gaining access to it. A stakeholder survey carried out within the ARIADNE project (Selhofer and Geser 2014) revealed that researchers are primarily struggling to know what data exists. Data accessibility is also a very important point as data appears difficult to find. In most cases data is not online, and when online, it is difficult to access. The lack of downloadable "raw data" for reuse, such as the databases used for the creation of a scholarly edition, or for the production of 3D models of artefacts, is also a major concern among archaeologists.

The evidence collected by the PARTHENOS literature review, which involved examining eighteen reports/deliverables, documented the enormous degree of fragmentation with re-

gard to data that might usefully be integrated, presented by a complex diversity of data habitats and different types of repositories.

As a general conclusion to the survey among the archaeological community, the major barriers to data accessibility are: (i) cost, e.g. for obtaining licences to use pictures, or for subscription fees, and (ii) the problem that relevant literature and data are often kept in lots of different places, e.g. in many different private collections of other researchers..

With regard to the international dimension, access to a wider geographical distribution of datasets seems to be important in facilitating collaboration between researchers at different institutes and in enhancing funding opportunities.

As regards scholarly activities in digital humanities, the work carried out within the project DM2E (Digitised Manuscripts to Europeana) and reported in (Hennicke et al. 2015), presents recommendations for future work on digital humanities virtual research environments (VREs). One of the main findings of the theoretical and empirical research is that scholars and computer scientists should refer to a model like the Scholarly Domain Model (SDM) to increase the sustainability of the VRE. The SDM consists of four different layers of abstraction including Areas, Scholarly Primitives, Scholarly Activities and Scholarly Operations. The proposed approach is essential for VREs if they are to comprehend the entire scholarly research process and to offer applications and services that can support the corresponding workflows.

Taking all requirements analysed into consideration, it is evident that PARTHENOS offers a broad field of opportunities for creating real value for users. While it is clear that the project cannot solve immediately all problems, it is true that PARTHENOS may have a high impact if the guidelines implemented can deliver improvement in any of the areas discussed above.

The reader should bear in mind that the present study offers a summary of what was collected by different contributors, with different backgrounds. We have tried to report the requirements as given in the sources analysed, and not to interpret or to add further comments. Nevertheless, we are aware that by opting for specific sources we have already made a selection. Therefore we want to emphasise that, although this analysis represents the best of our current knowledge on the basis of the sources used, it cannot be regarded as exhaustive and may not represent the concerns of the four communities identified by PARTHENOS to the same depth.

The following table provides an overview of the requirements collected among researchers representing the domains of History, Archaeology, Heritage and Applied Disciplines, and

Social Science and Language related studies. The matrix shows the individual requirements expressed by the various communities, and tries to indicate the level of importance of each of the requirements identified by PARTHENOS for the research data lifecycle. For this overview each of the use cases listed in the tables of the following section (1.1.2.) were analysed and grouped into 22 requirements (RQ) according to the commonalities of their scope. As the basis of data is heterogeneous and doesn't cover the humanities landscape for each domain to the same depth, the overview can provide only an indication of the importance of each requirement for a specific domain.¹⁰

| Requirements for re-search data lifecycle | History | Archaeology, Heritage and applied disciplines | Social Science | Language related studies |
|---|-------------------|--|-----------------------|---------------------------------|
| RQ-1 Production of appropriate and machine-readable metadata | n/a ¹¹ | n/a | ++ | n/a |
| RQ-2 Support multilinguality | n/a | n/a | + | n/a |
| RQ-3 Enriching data dissemination | n/a | n/a | + | n/a |
| RQ-4 Data quality | n/a | ++ | n/a | n/a |
| RQ-5 High quality metadata | + | +++ | n/a | n/a |
| RQ-6 Assign Persistent Identifiers | + | +++ | + | +++ |
| RQ-7 Research Data Preservation | + | +++++ | + | ++ |
| RQ-8 | + | ++ | + | ++ |

¹⁰ The use cases are described in more detail in the tables of Section 1.1.2 and references to the sources are provided.

¹¹ n/a: not available, i.e. there is no information on this specific point in the surveyed documents.

| | | | | |
|---|------|--------|------|--------|
| Foster use of a sustainability model | | | | |
| RQ-9 Transparency of available data | n/a | + | n/a | + |
| RQ-10 Harmonization of access regulation | + | + | + | n/a |
| RQ-11 Establish metadata standard for research process | + | + | + | + |
| RQ-12 Data access availability and control | ++++ | ++++++ | ++++ | ++++++ |
| RQ-13 Long-term scientific data re-use | + | + | + | n/a |
| RQ-14 Recognition of data sharing | n/a | + | n/a | n/a |
| RQ-15 International dimension | n/a | + | n/a | n/a |
| RQ-16 Implement system for role and rights management | +++ | +++ | +++ | +++ |
| RQ-17 Preserve documents in editable formats | + | + | + | + |
| RQ-18 Use of common metadata standards | + | + | n/a | +++ |
| RQ-19 Definition of common termi- | + | + | + | ++ |

| | | | | |
|---|-----|-----|-----|-----|
| nological resources | | | | |
| RQ-20 Use of friendly implementation and appropriate web services for the data | +++ | n/a | n/a | ++ |
| RQ-21 Add additional information to existing resources, e.g. interpretation, spatial and temporal context and changes in time & space to historical resources, bibliography and citations. | + | + | n/a | + |
| RQ-22 Use of formalized and free format, validation | +++ | +++ | +++ | +++ |

1.1.3. Results – Definition of Policy Requirements Concerning the Data Lifecycle

This section reports the requirements collected by PARTHENOS grouped according to the functions identified by the UKDA research data lifecycle model.

1.1.3.1. Creating data

This phase of the research data lifecycle includes “design research”, “plan data management (formats, storage)”, “plan consent for sharing”, “locate existing data”, “collect data (experiment, observe, measure, simulate)” and “capture and create metadata” (UK Data Archive, 2016).

1.1.3.1.1. Use Case #01_DDI: Production of machine-readable metadata (Arofan 2011)

Goal: To enable a mechanism for the identification of the metadata elements used by a given community or organization

Scope: Description of metadata as they are created and used throughout the data production lifecycle

Pre-conditions: Metadata is provided; standard formats and elements are used

Success End Conditions: A smaller agreed set of information can be identified for use

Failed End Condition: Impossible to retrieve dataset information

Primary actor: Research institutes conducting large-scale longitudinal or repeating cross-sectional surveys; research data centres and similar organizations; data producers

Trigger: Document the data sets that data managers will archive and disseminate to researchers

1.1.3.1.2. Use Case #02_DDI: Support multilingualism (Arofan 2011)

Goal: All human-readable text can be provided in different languages

Scope: Exchange of metadata and data between organizations

Pre-conditions: Multilingual thesauri used; translation tools applied

Success End Conditions: Retrieve dataset information in different languages

Failed End Condition: Impossible to retrieve dataset information in different languages

Primary actor: Research institutes; research data centres; data producers

Trigger: Search for metadata among different organizations

1.1.3.1.3. Use Case #03_DDI: Enriched data for dissemination (Arofan 2011)

Goal: Create metadata describing the process of collection & tabulation for enabling linking to disseminated aggregates

Scope: Create standard metadata model for microdata , and its processing and tabulation

Pre-conditions: Tools and services used to record microdata

Success End Conditions: Linked resources found

Failed End Condition: No links available

Primary actor: Research institutes; researchers; data centres; data producers

Trigger: Find linked resources

1.1.3.1.4. Use Case #01_MUSE: Accountability (Bird & Simons 2003)

Goal: Provide full documentation on which language descriptions are based.

Scope: Access to transcriptions and recordings

Pre-conditions: Application of: Recording tools; transcriptions tools; language verification tools

Success End Conditions: Full documentation available, i.e. a grammar is based on a text corpus

Failed End Condition: No or inadequate documentation

Primary actor: Researcher

Trigger: A researcher wants to provide the full documentation on which language descriptions are based

1.1.3.1.5. Use Case #12_ARIADNE: Standards for excavation and site/monument data (Papatheodorou et.al. 2013)

Goal: The community needs a set of international standards specifically for excavation and site/monument data

Scope: Develop tools and guidance based on international standards

Pre-conditions: International standards (that can be adapted as required)

Success End Conditions: Implementation of standards and use by the archaeology community

Failed End Condition: Lack of adoption of standards

Primary actor: Researchers; information manager

Trigger: A researcher is interested in achieving a high level of datasets integration fit for processing

1.1.3.2. Processing data

This phase of the research data lifecycle includes “enter data, digitize, transcribe & translate”, “check, validate, clean data”, “anonymize data where necessary”, “describe data” and “manage and store data” (UK Data Archive 2016).

1.1.3.2.1. Use Case #01_DCHRP: Ingestion (DCH-RP 2014)

Goal: Ingestion of different record types to an e-Infrastructure-based preservation system

Scope: System

Pre-conditions: A Cultural Institution has data and wants to ingest it in a storage preservation system

Success End Conditions: The data is ingested

Failed End Condition: The storage service is not able to ingest different types of records

Primary actor: A user in the role of content provider

Trigger: A user ingests data in a storage service

1.1.3.2.2. Use Case #02_DCHRP: Checking data (DCH-RP 2014)

Goal: Automatic checking of ingested data with tools able to verify integrity and consistency

Scope: System

Pre-conditions: Data from a Cultural Institution is ingested

Success End Conditions: Data is successfully checked

Failed End Condition: Data ingested is not complete or consistent

Primary actor: A user in the role of content provider

Trigger: A user launches a tool to check data

1.1.3.2.3. Use Case #03_DCHRP: Fixing information (DCH-RP 2014)

Goal: Data ingested is assigned a persistent identifier, which will allow to identify and to check file integrity

Scope: System

Pre-conditions: Data from a Cultural Institution is ingested and checked

Success End Conditions: Data is fixed

Failed End Condition: It is not possible to identify data after the fixing process

Primary actor: A user in the role of content provider

Trigger: A user launches a tool to fix data

1.1.3.2.4. Use Case #04_DCHRP: Storage (DCH-RP 2014)

Goal: An e-infrastructure-based preservation system should store files in a way that allows their full accessibility and usability

Scope: System

Pre-conditions: Information on formats and standards for raw data is available – appropriate metadata standards are in place as well as a trustworthy strategy for replacing obsolete technology

Success End Conditions: Data is accessible and usable

Failed End Condition: It is not possible to find adequate information

Primary actor: Collection manager

Trigger: A user looking for data ingested

1.1.3.3. Analysing data

This phase of the research data lifecycle includes “interpret data”, “derive data”, “produce research outputs”, “author publications” and “prepare data for preservation” (UK Data Archive 2016).

1.1.3.3.1. Use Case #01_AHC: Add interpretation to digital historical resources (Boonstra et al. 2004)

Goal: Interpretations need to be added to digital historical resources since a certain piece of data lacks meaning without interpretations

Scope: Interpreting historical resources

Pre-conditions: As interpretation is subjective, it needs to be added, in such a way that it can be separated from the original data in the source

Success End Conditions: Researchers are able to add interpretation to historical resources such that it exists separately from the source

Failed End Condition: Researchers are not able to add interpretation to historical resources or it is not separate from the source data

Primary actor: Historian in the role of data consumer

Trigger: Interpretation of historical resources to enhance existing data

1.1.3.3.2. Use Case #02_AHC: Add spatial and temporal context (Boonstra et al. 2004)

Goal: Spatial and temporal context needs to be added to data to understand its meaning

Scope: Linking sources

Pre-conditions: Availability of spatial and temporal information

Success End Conditions: Data can be accessed and analysed by researchers according to temporal and spatial criteria

Failed End Condition: Incomplete spatial and temporal contexts, lack of linking

Primary actor: Historian in the role of data consumer

Trigger: Research on linked sources

1.1.3.3.3. Use Case #03_AHC: Take into account the changes of time and space (Boonstra et al. 2004)

Goal: Since historical research deals with changes in time and space, analysis tools need to take into account the changes of time and space

Scope: Linking sources

Pre-conditions: Availability of time and space information

Success End Conditions: Tools available to researchers that perform analysis taking changes of time and space into account

Failed End Condition: Researchers are not able to perform analysis taking time and space changes into account.

Primary actor: Historian in the role of data consumer

Trigger: Research on linked sources

1.1.3.4. Preserving Data

This phase of the research data lifecycle includes “migrate data to best format”, “migrate data to suitable medium”, “back-up and store data”, “create metadata and documentation” and “archive data” (UK Data Archive 2016).

1.1.3.4.1. Use Case #05_DCHRP: Active digital preservation – Schedule-based integrity checking (DCH-RP 2014)

Goal: To verify, in an automatic way, the integrity of data on a regular basis

Scope: System

Pre-conditions: Data of a Cultural Institution is ingested and validated

Success End Conditions: Data is checked and they are valid

Failed End Condition: Uploaded data is not valid

Primary actor: A user in the role of content provider

Trigger: A user launches a tool to check data integrity

1.1.3.4.2. Use Case #06_DCHRP: Active digital preservation – De-referencing and deleting (DCH-RP 2014)

Goal: It is possible to de-reference and delete data

Scope: System

Pre-conditions: There is obsolete data or data with incorrect referencing

Success End Conditions: Data is deleted or dereferenced

Failed End Condition: The user has no rights to amend uploaded data

Primary actor: A user in the role of content provider

Trigger: To update the digital collections

1.1.3.4.3. Use Case #09_DCHRP: Active digital preservation – Data migration (DCH-RP 2014)

Goal: Migration of preserved files to new versions of software and/or hardware

Scope: System

Pre-conditions: A new version of a software / hardware is available

Success End Conditions: Data is migrated to a new platform / service

Failed End Condition: Data is not compatible with new platform / service

Primary actor: A user in the role of content provider

Trigger: Platform updating

1.1.3.4.4. Use Case #10_DCHRP: Active digital preservation – Possibilities to export data (DCH-RP 2014)

Goal: It is possible to export data in specific formats (CSV, XML, XSL)

Scope: System

Pre-conditions: Data is ingested and validated and is CSV, XML, XSL compatible

Success End Conditions: Data is exported in the requested format

Failed End Condition: Data is not available in a requested format

Primary actor: A user in the role of content provider

Trigger: A user download data from the platform

1.1.3.4.5. Use Case #11_DCHRP: Active digital preservation – Conversion and transformation of data (DCH-RP 2014)

Goal: The digital objects are converted into new standardised file formats

Scope: System

Pre-conditions: Data is available in a standard format (nonproprietary format)

Success End Conditions: Data ingested is converted and transformed in another format

Failed End Condition: The user is alerted with a message about the impossibility to process the document

Primary actor: A user in the role of data consumer

Trigger: A collection manager wants avoid technical obsolescence

1.1.3.4.6. Use Case #12_DCHRP: Active digital preservation – OAIS standard model (DCH-RP 2014)

Goal: Guarantee long term preservation of digital collection through an ISO standard model

Scope: System

Pre-conditions: Existing digital collections

Success End Conditions: Implementation of a digital library based on the OAIS 6 functions.

Failed End Condition: Data is not accessible

Primary actor: Digital content manager

Trigger: An institution wants to preserve digital collection

1.1.3.4.7. Use Case #01_AGORA: Involving scholars in assessing quality of digital content (Marras & De Grandis 2014)

Goal: Enhancement of existing resources and creation of new resources and tools for research

Scope: Scholars who question the scholarly quality of some digital content they had to use

Pre-conditions: 1. Bottom-up approach; 2. Adoption of two paradigms of software development: feature driven development (FDD) and value sensitive design (VSD)

Success End Conditions: Scholars are systematically involved at every stage of software design, development, testing and customization

Failed End Condition: Scholars are not involved in the design and development of software

Primary actor: Scholars; Software engineers

Trigger: The start of a collaboration with other institutions

1.1.3.4.8. Use Case #02_AGORA: Adequate training about the technical aspects of scholars' work (Marras & De Grandis 2014)

Goal: Enhancing existing resources; encoding the text according to the standards in use

Scope: Scholars and users collaborating with research institute

Pre-conditions: A set of standard code in the field of humanities; an expert/tutor to teach it to scholars and users

Success End Conditions: A cycle of lessons about theory and practice of the standard code in the field of humanities

Failed End Condition: It is impossible to organize adequate training

Primary actor: Scholars; users

Trigger: The organization of similar training programmes

1.1.3.4.9. Use Case #01_CLARIN: Appropriate metadata generation for the type of resource involved (Quochi et al. 2009)

Goal: Each of the resource types used in the social science research field needs to be appropriately described

Scope: Metadata creation and data storage

Pre-conditions: Semantic interoperability of metadata schemas

Success End Conditions: Use of ISO 24622-X or equivalent standard

Failed End Condition: Refusal of archives to cater for the user needs

Primary actor: Research infrastructures; data provider

Trigger: A researcher wants to retrieve information about a digital resource

1.1.3.4.10. Use Case #02_CLARIN: Reporting during the depositing workflow (Quochi et al. 2009)

Goal: A user submitting resources to a RI wants to know the status of the submission process

Scope: Depositing process

Pre-conditions: Existing depositing process (defined); archive manager in place

Success End Conditions: Successful usability test

Failed End Condition: Refusal of archives to implement

Primary actor: Archive

Trigger: Researchers require better control over data life cycle

1.1.3.4.11. Use Case #01_DASISH: High quality metadata (DASISH 2014b)

Goal: Improve metadata quality based on an analysis of the different metadata strategies of CLARIN, DARIAH and CESSDA

Scope: eScience infrastructure

Pre-conditions: Communication between involved actors

Success End Conditions: Apply new or update existing standards and re-enrich metadata to meet the changing needs of their target community

Failed End Condition: Data and metadata is not used by the target community

Primary actor: Infrastructure managers

Trigger: Changing needs of target community

1.1.3.4.12. Use Case #01_DISC: Use of common standards (D'lorio 2009)

Goal: The federation creates its own metadata standard to ensure the highest quality bibliographic description for all materials in its nodes

Scope: Research and educational institutions in the field of humanities, with a special regard to philosophy

Pre-conditions: Metadata sets exist for 1. the digitisation of historical materials, or 2. the creation of original, born-digital works

Success End Conditions: The federation metadata standard allows project partners to catalogue works

Failed End Condition: There is a limited description of the texts found in the platforms of the federation

Primary actor: European research institutions; researchers

Trigger: The start of a collaboration with other institutions

1.1.3.4.13. Use Case #02_DISC: Stable/persistent web address (D'lorio 2009)

Goal: A federation of semantic digital libraries in the field of philosophy will collect resources (philosophical texts, primary sources, videos); the federation will ensure the reliability of scholarly reference, meaning that resources can be suitable for scholarly quotations

Scope: Research and educational institutions in the field of humanities, with a special regard to philosophy

Pre-conditions: A stable URI identifies all resources published

Success End Conditions: A software platform able to handle a wide range of resources including texts, images, and videos identified by a stable URI

Failed End Condition: The software platform is not suitable to the different needs of any member of federation

Primary actor: European research institutions

Trigger: The start of a collaboration with other institutions

1.1.3.4.14. Use Case #03_DISC: Semantic annotation of documents with dedicated tools (D'Iorio 2009)

Goal: Creating a common ontology by merging existing narrower domain source ontologies from cultural heritage institutions

Scope: Research and educational institutions in the field of humanities, with a special regard to philosophy

Pre-conditions: 1. Semantic web applications; 2. guidelines about ontologies construction, particularly in the field of humanities

Success End Conditions: Each federation website exposes to various semantic web applications a set of public recommended ontologies using in semantic annotations; scholars can use the suggested ontologies or are able to extend them with new concepts and relationships or to design their own ontologies thus enabling personal annotation environments

Failed End Condition: Web semantic applications are not incorporated in the web site or the ontologies remain incomplete

Primary actor: European research institutions; researchers

Trigger: The start of a collaboration with other institutions

1.1.3.4.15. Use Case #01_DM2E: Evolving VRE (Hennicke et al. 2015)

Goal: Define a multi-layered scholarly domain model (SDM)

Scope: Virtual research environments (VREs)

Pre-conditions: Buyin from the humanities community, developments are user-focused and not technology driven

Success End Conditions: SDM "as a reference model for the discussion, evaluation and development of digital research infrastructures for the humanities"

Failed End Condition: No adaption of a VRE "to evolving scholarly practices" (p. 26)

Primary actor: VRE managers

Trigger: Evolving scholarly practices

1.1.3.4.16. Use Case #03_ESFRI: Preservation of public-funded research data for long-term scientific reuse (SSH RWG 2008)

Goal: Data should be available over the long term, and someone or some organization must ensure this

Scope: Researchers are able to find and reuse data readily even if it was created some time ago

Pre-conditions: Knowledge of where the repositories are, existence of repositories with solid funding models and strong technical infrastructure (including migration capacity)

Success End Conditions: Data is available when researchers require it

Failed End Condition: Data succumbs to institutional failure or 'bit rot'

Primary actor: Data repository, but also the public agencies likely to fund such repositories

Trigger: Researcher deposits data (then lots of time passes and it is still available)

1.1.3.4.17. Use Case #12_MUSE: Coverage (Bird & Simons 2003)

Goal: Document the 'multimedia linguistic field methods' that were used

Scope: Access to multimedia linguistic field methods

Pre-conditions: Linguistic tools

Success End Conditions: Make rich records of rich interactions, especially in the case of endangered languages or genres.

Failed End Condition: Impossibility to create a documentation about the multimedia linguistic field methods

Primary actor: Researcher

Trigger: A researcher is interested in a study about the multimedia linguistic field methods

1.1.3.4.18. Use Case #01_PARSE: Research data preservation (Kuipers & Hoeven 2009)

Goal: Build an international e-infrastructure for data preservation and access

Scope: Long-term availability of research data

Pre-conditions: Funding (mainly public) training, more expertise, more resources, more digital repositories

Success End Conditions: The results become public property and properly preserved; reanalysis of existing data; interdisciplinary collaborations; advancement of science (new research can build on existing knowledge); validation; economic value

Failed End Condition: Users may be unable to understand or use the data Lack of sustainable hardware, software or support of computer environment

Primary actor: Data Managers (data centres, digital archives, etc.); other actors Publishers Researchers Funders

Trigger: Digital resources must persist and remain findable, accessible, and understandable

1.1.3.4.19. Use Case #01_FLaReNet: Foster use of a sustainability model (Soria et al. 2012)

Goal: The developed resource remains accessible and usable in a long-term perspective

Scope: Usage of resources

Pre-conditions: Resources are maintained in a long term repository: data is not lost; standards and technologies are updated over time

Success End Conditions: Long after the development of the resource, the resource remains accessible to the public; people know about its existence and can still use it as it was conceived

Failed End Condition: Some time after the development of the resource it is no longer accessible to the public; or people don't know about its existence; or they can no longer use it as it was conceived

Primary actor: Producer

Trigger: Language resources (LR) production / publication

1.1.3.4.20. Use Case #01_ARIADNE: Transparency of available data (Selhofer & Geser 2014)

Goal: To reduce the lack of data transparency reported by the archaeological community

Scope: Data access

Pre-conditions: Metadata is available

Success End Conditions: Provenance information is available

Failed End Condition: Adoption of inconsistent interfaces, insufficient provenance information and scattered and heterogeneous resources

Primary actor: Archaeologist in the role of data consumer

Trigger: An archaeologist wishes to use data

1.1.3.4.21. Use Case #01_HumaNum: Use or migrate data in open source and free formats (Rouchon et al. 2011)

Goal: Clean, migrate and preserve data from proprietary formats to open source and free formats

Scope: Audio (sounds, speeches, etc.) and audiovisual data

Pre-conditions: Check data integrity (bits packets, metadata)

Success End Conditions: To be able to migrate the data

Failed End Condition: Migration fail and loss of data or of data quality

Primary actor: The data provider (researcher, data producer)

Trigger: To analyse the data quality and define a “target format”

1.1.3.4.22. Use Case #02_Huma-Num: Use normalized formats (Rouchon et al. 2011)

Goal: Normalize the data to promote interoperability (through metadata)

Scope: Audio (sounds, speeches, etc.) and audio-visual data

Pre-conditions: Define or find a corresponding norm

Success End Conditions: To find a norm corresponding to the data you treat

Failed End Condition: Inability to find a norm or outdated norm

Primary actor: The data provider (researcher, data producer) AND a normalization process or an operator that provides the norm

Trigger: Having set the encoding level of the data relative to the selected standard and defined as a target standard

1.1.3.4.23. Use Case #03_Huma-Num: Check the formats (Rouchon et al. 2011)

Goal: Check the quality data and formats

Scope: Audio (sounds, speeches, etc.) and audio-visual data

Pre-conditions: Define the formats' validation level (quality criteria) to thin the data

Success End Conditions: To be able to validate the format

Failed End Condition: To be faced with “not clean” data and formats

Primary actor: The data provider (researcher, data producer) AND the provider of the tool's algorithm (Ex : the CINES in France)

Trigger: Having defined the quality and data rejection level in relation to the verification format

1.1.3.4.24. Use Case #01_GOEDOC: Assigning persistent identifiers (Puhl et al. 2015)

Goal: Enable data sustainability

Scope: Long-term availability of research data

Pre-conditions: Long-term data storage and archival procedures in place, funding support

Success End Conditions: Improved access to data, increasing amount of resources

Failed End Condition: Data is lost, no improvements in sustainability

Primary actor: Data managers (data centres, digital archives, etc.); publishers; researchers; other actors

Trigger: Researchers accessing data and resources through repositories, references, etc.

1.1.3.4.25. Use Case #02_GOEDOC: Establish metadata standards for research processes (Puhl et al. 2015)

Goal: Normalize the data to promote interoperability (through metadata)

Scope: Metadata to support the data life cycle

Pre-conditions: Appropriate metadata standards exist

Success End Conditions: Metadata standards integrated into the data lifecycle

Failed End Condition: No or little metadata about the research process

Primary actor: Data producers

Trigger: Researchers needing context and information about data

1.1.3.4.26. Use Case #03_GOEDOC: Documents preserved in editable formats (Puhl et al. 2015)

Goal: Enable data sustainability

Scope: Long-term availability of research data and documents, reusability

Pre-conditions: Repositories etc. policies allow document storage in editable formats.

Success End Conditions: Researchers are able to reuse and edit resources

Failed End Condition: Researchers are not able to reuse and edit resources

Primary actor: Data Managers (data centres, digital archives, etc.)

Trigger: Editing of documents, updates

1.1.3.4.27. Use Case #02_MUSE: Terminology (Bird & Simons 2003)

Goal: Map the terminology, element tags and symbols and abbreviations used in description to a common ontology of linguistic terms

Scope: Access to a common ontology of linguistic terms

Pre-conditions: Linguistic and ontology tools

Success End Conditions: Have a map for the terminology, element tags and symbols and abbreviations used in description to a common ontology of linguistic terms

Failed End Condition: No common ontologies available.

Primary actor: Researcher

Trigger: A researcher wants to map several elements used in descriptive markup to a common ontology of linguistic terms

1.1.3.4.28. Use Case #03_MUSE: Existence (Bird & Simons 2003)

Goal: 1. List all language resources with an OLAC repository; 2. any resource presented in HTML on the web should contain metadata with keywords and description for use by conventional search engines

Scope: Access to OLAC repository

Pre-conditions: HTML customization tool; metadata customization tool

Success End Conditions: It is possible to list all language resources with an OLAC repository

Failed End Condition: Language resources not available in an OLAC compliant format

Primary actor: Researcher

Trigger: The researcher wants to have the data in an OLAC-compliant format

1.1.3.4.29. Use Case #04_MUSE: Relevance (Bird & Simons 2003)

Goal: Follow the OLAC recommendations on best practice for describing language resources using metadata, especially concerning language identification and linguistic data type

Scope: Access to OLAC repository and OLAC recommendations

Pre-conditions: Metadata customization tool

Success End Conditions: The highest possibility of discovery by interested users in the OLAC union catalogue hosted on the LINGUIST List site

Failed End Condition: No improved discovery

Primary actor: Researcher

Trigger: The researcher wants to identify language and linguistic data types

1.1.3.4.30. Use Case #01_SURFSHARE: Preservation of data (Feijen 2011)

Goal: Preservation of research data after the publication phase

Scope: Long-term availability of research data and documents

Pre-conditions: Storage during the research project phase has been well managed

Success End Conditions: Research data remains available and accessible to researchers in the long term

Failed End Condition: Research data unavailable after publication phase

Primary actor: Data Managers (data centres, digital archives, etc.)

Trigger: Researchers want access to research data after publication phase

1.1.3.5. Giving Access to data

This phase of the research data lifecycle includes “distribute data”, “share data”, “control access”, “establish copyright”, and “promote data” (UK Data Archive 2016).

1.1.3.5.1. Use Case #02_DASISH: Common list of metadata elements (DASISH 2014b)

Goal: Harvest metadata and make it available in a metadata catalogue for browsing and searching

Scope: Different eScience infrastructures

Pre-conditions: Definition of “a common list of metadata elements that could be deployed across the different communities”

Success End Conditions: Development of a Joint Metadata Domain (DASISH JMD)

Failed End Condition: No aggregation of metadata between eScience infrastructures

Primary actor: Researchers from different research communities

Trigger: Possibility for cross-fertilisation between eScience infrastructures

1.1.3.5.2. Use Case #03_CLARIN Ease of interaction with repositories/resource registries (Quochi et al. 2009)

Goal: For social science and humanities users provision of data and reuse of data is still “alien” hence the support of the users’ needs to take account of a low threshold

Scope: Research Infrastructures

Pre-conditions: Usable “products” of the RI

Success End Conditions: Successful usability tests

Failed End Condition: Devastating usability tests

Primary actor: Research Infrastructure

Trigger: Implementation of data sharing and reuse policies

1.1.3.5.3. Use Case #04_CLARIN: Interaction of resources with appropriate web services for the data (Quochi et al. 2009)

Goal: Reusing data often requires appropriate analysis tools which may be complicated for the novice or intermediate user

Scope: Research Infrastructures

Pre-conditions: Analysis tools

Success End Conditions: Robustness of tools e.g. to solve the problem of installation and reuse, web services are appropriate. Based on REST or SOAP they can even be integrated in standalone tools

Failed End Condition: Unavailability of tools

Primary actor: Research Infrastructure

Trigger: Research Infrastructure (RI) supporting and encouraging users who are novice or intermediate users

1.1.3.5.4. Use Case #01_ESFRI: Harmonization of access regulations (SSH RWG 2008)

Goal: Policies regulating access to data should be harmonised

Scope: Researchers should be able to understand and use access policies across data stores

Pre-conditions: Data repositories need to work together to ensure less variance in their access policies

Success End Conditions: A common set of practices and policies that can be applied widely

Failed End Condition: That no harmonisation occurs and researchers cannot access data because they either do not understand, cannot navigate or cannot fulfil access conditions

Primary actor: Data repositories

Trigger: When a researcher needs data from multiple sources

1.1.3.5.5. Use Case #02_ESFRI: Free data access (SSH RWG 2008)

Goal: Researchers should have unpaid access to the research data of their peers

Scope: Researchers are able to find and reuse data readily and/or deposit their research data for others to use at the conclusion of a project

Pre-conditions: Knowledge of how to publish in a free and open manner; repositories where data can be stored and found

Success End Conditions: Researchers are aware of existence of policies of open repositories and use them easily

Failed End Condition: Either researchers do not know of repositories or do not use them

Primary actor: Researcher (though there are many other peripheral players, including repositories and funding agencies)

Trigger: Researcher either needs data or has created data and is looking to make it available

1.1.3.5.6. Use Case #01_WDL: High definition digital object and high quality metadata Participating to a world digital library with some digital objects (WDL Content Selection Committee 2015)

Goal: An institution would like to make available rare copies of philosophical texts, granting the preservation

Scope: Research and educational institutions in the field of humanities, with a special regard to philosophy, could access to the texts

Pre-conditions: High definition digital object and high quality metadata, including place, time period, date created, item type, topic, contributing institution and language

Success End Conditions: A description, written by scholars in jargon-free language, accompanies every item, explaining what the item is and why it is important.

Failed End Condition: It is not possible to make a suitable description of the item to present complex material to students as well to scholars

Primary actor: Research institution in the field of philosophy

Trigger: A world digital library makes freely available resources for use and reuse by students and scholars

1.1.3.5.7. Use Case #02_ARIADNE: Improvement of data accessibility (Selhofer & Geser 2014)

Goal: Improvement of data accessibility; data appears as difficult to find, because not online, and when online, difficult to access

Scope: Data access

Pre-conditions: Availability of: reference tools, cross search tools, knowledge on the structure of the different resources involved

Success End Conditions: The researcher obtains a list of relevant datasets

Failed End Condition: Information about relevant datasets is not accessible; user must perform many different searches

Primary actor: Archaeologist in the role of data consumer

Trigger: Researcher seeking data

1.1.3.5.8. Use Case #03_ARIADNE: Perceived lack of recognition for data sharing (Selhofer & Geser 2014)

Goal: Have a common practice for publishing and sharing of data in national data archives or international repositories

Scope: Data sharing

Pre-conditions: Data providers willing to implement new policies and solve IPR issues

Success End Conditions: More data available, improved access and sharing

Failed End Condition: Data providers not willing to implement new policies and solve IPR issues

Primary actor: Archaeologist in the role of data producer

Trigger: Researcher seeking data

1.1.3.5.9. Use Case #04_ARIADNE: Free access to data (Selhofer & Geser 2014)

Goal: Improved access to data

Scope: Data access

Pre-conditions: IPR allows free access

Success End Conditions: More data available, improved access and sharing

Failed End Condition: Data providers not willing to implement new policies and solve IPR issues.

Primary actor: Archaeologist in the role of data consumer

Trigger: Researcher seeking data

1.1.3.5.10. Use Case #05_ARIADNE: International dimension (Selhofer & Geser 2014)

Goal: Access to a wider geographical dataset will help facilitate crosscollaboration and enhance funding opportunities between researchers from different institutes

Scope: Data access

Pre-conditions: Data providers willing to implement new policies and solve IPR issues

Success End Conditions: More data available, improved access and sharing across national boundaries

Failed End Condition: Lack of data access across national boundaries, no improvement

Primary actor: Archaeologist in the role of data producer

Trigger: Researcher seeking data

1.1.3.5.11. Use Case #04_GOEDOC: Implement system for role and rights management (Puhl et al. 2015)

Goal: Enable appropriate access to data, ensure IPR integrity

Scope: Data access

Pre-conditions: IPR information is available, uniform and consistent for data

Success End Conditions: More data available to researchers with clear use conditions

Failed End Condition: IPR remains a barrier to data access

Primary actor: Data repositories

Trigger: Researcher requires IPR status of data

1.1.3.5.12. Use Case #05_MUSE: Citation Bibliography (Bird & Simons 2003)

Goal: 1. Provide complete bibliographic information in the metadata for all language resources created; 2. provide complete citations for all language resources used; 3. Use the metadata record of a language resource to document its relationship to other resources (e.g. in the OLAC context, use the RELATION element)

Scope: Metadata manipulating tool

Pre-conditions: Metadata manipulation tools

Success End Conditions: Provide a complete bibliographic data in the metadata for all language resources created, with complete citations and the possibility to build relationships between different resources

Failed End Condition: Lack of or incomplete citation and bibliography information for language resources

Primary actor: Researcher

Trigger: A researcher

1.1.3.5.13. Use Case #06_MUSE: Persistence (Bird & Simons 2003)

Goal: Ensure that resources have an unique and possibly persistent identifiers, such as an ISBN, an OAI identifier, or a DOI (digital object identifier)

Scope: Data preservation and sustainability

Pre-conditions: Repository and data centres implement and support persistent identifiers

Success End Conditions: Provide a unique identifier to each object avoiding the possibility of ambiguity

Failed End Condition: Data is lost or becomes inaccessible

Primary actor: Data provider

Trigger: Data provider wants each element to have assigned an identifier which is guaranteed to be unique among all identifiers used for those objects and for a specific purpose

1.1.3.5.14. Use Case #07_MUSE: Balance (Bird & Simons 2003)

Goal: Limit any stipulations of sensitivity to the sensitive sections of the resource, permitting nonsensitive sections to be disseminated more freely

Scope: Data access

Pre-conditions: Local legal constraints access tools

Success End Conditions: In case of local legal constraints block, provide a partial access instead of a completely close access

Failed End Condition: Data remains blocked

Primary actor: Data provider

Trigger: If some content is partially blocked (due to local legal constraints) data providers prefer to grant partial access to that content

1.1.3.5.15. Use Case #01_VRE: Access to data (Carusi & Reimer 2010)

Goal: Access to data, tools, computational resources and collaborators leads to faster research results and novel research directions

Scope: Data access

Pre-conditions: Data available online

Success End Conditions: More data available, improved access

Failed End Condition: Lack of downloadable "raw data" for reuse

Primary actor: Researcher in the role of data consumer

Trigger: Researcher seeking data

1.1.3.5.16. Use Case #02_VRE: Sustainability (Carusi & Reimer 2010)

Goal: VREs have to be supported and used by research communities in order to be viable

Scope: VRE

Pre-conditions: Support and funding for VRE and researchers

Success End Conditions: VREs become integrated into the data life cycle

Failed End Condition: VREs not used and supported enough, not the norm

Primary actor: Researcher in the role of data consumer

Trigger: Researchers seeking to collaborate

1.1.3.5.17. Use Case #03_VRE: Data usability (Carusi & Reimer 2010)

Goal: In order to support collaborative and cooperative activities, it is important that virtual environments offer the means to access appropriate information as well as communication

Scope: VRE

Pre-conditions: VREs are designed and implemented according to researchers requirements

Success End Conditions: Possibility to use and reuse data

Failed End Condition: No possibility to access data

Primary actor: Researchers in the role of data consumers

Trigger: Researchers seeking to collaborate

1.1.3.5.18. Use Case #04_VRE: authentication and rights management (Carusi & Reimer 2010)

Goal: Providing general VRE frameworks that can be used to develop and host different VREs

Scope: VRE

Pre-conditions: Metadata to support rights management

Success End Conditions: Provision of core services (such as authentication and rights management; repositories; project planning, collaboration and communication tools)

Failed End Condition: VRE are not flexible enough

Primary actor: Researchers

Trigger: Researchers seeking to collaborate

1.1.3.5.19. Use Case #05_VRE: Formation of common vocabularies (Carusi & Reimer 2010)

Goal: A major shift in research practices will occur through the formation of common vocabularies as researchers collaborate with others across disciplinary, institutional and national boundaries

Scope: VRE

Pre-conditions: This will occur through the production of common taxonomies, data standards and metadata.

Success End Conditions: Semantic web approaches are seen as helpful in this context

Failed End Condition: Lack of common standards, too many disparate solutions

Primary actor: Researchers

Trigger: Increasing need for multinational and cross-discipline research

1.1.3.5.20. Use Case #06_VRE: Promote a set of policies and legal frameworks (Carusi & Reimer 2010)

Goal: It is extremely important that all stakeholders in the development of VREs come together to promote a set of policies and legal frameworks that will allow sharing of data and other resources in a transparent and comprehensible way

Scope: VRE

Pre-conditions: Investment in the research, development and implementation of agreed policies and legal frameworks

Success End Conditions: Agreement on policies and legal frameworks, implementation of these in VREs and RIs

Failed End Condition: Lack of common policies and legal frameworks; too many disparate solutions

Primary actor: Stakeholders in the role of data providers

Trigger: Increasing need for multinational and cross-discipline research

1.1.3.5.21. Use Case #02_SURFSHARE: Data access control (Feijen 2011)

Goal: Leave to the researcher/data provider/data owner the possibility to remain in control of his data also after transferring the data to another party

Scope: Data access

Pre-conditions: IPR of data is clear, common policies and standards are adopted

Success End Conditions: Researchers able to deposit data with confidence

Failed End Condition: No confidence; data not made available

Primary actor: Data managers (data centres, digital archives, etc.); researchers

Trigger: Researcher makes data available and for reuse

1.1.3.6. Re-using data

This phase of the research data lifecycle includes “follow-up research”, “new research”, “undertake research reviews”, “scrutinise findings”, and “teach and learn” (UK Data Archive 2016).

1.1.3.6.1. Use Case #04_CLARIN: Interaction of resources with appropriate web services for the data (Quochi et al. 2009)

Goal: Get robust and appropriate analysis tools for data reusing suitable also for novice or intermediate users

Scope: Research Infrastructures

Pre-conditions: Analysis tools

Success End Conditions: Robustness of tools e.g. to solve the problem of installation and reuse, web services are appropriate. Based on REST or SOAP they can even be integrated in standalone tools

Failed End Condition: Unavailability of tools

Primary actor: Research Infrastructure

Trigger: Reusing data

1.1.3.6.2. Use Case #08_MUSE: Use and reuse (Bird & Simons 2003)

Goal: Publish documentation and descriptions in such a way that users can gain access to the files to manipulate them in novel ways

Scope: Foster data reuse; access to documentation and data description

Pre-conditions: Data manipulation tools

Success End Conditions: Users can gain access to the files to manipulate them in novel ways, i.e.: not only traditional forms of publication through a fixed user interface like a web search form, or a fixed presentation view like a PDF file

Failed End Condition: Documentation and descriptions are published through a fixed user interface like a web search form, or a fixed presentation view like a PDF file which cannot be manipulated

Primary actor: Researcher in linguistic studies, in the role of data provider

Trigger: Data Provider wants to foster data reuse

1.1.3.6.3. Use Case #09_MUSE: Immutability (Bird & Simons 2003)

Goal: Distinguish multiple versions with a version number or date, and assign a distinct identifier to each version

Scope: Versioning of digital assets

Pre-conditions: Different working versions of digital resources spanning different person, places and time

Success End Conditions: Unique identification of digital resources and their versions

Failed End Condition: New versions of digital resources are not detectable in the authoring process

Primary actor: Researcher in linguistic studies, in the role of data provider

Trigger: Data provider wants to create a file versioning of the objects according to the dates and the authors that have modified it

1.1.3.6.4. Use Case #10_MUSE: Longevity (Bird & Simons 2003)

Goal: Commit all documentation and description to a digital archive that can credibly promise longterm preservation and access

Scope: Longterm availability of research data

Pre-conditions: Longterm data storage and archival procedures in place, funding support

Success End Conditions: Improved access to data; increasing amount of resources

Failed End Condition: Data is lost; no improvements in sustainability

Primary actor: Researcher in linguistic studies, in the role of data provider

Trigger: Data Provider wants to grant a longterm preservation for the data

1.1.3.6.5. Use Case #11_MUSE: Safety (Bird & Simons 2003)

Goal: Ensure that copies of archived documentation and description are kept at multiple locations

Scope: Longterm availability of research data

Pre-conditions: Backup tools

Success End Conditions: Have periodic copies of archived documentation

Failed End Condition: Data is lost, no improvements in sustainability

Primary actor: Researcher in linguistic studies, in the role of data provider

Trigger: Data provider wants to have an available backup copy in case of loss or failure

1.2. Definition of Policy Requirements on Quality Assessment of Digital Repositories and Quality Assurance of Data and Metadata Items

Main authors: Paola Ronzino (PIN), formerly Juliane Stiller (FHP), final edition: Claus Spiecker (FHP)

1.2.1. Introduction

An important aspect of the relationship between repositories and their stakeholders is trust. This holds particularly for data producers and consumers, who need to be assured that the archive or repository where their data is stored preserves the authenticity and integrity of the data.

To demonstrate their trustworthiness to stakeholders, repositories have conducted assessments and self-audits against standards and criteria catalogues. As reported in the CESSDA “User guide for digital preservation” (CESSDA 2015), after the publication of the OAIS Reference Model in 2002, archives and repositories started to refer to themselves as “OAIS-compliant”. This served to demonstrate that they could be trusted regarding preservation and dissemination of digital assets. Progressively, various checklists and criteria catalogues were developed to assess the trustworthiness of digital archives and for the purpose of certification. The list includes:

- Trusted Digital Repositories: Attributes and Responsibilities. An RLG-OCLC Report (RLG, 2002)
- DRAMBORA: Digital Repository Audit Method Based on Risk Assessment (DCC & DPE, 2007)
- Trustworthy Repositories Audit & Certification: Criteria and Checklist (OCLC & CRL, 2007)
- Nestor criteria. Catalogue of Criteria for Trusted Digital Repositories. Version 2 (Nestor, 2009)
- European Framework for Audit and Certification of Digital Repositories (2010)
- Audit and Certification of Trustworthy Digital Repositories. Recommended Practice (CCSDS, 2011)
- Data Seal of Approval. Quality Guidelines for Digital Research Data (2009, 2013)
- DIN 31644: Criteria for trustworthy digital archives (2012)
- ISO 16363: Audit and certification of trustworthy digital repositories (2012)

Repositories can acquire a basic certification through the Data Seal of Approval (DSA), which consists of a set of 16 guidelines relating to data producers, repositories, and users. To obtain the DSA, repositories carry out a self-assessment using the guidelines. The procedure is finalized by a member of the DSA board who reviews the assessment and documentation provided.

1.2.2. Overview of the Analysis of Quality Assessment of Digital Repositories

The following section provides a summary of the results derived from the conducted analysis. The possibility of storing research data in reliable repositories is seen by researcher communities as an important consideration for quality assurance. Inclusion of repositories in everyday scientific work, as well as their contribution to quality assurance, varies according to discipline.

From the analysis of the collected requirements, it emerges that to support researchers in quality assurance of data it is necessary to establish discipline-specific services of data management, which are in line with the discipline's requirements.

Great importance is attached within data management to the selection and verifiability of data in standardized form. The continued development of this process is very important.

In the social sciences and humanities there is a huge variety of resource types, from text collections, to video (e.g. filmed situations in laboratories from various angles, or remotely

in a rural village), from experimental measurements (such as EEG) to manual annotations. Each of these types needs to be described appropriately. However, as the descriptions have mainly to be produced by users; the descriptive features used need to be easy to understand, while those not needed by the user may be left out. The selection and promotion of high-quality deposit services is very important for researchers in the social sciences and humanities. Suggestions for service improvements therefore include:

- use of a PID system
- use of a Federated Identity

Moreover, it is required that for the management system, a model based on the data-PASS model (Data Preservation Alliance for the Social Sciences, see references) should be developed, together with clear guidelines and procedures for management, archiving, and sharing of data. Even though this model was originally developed for the Social Sciences, it is important for the humanities too, e.g. to keep a clear track of the origin of the data.

1.2.3. Results of the quality assessment

The requirements collected in this section have been split into two parts – one for data producers and one for repositories (as specified by the DSA).

1.2.3.1. Data producers

1.2.3.1.1. Use Case #05_CLARIN: Metadata needs to be flexible and adjustable to the needs of repositories, resources and users (Quochi et al. 2009)

Goal: The huge variety of resource types produced in the SS need to be appropriately describable. The descriptions have to be mainly produced by users; the descriptive features used need to be easy to understand by users

Scope: Metadata creation (data creator); archive manager (consulting the depositor which metadata to use); infrastructure provider/archive (supporting different types of resources and allow for a highly adaptable metadata schema)

Pre-conditions: Semantic interoperability of metadata schemas, i.e. descriptive categories need to be reused where possible; reuse of groups of descriptive categories.

Success End Conditions: Use of ISO 24622X or equivalent standard

Failed End Condition: Refusal of archives to cater for the user needs

Primary actor: Research infrastructures; data provider

Trigger: Retrieve information about repositories, resources and actors' role

1.2.3.1.2. Use Case #06_CLARIN: Use of multiple metadata schemas to avoid huge complexity based on unified schemas for vast amount of different data types (Quochi et al. 2009)

Goal: Too complex structures for metadata lend themselves to tag abuse and have to be avoided; RDF as a maintenance format is unusable for data providers, if there is no appropriate tool; especially when using a variety of editors RDF will result in inconsistencies

Scope: Metadata creation (data creator); archive manager (consulting the depositor which metadata to use); infrastructure provider/archive (supporting different types of resources and allow for a highly adaptable metadata schema)

Pre-conditions: Semantic interoperability of metadata schemas, i.e. descriptive categories need to be reused where possible; reuse of groups of descriptive categories

Success End Conditions: Use of ISO 24622X or equivalent standard

Failed End Condition: Refusal of archives to cater for the user needs

Primary actor: Research infrastructures; data provider

Trigger: Retrieve information about the semantics of digital resources

1.2.3.1.3. Use Case #07_CLARIN: Metadata description appropriate for researchers of different disciplines (Quochi et al. 2009)

Goal: SSH is interdisciplinary per definition, descriptions of services and resources need to take into account that different terminology and descriptions might be required and even self evident information may have to be made explicit

Scope: RI and domain experts

Pre-conditions: Willingness to interact

Success End Conditions: List of key concepts with general definition

Failed End Condition: Lack of general information

Primary actor: Research infrastructure, Data providers

Trigger: Retrieve information about digital resources in a crossdisciplinary environment

1.2.3.1.4. Use Case #09_CLARIN: Support for legacy data (Quochi et al. 2009)

Goal: There are vast amounts of legacy data out there with various file formats. Legacy data can be extremely labour intensive to bring it up to state of the art

Scope: Research Infrastructure + data providers

Pre-conditions: Recommended data formats

Success End Conditions: Digital resource preservation

Failed End Condition: Impossible to access or process data stored in an obsolete format

Primary actor: Research infrastructure, Data providers

Trigger: Data storage

1.2.3.2. Repositories

1.2.3.2.1. Use Case #08_CLARIN: Multilingual support (Quochi et al. 2009)

Goal: Get multilingual support for SSH resources and communities

Scope: Research Infrastructure

Pre-conditions: Multilingual support in tools and data formats; utf8 support, xml:lang (or equivalent)

Success End Conditions: Resources with multiple languages in them can be processed

Failed End Condition: Only monolingual resources or even worse: only one language support

Primary actor: Research Infrastructure

Trigger: Access multilingual resources

1.2.3.2.2. Use Case #10_CLARIN: Tools available as web services (Quochi et al. 2009)

Goal: Reusing data often requires appropriate analysis tools which may be complicated for the novice or intermediate user. To solve the problem of installation and reuse, web services are appropriate. Based on REST or SOAP they can even be integrated in standalone tools

Scope: Research Infrastructures and Community

Pre-conditions: Existing tools

Success End Conditions: Recommendation of tools, especially online available

Failed End Condition: Established proprietary tools with idiosyncratic data formats unintelligible to RIs

Primary actor: Research Infrastructure

Trigger: Access and process digital resources

1.2.3.2.3. Use Case #03_DASISH: Providing a data repository (DASISH 2012c)

Goal: An institution wants to provide a trustworthy environment for data management and data curation

Scope: Institutions in the field of social sciences, arts and humanities

Pre-conditions: A linear step-by-step implementation tool for repository building is available

Success End Conditions: The institution builds a repository

Failed End Condition: No repository is built

Primary actor: Research institution

Trigger: Need for a trustworthy environment for data management and data curation

1.2.3.2.4. Use Case #04_DASISH: Making an institutional repository available for external researchers (DASISH 2014d)

Goal: An institution would like to open its repository for use by researchers from external institutions

Scope: Researchers and institutions in the field of social sciences, arts and humanities

Pre-conditions: A persistent identifier system and a federated identity management system are available

Success End Conditions: External researchers are granted access to the institutional repository

Failed End Condition: The systems do not work and the institutional repository is not accessible to external researchers

Primary actor: Institution hosting a repository

Trigger: PI system and a federated identity management system are available and make it possible to open a repository for external researchers

1.2.3.3. Data and metadata quality assessment

Quality, uniqueness, risk of loss, repeatability, production costs, and potential for reuse are identified as factors that confer value to research data. The deep study carried out within the ARIADNE project (Geser and Selhofer 2015) is a particularly valuable source of information for the innovation needs in archaeology, concerning, among other things, open data sharing, digital archives and research infrastructures and services. The report informs us about the advice that the archive community gives to holders of research data. In particular, this includes recommending that holders of research data should engage in the general evaluation and selection criteria for valuable data, and that data should be curated and made accessible to the wider research community.

To achieve an increase in value, research data needs to be shared openly and reused, a conclusion supported by Palmer et al. (2011) and Weber et al. (2012). These sources confirm the potential of scientific data as it is central and of high value for the reuse of the findings, and that data increases in value through exposure to diverse contexts of use. When researchers see a potential benefit from open data sharing, data becomes more valuable.

Unfortunately, citation of data producers is still an uncommon practice when research data is made available, whether the data is associated with publication or released by researchers independently. A possible solution could be that scientists or data providers in general who share their data are invited as co-authors of publications that build on the data. Co-authorship is, possibly, more welcome if the researchers that provide the data, are also involved in the projects that reuse their data (Geser and Selhofer 2015).

The question of metadata starts to become fundamental when researchers share data through a repository or an archive. The ARIADNE survey found out that, when data is shared through digital repositories, researchers consider the effort required to provide the metadata as a barrier to open data sharing. While data repositories and users would benefit from rich and complete metadata, data providers usually prefer not to invest much effort on providing metadata. This unwillingness results in fewer contributions with metadata that is insufficient to allow data reuse. Potential re-users need metadata that is rich enough to help them to understand the provenance and context of the data, and to enable them to evaluate of the relevance of the data, and use it properly to prevent incorrect conclusions. Fulfilling the request for high-quality metadata seems currently to be possible only for domain-based archives that are mandated or recognized as the best place to share valuable data according to community standards.

Examples of data archives that devote special attention to measures for high quality metadata, trust and credibility are:

the Archaeology Data Service (ADS) in the UK¹², the Data Archiving and Networked Services (DANS)¹³ established in the Netherlands in 2005¹⁴, and other digital archives that are certified according to the DSA criteria.

1.2.4. Overview of the Analysis of Data and Metadata Assessment

The analysis of the requirements concerning data and metadata assessment, shows that there is a considerable demand from the different communities for good quality metadata. In practice we have seen that there are many repositories that have only weak metadata, in the sense that it is general-purpose and only specific for a certain university and not standard to all (Geser and Selhofer 2015).

¹² ADS is the mandated archive for data of many projects funded by the Arts and Humanities Research Council and the Natural Environment Research Council as well as the archive recommended by the British Academy, Council for British Archaeology, English Heritage and the Society of Antiquaries.

¹³ The DANS-EASY system includes the E-Depot Dutch Archaeology 'EDNA' (DANS 2016); over 80% of the data deposited in EDNA are publicly accessible.

¹⁴ DANS stores data from archaeologists since 2007, according to the Quality Standard for Dutch Archaeology (Kwaliteitsnorm Archeologie).

One of the use cases proposed by the archaeological community¹⁵ requires that domain-based, specialized and mandated archives should set high-quality metadata standards, which depositors will be obligated to accept and follow, guided by archive curators.

Moreover, defining a common set of standards that meet the needs of the sector is seen as the key to interoperability for the archaeological community (Papatheodorou et al. 2013).

As regards the reuse of available data from other researchers, the major requirement is that data must be relevant, understandable and trustworthy. The description of the data should be rich enough to allow users to discover, understand the provenance, and evaluate the trustworthiness and quality of the data. In order to be reused, data must be available under an adequate licence. Furthermore, data should be open to the greatest extent and with the fewest constraints possible, especially when it comes from publicly funded research (Geser and Selhofer 2015).

As regards data accessible online, researchers of the archaeological community report that sometimes data is not as useful as it could be, due to the lack of standardization, or because data is structured in different ways, is not up to date, or is incomplete or lacking important details.

A common issue addressed by each community involved in the collection of requirements is the effectiveness of metadata for cross-domain data reuse. Research datasets that are shared through repositories must be provided with metadata, to enable data discovery and access. For effective reuse, data should be machine-readable and in open, non-proprietary formats. In some research areas the availability of the software used to compute a specific analysis is very important, as it allows researchers to reproduce and validate research results and repurpose the data.

For the language related studies community, metadata harmonization is intended as the challenge of verifying that there is structural and syntactic interoperability between the resources that use the same property, for example Dublin Core's language property. Furthermore, the LR community wishes to detect duplicates that occur, either due to the original representation or as a result of combining multiple sources. When a large number of records describe the same resource, then queries for that resource will return too many records, which may lead to errors on the part of the users (McCrae et al. 2015).

¹⁵ Although many examples are taken from the archaeological community, due to the fact that the data from the ARIADNE community was very well known to the PARTHENOS project partners, the findings are very similar in other communities.

A controlled vocabulary service has the potential to improve the quality and consistency of metadata descriptions. Such a list of terms could for instance contain organization names and MIME types, and guide users when generating metadata.

In the context of META-SHARE, the network that aims to provide an open and secure infrastructure for the language technology domain, the term 'metadata' refers to descriptions of language resources (LR), including both data and technologies (tools/services)¹⁶ used for their processing. The mechanism adopted by META-SHARE is the component-based mechanism (Component Metadata Infrastructure, CMDI), where semantically coherent elements are grouped together to form components (Broeder et al., 2014, Choukri et al. 2011). In this context, elements are used to encode specific descriptive features of the LRs. Furthermore, links to conceptually similar existing elements in the Dublin Core (DC 2016) and the ISO Data Category Registry (ISO DCR, [ISO 12620, 2009]) with other related schemas and models, are provided to cater for semantic consistency.

A study on the current state of research and practice regarding metadata quality, focusing on the functional perspective and on evaluation criteria, and examining mechanisms for improving metadata quality, has been carried out by Drexel University, Philadelphia (Jung-Ran Park 2009). According to the author, metadata quality is a reflection of the degree to which the metadata supports functionality relating to discovery, use, provenance, currency, authentication, and administration. The study emphasises that accuracy, completeness, and consistency are the most common criteria used in measuring metadata quality. The results of the study show (i) that there is a pressing need for a common data model to support interoperability of data across digital repositories, and (ii) that the inclusion of guidelines within Web forms or templates for entering metadata is of great value for improving metadata quality.

We conclude from our analysis that the completeness of metadata should be assessed in terms of its intended functionality. Metadata should at least indicate the type of resource described and its relation to the local collection, as well as the guidelines followed in creating the metadata. Metadata should provide a description of the resource that is accurate, and consistent at both a conceptual and structural level, and moreover it should be sufficient to support functionality relating to discovery, use, provenance, currency, authentication, and administration.

¹⁶ These are also found in the literature as Language Resources and Technologies (LRTs).

1.2.5. Results – Metadata and Data Quality Assessment

1.2.5.1.1. Use Case #03_AGORA: TEI schema for secondary materials Partners in a project have to apply a standard when converting existing documents and creating new resources for inclusion within the project framework (Burnard 2011)

Goal: To mark-up with XML-TEI all secondary material (existing published or unpublished contemporary critical material such as articles or reports) to be integrated in the project scholarly space

Scope: Content providers in a scholarly or project digital archive

Pre-conditions: A common XML-TEI schema according to different secondary material in the field of philosophy

Success End Conditions: The defined schema corresponds to the data to be encoded and the secondary materials can be upload in the digital archive

Failed End Condition: The secondary material is only partially encoded

Primary actor: Partners in a project

Trigger: The start of a collaboration with other institutions in a project

1.2.5.1.2. Use Case #06_ARIADNE: Definition of a common set of standards (Papathodorou et al. 2013)

Goal: The key for interoperability is the definition of a common set of standards which meet the needs of the sector. Data that are accessible online, sometimes is not as useful as it could be, because data is structured in different ways, not up to date, incomplete or lack important details

Scope: Interoperability

Pre-conditions: Mapping data to a common standard

Success End Conditions: The common set of standards support the discovery of similar resources

Failed End Condition: The system does not provide any useful result

Primary actor: Data managers

Trigger: A user wants to access online data

1.2.5.1.3. Use Case #01_RIN (RIN 2008)

Goal: The datasets must provide an appropriate record of the work that has been undertaken, so that it can be checked and validated by other researchers

Scope: System

Pre-conditions: The dataset is checked by a peer reviewer

Success End Conditions: In the record all the information about review process are registered

Failed End Condition: The record has no links with the dataset

Primary actor: A user in the role of peer reviewer

Trigger: A user uses a tool to put information about the review process

1.2.5.1.4. Use Case #02_RIN (RIN 2008)

Goal: Datasets are discoverable, accessible and re-usable

Scope: System

Pre-conditions: Datasets are indexed by PARTHENOS

Success End Conditions: Datasets are showed in a repository with information about re-usability

Failed End Condition: Datasets cannot be reused due to copyright status

Primary actor: A user in the role of data consumer

Trigger: A user searches / browses the datasets of a content provider

1.2.5.1.5. Use Case #03_RIN (RIN 2008)

Goal: Division of datasets to make easier the review of them entirely

Scope: Guideline / system

Pre-conditions: Dataset has a critical mass of data

Success End Conditions: Dataset is split in different subsets

Failed End Condition: The split datasets are too large

Primary actor: A user in the role of Content Provider

Trigger: A user wants to upload a large dataset and reads the guideline to split it in different subsets

1.2.5.1.6. Use Case #04_RIN (RIN 2008)

Goal: Instituting a formal process of review in two stages with a focus on content and technical merit

Scope: Guideline / System

Pre-conditions: Data is ingested

Success End Conditions: The two step of review: 1. review of content by peer reviewer; 2. automatic review of data structure

Failed End Condition: The formal process of review failed

Primary actor: A user in the role of a peer reviewer

Trigger: A peer reviewer reads the guideline to check data according to a defined formal process; a peer review launches a tool to check the data structure

1.2.5.1.7. Use Case #01_LREC: Community should help in documenting existing language resources (Calzolari et al. 2012)

Goal: All LRs should get at least a minimal description and documentation in a basic catalogue, including minor ones, and those still in progress; this is achieved by crowd-sourcing information on LRs used in papers at major conferences

Scope: Documentation

Pre-conditions: The author of a paper has used a resource, and is willing to help documenting it as well as acknowledging its use; organisers of a conference have put a system in place (such as the LRE-map) to allow authors to document with basic metadata all LRs used in their own paper

Success End Conditions: The author(s) of a paper, when submitting to a conference, accept to spend some minutes filling up the LRE MAP, thus documenting and describing with small set of pre-defined metadata all languages used in their paper, whether they are own LRs or by others, finalised or in progress

Failed End Condition: LRE MAP not available for the given conference; authors refuse to fill up LRE MAP

Primary actor: Author of a paper Organizers of conference

Trigger: Paper submission

1.2.5.1.8. Use Case #07_ARIADNE: Data quality (Papatheodorou et al. 2013)

Goal: Users often complain about the lack of usefulness of data because data is structured in different way, or is incomplete, or lacks important details, or is not up to date

Scope: Digital repositories

Pre-conditions: Specification of the quality requirements for specific collections or data sets to be integrated in the e-infrastructure, so that the users regard the resulting services as valuable

Success End Conditions: The required datasets are available in an uncomplicated way

Failed End Condition: Datasets contain obsolete data and lack important details

Primary actor: User in the role of data consumer

Trigger: A user wants to find up to date information

1.2.5.1.9. Use Case #08_ARIADNE: Metadata quality (Papatheodorou et al. 2013)

Goal: Domain-based repositories can set high requirements for metadata, which the depositors will accept and follow, guided by archive curators, if necessary

Scope: Digital repositories

Pre-conditions: Specification of the quality requirements for specific collections or data sets to be integrated in the e-infrastructure, so that the users regard the resulting services as valuable

Success End Conditions: The available datasets are well described

Failed End Condition: The available datasets are missing important information

Primary actor: Data managers

Trigger: Metadata quality

1.2.5.1.10. Use Case #01_ACLweb: Metadata harmonization (McCrae et al. 2015)

Goal: When collecting metadata from multiple resources, there are two principal challenges: 1. property harmonization and 2. duplication detection

Scope: Harmonization is the challenge of verifying that there is not only structural and syntactic interoperability between the resources in that they use the same property, for example Dublin Core's language property, but also that they use the same value; we wish to detect duplicates that occur either due to the original representation or from multiple sources; it is clear that if a large number of records in fact describe the same resource then queries for that resource will return too many resources that may lead to errors for users

Pre-conditions: The application of NLP techniques allows to provide common metadata that will better enable users to find language resources for their specific applications

Success End Conditions: NLP enables data holders to provide cleaner federated data, resulting in better access and usability for research users

Failed End Condition: Data remains full of duplications and errors, making federated usage impossible

Primary actor: Data manager

Trigger: Complementary sets of data that would make a valuable federated resource are identified

1.2.5.1.11. Use Case #01_APARSEN: Exchange of data in reproducible form (Pampel et al. 2012)

Goal: To support quality assurance of data standards have to be developed in many disciplines which enable exchange of data in a reproducible form

Scope: Reviewing of data

Pre-conditions: Data is accessible in a re-usable form

Success End Conditions: Innovative publication strategies such as data publications are considered to be a positive contribution

Failed End Condition: Data is not available for reuse

Primary actor: Reviewer

Trigger: Assessment of data quality

1.2.5.1.12. Use Case #02_APARSEN: Development of incentive and reward system for quality assurance through scientists (Pampel et al. 2012)

Goal: The development of incentive and reward systems can help to increase recognition of quality assurance activities

Scope: Data quality

Pre-conditions: Quality assessment of research data

Success End Conditions: Increase of data quality

Failed End Condition: No possibility to increase data quality

Primary actor: Researcher in the role of data producer

Trigger: Assessment of data quality

1.2.5.1.13. Use Case #03_APARSEN: Establish discipline-specific services of data management (Pampel et al. 2012)

Goal: To support scientists in quality assurance of data it is necessary to establish discipline-specific services of data management, which are in line with scientific requirements

Scope: Data quality

Pre-conditions: Quality assessment of research data

Success End Conditions: Cooperation with publishers in developing data journals

Failed End Condition: Data management services not in line with scientific requirements

Primary actor: Data manager

Trigger: Assessment of data quality

1.2.5.1.14. Use Case #04_APARSEN: Quality assurance in the data creation process (Pampel et al. 2012)

Goal: To secure quality of data during data collection, scientists are required to apply methods and tools in a qualified and professional manner

Scope: Data quality

Pre-conditions: Availability of methods and tools for data quality assurance

Success End Conditions: Data quality assurance is performed in a qualified manner

Failed End Condition: Methods and tools not available for data quality assurance during the creation process

Primary actor: Researcher in the role of data producer

Trigger: Assessment of data quality during the creation process

1.2.5.1.15. Use Case #05_APARSEN: Development of certifications and audits (Pampel et al. 2012)

Goal: Certification and audit secure the quality of data repositories and affect the quality assurance of data

Scope: Data quality

Pre-conditions: Creation of reliable data repositories designed in accordance with disciplinary requirements

Success End Conditions: Producers of data benefit from opening it to a broad access

Failed End Condition: Data producers do not trust the repository

Primary actor: Data managers

Trigger: Contribute on data quality assurance

1.2.5.1.16. Use Case #06_APARSEN: Data management planning (Pampel et al. 2012)

Goal: Infrastructure facilities such as libraries and data centres can contribute to quality assurance of data via measures of research data management

Scope: Data management

Pre-conditions: Data must be provided in a reusable form

Success End Conditions: Data can be accessed in a reusable form

Failed End Condition: No possibility to reuse data

Primary actor: Data managers

Trigger: Contribute on data quality assurance

1.2.5.1.17. Use Case #07_APARSEN: Quality assessment of datasets (Pampel et al. 2012)

Goal: Publishers and journals can support quality assurance of data by demanding specific handling of data which form the basis of an article (e.g. within editorial policies)

Scope: Quality assurance of data

Pre-conditions: Existence of agreed method on quality assurance

Success End Conditions: Publishers and journals can contribute to quality assured publication of research data by operating data journals in cooperation with repositories

Failed End Condition: No agreements established between publishers and repositories

Primary actor: Publishers

Trigger: Contribute on data quality assurance

1.2.5.1.18. Use Case #02_SURFSHARE: Quality data checking and quality data enrichment (Feijen 2011)

Goal: It is relevant to underline, for both quality data checking and quality data enrichment, that the human intervention is an essential part of the checking process

Scope: Data checking

Pre-conditions: Data is store in digital archives and is accessible

Success End Conditions: When the data is transferred to another party, researchers wish to remain in control of their data

Failed End Condition: The researcher lose any possibility to remain in control of their data

Primary actor: Researcher

Trigger: Users want to check and enrich their data

1.2.5.1.19. Use Case #01_VLO: Metadata harvesting process (Van Uytvanck et al. 2012)

Goal: Compose a tailored metadata schema that relies on pre-canned components with explicit semantic declarations

Scope: Metadata harvesting

Pre-conditions: The challenge that comes with this approach is providing a uniform and easy to use interface to search in the resulting metadata records

Success End Conditions: Gather a large collection of varied metadata records and make them accessible using the CMDI infrastructure as the semantic backbone

Failed End Condition: The repository is not compliant with CMDI infrastructure

Primary actor: Researcher in the role of data consumer

Trigger: A researcher needs efficient ways to navigate to the language resources that really matter

1.2.5.1.20. Use Case #02_VLO: Direct access to language resource (Van Uytvanck et al. 2012)

Goal: Users want to have direct access to the language resources

Scope: Access to resources

Pre-conditions: This can be addressed by adding links to language resources

Success End Conditions: Access to resources

Failed End Condition: No links to resources are available

Primary actor: Researcher in the role of data consumer

Trigger: Users want to have direct access to the language resources

1.2.5.1.21. Use Case #03_VLO: Controlled vocabulary (Van Uytvanck et al. 2012)

Goal: Establish a controlled vocabulary service that has the potential to improve the quality of the metadata descriptions

Scope: Metadata quality

Pre-conditions: Availability of controlled vocabularies or resources to provide them

Success End Conditions: Data can be enriched through the use of a controlled vocabulary

Failed End Condition: Controlled vocabularies are not available for a certain research domain

Primary actor: Researcher in the role of data producer

Trigger: Improve metadata quality

1.3. Definition of Policy Requirements Concerning IPR, Open Data and Open Access

Main authors: Sara di Giorgio, with support from Antonio Davide Madonna and Marzia Piccininno (all MIBACT-ICCU)

1.3.1. Introduction – Overview of IPR

This section describes the requirements for the definition of policy requirements for IPR, Open Data and Open Access. Each of these topics is presented in separate sub-sections. Attention to IPR has increased in past years, as demonstrated by the laws that have been enacted at European and national levels. Even though IPR includes several themes such

as patents and trademarks, it is clear that the most relevant, for the research communities involved in the project, is the copyright of the data and related issues.

Technological advances giving rise to new methods of data collection and data creation amplify copyright and ethics issue which need to be addressed according to the requirements for researchers in relation to legal provision. Researchers are able through their work to make available large amounts of data that have copyright conditions that must be presented in a clear way, stating what can and cannot be done by human and machine agents.

However, cultural institutions and researchers can have some difficulties in establishing whether the data is freely accessible and reusable or subject to legal constraints. Identifying the correct copyright status of the resources is the primary need of the research communities. Sometimes, in fact, not all the necessary information is available for defining the copyright. It would be appropriate to have dedicated tools able to guide the collection manager in the choice of the licence to be taken. To proceed in this way, however, it is desirable to have a framework of licences that standardises and harmonises rights. The provision of a Licensing Framework, as already happens in other European projects, such as Europeana¹⁷, would bring clarity to a complex area, and make transparent the relationship between end users and the institutions that provide data.

Once the framework of licences and correct copyright status are identified, research communities have expressed the need to assign, in an automatic way, the licence to the data (or to the collections) they intend to make available for research purposes, making them searchable. To establish the correct licence to be assigned, if the resources made available are protected by copyright, then it is necessary to understand the level of information that can be made publicly available. Therefore, within the research infrastructure, it will be necessary to establish criteria to define the permitted reuse of resources regarding content and metadata.

1.3.2. Overview of the analysis of IPR, Open Data and Open Access requirements

The work done within ST2.1.3 on gathering requirements and identification of needs on IPR, Open Data and Open Access represent a strong starting point for WP3, even if it should be necessary for WP2 to undertake further analysis for some topics later in the project.

¹⁷ See: European Licensing Framework: <http://pro.europeana.eu/get-involved/europeana-ipr/the-licensing-framework>

IPR is, without a doubt, the topic on which the research communities worked most; for this reason it is possible to identify, in a timely manner, the commonalities.

In particular, the most relevant point concerns the identification of a framework of licences that can help the researcher to select the right copyright status and to avoid any legal issues as far as possible. Another hot topic is the developing of an AAI (Authentication and Authorization Infrastructure), guaranteeing controlled access to data by the users of the research infrastructure.

Regarding Open Access, research communities showed a common vision, considering this system as a powerful instrument for sharing research results. In this regard, it will be necessary to identify sustainable models model, not only from an economic point of view, but also one that ensures the best possible results in terms of quality and dissemination.

Open Data, in particular, deserves a special mention. On the one hand, research communities are in agreement on the need to develop a system for sharing their data freely according to defined standards; on the other hand, they have difficulties in overcoming problems that often arise, because data is commercially valuable or can be aggregated into works of value.

So, the documents produced by the research communities on Open Data are fewer than for the other topics discussed. As a consequence, if the requirements on IPR and Open Access can be said to represent a common vision, the requirements for Open Data should be investigated further, including also the PSI Directive and its application at national level. For this reason, ST2.1.3 intends, in the near future, to focus its attention on this topic, involving the project partners through surveys and interviews.

Last but not least, the absence of the history community deserves a mention: probably, this situation is due to the low degree of interest showed by this community in these topics. A need expressed by the research communities, in the IPR field, is the means to manage restricted access to protected resources by users. From this point of view, a better solution is represented by the AAI (Authentication and Authorization Infrastructure). Thanks to this system, for safeguarding privacy and data protection, it is possible to define different user levels and allow limited access to the resources that don't have a level of public dissemination.

The following table summarises the IPR requirements for each of the four communities identified by PARTHENOS. It shows the number of individual requirements from the various communities thus indicating the level of importance for each of the IPR requirements identified by PARTHENOS.

| IPR Requirement | History | Archaeology, heritage and applied disciplines | Social Science & Humanities | Language related studies |
|--|---------|---|-----------------------------|--------------------------|
| Definition of a framework of licences to adopt for data in Portal-Infrastructure's Community | | +++ | ++ | + |
| Creation of a tool to identify the copyright status for data or collections | | +++ | + | ++ |
| Creation of a tool to associate the identified copyright status with data within Portal-Infrastructure's community | | +++ | ++ | ++ |
| Creation of a AAI (Authentication and Authorization Infrastructure) | | +++ | ++ | |

1.3.3. Results – the IPR requirements

According to gathered requirements, it is possible to identify the following workflow: 1. identification of a common licences framework; 2. identification and association of the right license with data and metadata; 3. data publication within PARTHENOS portal; 4. definition of a specific process to ensure sensible data.

1.3.3.1.1. Use Case #ipr_01_Europeana_projects: IPR management¹⁸

Goal: Identification of IPR/copyright status for the resources (content, data, metadata)

Scope: Guideline

Pre-conditions: A content provider / researcher has a digital resource / collection

¹⁸ Requirements on IPR management are included in all the following documents: ATHENA (2013b), Tsolis (2013), Minerva Working Group (2008). Choukri et al. (2013), Fernie (2014).

Success End Conditions: A content provider / researcher identifies the right IPR/Copyright of the resources

Failed End Condition: A content provider is not able to identify the IPR/copyright status

Primary actor: Researcher / Institution who manages digital collections

Trigger: The content provider refers to Portal / Infrastructure's guideline

1.3.3.1.2. Use Case #ipr_02_Europeana_projects: IPR management¹⁹

Goal: Updated information on IPR legal framework across Europe

Scope: Guideline

Pre-conditions: A content provider / researcher has a digital resource / collection

Success End Conditions: A content provider / researcher consults national and EU legislation on IPR

Failed End Condition: A content provider is not able to identify the regulations for the copyright

Primary actor: Researcher / Institution who manages digital collections

Trigger: The content provider refers to Portal / Infrastructure's guideline

1.3.3.1.3. Use Case #ipr_03_Europeana_projects: Licensing framework (ATHENA 2013b)

Goal: Provide a licensing framework from a list of rights statements

Scope: Guideline

Pre-conditions: A content provider / researcher uploaded data in Portal-Infrastructure

Success End Conditions: A content provider / researcher associates the right licence to the data

Failed End Condition: The researcher is not able to associate the appropriate licence to the data

Primary actor: Researcher / Institution who manages data

Trigger: The content provider refers to Portal-Infrastructure's guideline with an exhaustive licensing framework

1.3.3.1.4. Use Case #ipr_0_Europeana_projects: licensing tool (ATHENA 2013b)

Goal: Assignment of the right licence for resources and, if it is necessary, limitation to re-use via online tool

Scope: System

¹⁹ Ibidem

Pre-conditions: The researcher knows some information about resources, but he isn't able to identify IPR/Copyright status

Success End Conditions: The researcher fills the online tool with the information required and assigns a copyright status to the resources

Failed End Condition: The researcher is not able to assign a copyright status because of a lack of information

Primary actor: Researcher

Trigger: Access to the online tool

1.3.3.1.5. Use Case #ipr_05_Europeana_projects: Creative Commons framework (ATHENA 2013b)

Goal: Adoption of Creative Commons Licence as a part of licensing framework

Scope: Guideline

Pre-conditions: A content provider / researcher uploaded data in a Portal-Infrastructure

Success End Conditions: The content provider selects a Creative Commons Licence for its data

Failed End Condition: The content provider is not able to identify the Creative Commons Licence for its contents

Primary actor: Content provider / researcher

Trigger: The content owner wants to share its data with the research community by using a Creative Commons Licence

1.3.3.1.6. Use Case #ipr_06_Europeana_projects: Images reuse (ATHENA 2013b)

Goal: Identification of image free for reuse

Scope: Guideline

Pre-conditions: The content provider / researcher manages images of digital objects in low resolution

Success End Conditions: The content provider / researcher shares the images of digital objects free for reuse

Failed End Condition: The content provider / researcher is not able to share its images because the size of the images

Primary actor: Content provider / researcher

Trigger: The content provider / researcher refers to the guideline for free reuse of images

1.3.3.1.7. Use Case #ipr_07_europeana_projects: Policy for orphan works and out-of-commerce (ATHENA 2013b)

Goal: Identification of the correct reuse policies for orphan works and out-of-commerce resources

Scope: Guideline

Pre-conditions: The content provider manages orphan works and out-of-commerce resources

Success End Conditions: The content provider shares under public domain orphan works and out-of-commerce resources

Failed End Condition: The content provider doesn't share orphan works and out-of-commerce resources

Primary actor: Content provider / researcher

Trigger: The content provider / researcher refers to Portal-Infrastructure's guideline to know the policies about the orphan works and out-of-commerce

1.3.3.1.8. Use Case #ipr_08_Europeana_projects: Level publishing of Metadata (ATHENA 2013b)

Goal: Identification of different levels to publish metadata: 1. open 2. restricted access

Scope: System

Pre-conditions: The content provider / researcher ingested metadata with protected information

Success End Conditions: The content provider / researcher chooses to publish a minimal, intermediate or full set of metadata

Failed End Condition: The content provider can't choose a publication level for its metadata

Primary actor: Content provider / researcher

Trigger: The metadata uploaded contains sensitive information that the content provider doesn't want to share at a public level

1.3.3.1.9. Use Case #ipr_09_europeana_projects: User-Generated contents (ATHENA 2013b)

Goal: Provide clear terms of use, to which users must consent, before they create content on the site

Scope: Guideline

Pre-conditions: The contents are available online within Portal-Infrastructure's community

Success End Conditions: The user reads and accepts the terms of use

Failed End Condition: The user doesn't accept the terms of use

Primary actor: Users

Trigger: A user creates contents within Portal-Infrastructure's community

1.3.3.1.10. Use Case #ipr_10_europeana_projects: Reuse of donor item (ATHENA 2013b)

Goal: Identification of restrictions for donor items

Scope: Guideline

Pre-conditions: A content provider / researcher receives a donor resource

Success End Conditions: The item has no restriction and can be published

Failed End Condition: The item has restriction and can't be reused

Primary actor: Researcher / Institution who manages item

Trigger: A Content provider / researcher wants to publish online a donor item and refers to guideline

1.3.3.1.11. Use Case #ipr_11_Publish_METASHARE: An annotated corpus (Choukri et al. 2011)

Goal: Create an annotated version of a textual corpus, containing some additional layers of linguistic information, and make it publicly available in a standardised format

Scope: Guideline

Pre-conditions: The initial corpus must be free of all restrictions, either available in public domain or available under such a licence (e.g. Creative Commons) that allows reuse of the data and publication of derivatives; moreover, the copyright owners have to be identified for attribution; finally, the absence of sensitive data has to be verified

Success End Conditions: The item has no restriction and can be published

Failed End Condition: The item has restriction and can't be annotated and republished

Primary actor: Researcher/ Institution

Trigger: Researcher/ Institution who wants to make corpus available and refers to guideline

1.3.3.1.12. Use Case #ipr_12_DRI: System Licences association (Webb & McGowan 2015)

Goal: The system must map copyright statements and reuse licences to digital objects

Scope: System

Pre-conditions: The provider of the content gives clear statements of copyright and licensing

Success End Conditions: Researcher / user is able to determine which items are reusable

Failed End Condition: Researcher can't determine licensing and either ignores the content, or carries on and uses anyway, potentially breaking licensing and copyright restrictions

Primary actor: Content providers

Trigger: Content provider wants to ensure users know what they can and can't reuse, allocating licences as appropriate

1.3.3.1.13. Use Case #ipr_13_DRI: System Licences association (Webb & McGoo-han 2015)

Goal: The system must enable a user to edit a collection in accordance with their access rights

Scope: System

Pre-conditions: Provider of content is willing for content to be edited or used, and has clearly written access statements

Success End Conditions: Researcher is able to download and reuse digital content for analysis

Failed End Condition: Researcher is unable to use digital content

Primary actor: Content provider

Trigger: Content provider wants to enable researchers to reuse their content as defined by the licence

1.3.3.1.14. Use Case #ipr _14_DRI: Information and updating of current legislation (Webb & McGoo-han 2015)

Goal: The system must adhere to current legislation on IPR

Scope: System

Pre-conditions: Provider is knowledgeable of current legislation, and actively checking for updates

Success End Conditions: Researchers are confident that they are using content legally

Failed End Condition: Provider lapses in legal obligations and either doesn't provide content that researchers have the right to, or provides content with incorrect licences, making data void

Primary actor: Content providers

Trigger: Content providers wants to make sure legal obligations are covered

1.3.3.1.15. Use Case #ipr _15_DRI: Indicate clearly which items are reusable (Webb & McGoohan 2015)

Goal: The system must map copyright statements and reuse licence to digital objects

Scope: System

Pre-conditions: The provider of the content provides clear statements of copyright and licensing

Success End Conditions: Researcher / user is able to determine which items are reusable

Failed End Condition: Researcher can't determine licensing and either ignores the content, or carries on and uses anyway, potentially breaking licensing and copyright restrictions

Primary actor: Content providers

Trigger: Content provider wants to ensure users know what they can and can't reuse

1.3.3.1.16. Use Case #ipr _16_DASISH: Identification of ethical aspects and legal requirements in the SSH domains (Schmidutz et al. 2013)

Goal: A researcher wants to collect data and linking research data with data from external source

Scope: Guideline

Pre-conditions: SSH sensitive data generated in the process of survey production and new data source e.g. internet and social media

Success End Conditions: Data protection is strengthened

Failed End Condition: Lack of information, knowledge to consent the measures to safeguard privacy

Primary actor: Researcher who collects, curates and disseminates new data type

Trigger: The user finds Portal-Infrastructure's guideline with a legal and ethical framework

1.3.3.1.17. Use Case #ipr_17_DASISH: User authentication and authorization (Schmidutz et al. 2013)

Goal: Creation of an Authentication and Authorization Infrastructure (AAI) to enforcing the user agreements

Scope: System

Pre-conditions: The Portal-Infrastructure's provides services devoted to registered users

Success End Conditions: The federated user is able to access the PARTHENOS Portal

Failed End Condition: The user is not federated to access the Portal-Infrastructure's Portal

Primary actor: Users

Trigger: A federated user wants to access to a reserved service

1.3.3.1.18. Use Case #ipr_18_DASISH: User authentication and authorization policy (Schmidutz et al. 2013)

Goal: Creation of different types of users to ensure that only users with right credentials get access to proper copyrighted or restricted resources

Scope: System

Pre-conditions: The users is authorized by Portal-Infrastructure

Success End Conditions: The users accepts the dedicated service

Failed End Condition: The user has not the permission to access the requested services

Primary actor: Users

Trigger: A user want to access restricted resources and services

1.3.4. Overview of Open Data

Open Data is one of the thorniest issues in the field of resource reuse. Open Data, in fact, is a new discipline that cannot claim a shared vision yet, as it is tied to the web and large amounts of data have only become available over the last few years. Moreover, in many cases there is a risk of confusing the free access to data with Open Data, which, in addition to being freely available, is free from any kind of restriction and can also be edited and reused for commercial purposes²⁰. It is also important to emphasise that Open Data, Linked Data (LD) and Linked Open Data (LOD) are not the same. For LD and LOD, in fact, the focus is on connection and methods rather than on IPR issues²¹. Several institutions, especially public ones, still provide data under copyright even if they should make available according to the principle "open by default". For this reason, the collected requirements are not enough to cover the complexity of this topic, so it will be necessary to proceed with a dedicated survey to fill this gap. Thus, the main need of the research communities is in

²⁰ Open definition by the Open Knowledge International Initiative 2016 <http://opendefinition.org/od/2.1/en/>

²¹ The goal of the W3C SWEOL Linking Open Data community project is to extend the Web with a data commons by publishing various open data sets as RDF on the Web and by setting RDF links between data items from different data sources. See more at <https://www.w3.org/wiki/SweolG/TaskForces/CommunityProjects/LinkingOpenData>

regard to the definition of a set of rules which defines the level of information needed to be shared in order to meet open data criteria. Another need is the ability to search for data according to their status of open data in an in-system search engine, allowing the user to know immediately the data status and the reuse possibility.

| Open Data Requirement | History | Archaeology, heritage and applied disciplines | Social Science & Humanities | Language related studies |
|---|---------|---|-----------------------------|--------------------------|
| Definition of a minimum set of data to share under a CC0 or CC-BY licence | | +++ | ++ | |
| Guarantee the searchability of data | | ++ | ++ | |

1.3.5. Results – the Open Data requirements

The most relevant requirements for Open Data are the definition of what can be shared and the way to harvest this information in an easy way.

1.3.5.1.1. Use Case #od_01_DASISH: Access and data re-use (Bøe et al. 2014)

Goal: Define access and re-use restrictions of these data

Scope: 1. Access to repositories; 2. access to data

Pre-conditions: Documentation about licences, such as Creative Commons, Open Data Commons, and GNU General Public Licence

Success End Conditions: Attach a licence to the data providers' data which defines access and reuse restrictions of these data

Failed End Condition: Repositories, projects or sectors have prepared their own licence agreements

Primary actor: Data providers

Trigger: Data providers are encouraged or required to attach a licence to their data

1.3.5.1.2. Use Case #od_02_Europeana_cloud: Sharing research data in open and trustful way (Zeinstra et al. 2013)

Goal: Establishing a set of minimum metadata fields to publish under CC0 (or CCBY)

Scope: Guideline

Pre-conditions: The researcher/institution manages the IPR of the data

Success End Conditions: The researcher/institution can publish and disseminate open data

Failed End Condition: Data is protected by copyright

Primary actor: Researcher / institutions with a role of data manager

Trigger: The user finds Portal-Infrastructure's guideline which provides a frame-work that regulates the participation in Portal-Infrastructure by providing a service level agreement between Portal-Infrastructure and data providers for disseminating open data

1.3.5.1.3. Use Case #od_03_Europeana_cloud: Persistent identification of datasets ingested (Zeinstra et al. 2013)

Goal: Allocation of persistent identification of datasets ingested in Portal-Infrastructure; the system used should be capable of identifying subsets

Scope: System

Pre-conditions: Depositing research data into a data repository

Success End Conditions: The resources has been identified by a DOI or equivalent

Failed End Condition: The system doesn't assign a persistent identifier

Primary actor: Researcher with a role of data manager

Trigger: Data uploading in Portal-Infrastructure's repository

1.3.5.1.4. Use Case #od_04_Europeana_cloud: Common method of data citation (Zeinstra et al. 2013)

Goal: Encouraging researchers to share access to their datasets

Scope: System

Pre-conditions: Depositing research data into a data repository or archive with a unique identifier

Success End Conditions: Citation is permanently associated with data

Failed End Condition: The system doesn't allow a permanent association

Primary actor: Researcher with a role of data manager

Trigger: Data uploading in Portal-Infrastructure's repository

1.3.5.1.5. Use Case #od_05_Europeana_cloud: API Service (Zeinstra et al. 2013)

Goal: The system must support user client development through a REST based API

Scope: System

Pre-conditions: Data is suitable for API use

Success End Conditions: The system has a useable API that can be used to retrieve large amounts of data from the repository

Failed End Condition: The API is not suitable for use with the data

Primary actor: Researcher with a role of data manager

Trigger: Content Provider wants data to be made available in large quantities

1.3.5.1.6. Use Case #od_06_Europeana_cloud: List of digital objects (Zeinstra et al. 2013)

Goal: The system shall retrieve a list of digital objects based on the search criteria/query and the user and the digital objects access policies

Scope: System

Pre-conditions: Metadata for items is such that they can be accurately retrieved from the insystem search engine (as well as external search engines)

Success End Conditions: Users are able to obtain items that are relevant to their search

Failed End Condition: Users are given too many items that are not relevant to their search (or not enough data)

Primary actor: Cataloguer; content provider

Trigger: Content provider wants data to be searchable

1.3.6. Overview of the analysis of Open Access requirements

Open access has received a significant boost in recent years. The possibility of publishing academic articles online has made easier the dissemination of research results that were previously the prerogative of a small group of people. This new approach has led to a substantial change in the publishing process, giving everyone the chance to disseminate their work. For this reason, the quality of the publication has become one of the hot topics related to open access. Research communities, however, have worked hard to solve this problem with the definition of best practice and the introduction of an in-depth level of checking by reviewers before the publication of an article / journal. Regarding the definition of a methodological approach for ensuring the quality of publications, the most relevant need of research communities is for a sustainability model for open access publications. In this way they can not only ensure the sustainability of online publication over the long term, but can also provide a boost for traditional publication processes, with the adoption of specific models. Another requirement is for the creation of restricted and/or controlled access with-

in the open access model, to protect, for example, data involving human subjects or data that is copyrighted.

| Open Access Requirement | History | Archaeology, heritage and applied disciplines | Social Science & Humanities | Language related studies |
|--|---------|---|-----------------------------|--------------------------|
| Definition of best practice for reviewing an academic article before publication | | ++ | ++ | |
| Identification of different sustainability models | | ++ | +++ | |
| If there is also a traditional publication process, define the right strategy for ensuring the optimal dissemination of the open access publication. | | ++ | +++ | |
| Creation of controlled access to protected data | | ++ | ++ | |
| Common method of data citation | | +++ | ++ | |

1.3.7. Results – Open Access

The requirements for Open Access focused on the best way to review articles before publication, matching data from different repositories, and providing a sustainable model to ensure the wide dissemination of publications.

1.3.7.1.1. Use Case oa_01_DRI: Stable access to the Repository (Webb & McGoohan 2015)

Goal: The repository shall provide 'reliable, long-term access to managed digital resources to its designated community, now and in the future'

Scope: System

Pre-conditions: Sustainable long-term funding is assured

Success End Conditions: The repository is able to provide access to digital content into the long-term future

Failed End Condition: The repository ceases to exist after initial funding ends, resulting in a loss of data access, and also potentially of the data itself

Primary actor: System

Trigger: The mission of the Repository is sustainability

1.3.7.1.2. Use Case oa_02_DRI: Shared formats (Webb & McGoochan 2015)

Goal: The system shall provide a suite of research tools for data that share formats and conventions

Scope: System

Pre-conditions: Formats and conventions have been determined, and the data already matches them

Success End Conditions: Users are able to combine data they already have with data obtained from the repository in order to use repository-hosted tools thanks to shared formats

Failed End Condition: Users have to spend a significant amount of time standardising data they obtain from the repository, some of which might not be in any way usable with data they already have

Primary actor: Content provider

Trigger: The researchers wants to be able to combine their data in the repository with other data

1.3.7.1.3. Use Case oa_03_DRI: Interaction with other EU research infrastructures (Webb & McGoochan 2015)

Goal: The system shall interface with similar developing EU research infrastructures

Scope: System

Pre-conditions: The system already has data formats that conform to a common standard

Success End Conditions: The implemented system is able to interact with other EU research infrastructure platforms

Failed End Condition: The system is unable to interact with other infrastructures, and the data doesn't get used

Primary actor: Developer

Trigger: Multiple platforms make data access difficult; easy interface access allows for use and reuse of data with other sources

1.3.7.1.4. Use Case oa_04_DRI: Secure AAI (Webb & McGoohan 2015)

Goal: The system shall manage access to digital objects through authentication and authorization mechanisms

Scope: System

Pre-conditions: The system is already secure at every level to enable people to set up passwords

Success End Conditions: Users access the data via a secure authentication process that they know they can trust; content providers are also happy for their data to be available via the system

Failed End Condition: The system is not secure and therefore content providers don't provide data for users.

Primary actor: Developer

Trigger: Content providers are reluctant to discuss ingestion of data unless they know the system will be secure

1.3.7.1.5. Use Case oa_05_DRI: Downloading data (Webb & McGoohan 2015)

Goal: The system shall allow users to download files to their local drive in accordance with their access rights and the object's access rights

Scope: System

Pre-conditions: The system allows for downloads and the content providers also allow for download of data

Success End Conditions: Users are able to download materials for reuse if the data content rights access allows

Failed End Condition: Users can download anything to their local drives regardless of access rights; users are unable to download anything and become frustrated with the system

Primary actor: Developer

Trigger: Users want to be able to download content locally for use

1.3.7.1.6. Use Case oa_06_DRI: Edit digital objects (Webb & McGoohan 2015)

Goal: The system shall enable a user to edit digital objects in a collection in accordance with their access rights

Scope: System

Pre-conditions: User knows how to make edits to collection within the platform / Access Rights are clearly stated

Success End Conditions: User is able to make edits within a collection, knowing that they are within their access rights

Failed End Condition: Users are unable to edit collection items, thus not enabling them to get the information they need

Primary actor: Developer

Trigger: Users want to be able to edit items within collections

1.3.7.1.7. Use Case #oa_07_ESFRI: Provision of access across data repositories (ESFRI 2008)

Goal: Provide seamless access to data across repositories, nations and research purposes

Scope: System Data repositories, repository management, repository hosts, Data Access

Pre-conditions: 1. Metadata standardization; 2. Interoperability; 3. Data harmonization; 4. Central data access (Access to data collection)

Success End Conditions: Seamless access to data across repositories

Failed End Condition: Access to data repositories is denied or only partial access is granted

Primary actor: Public/Private institutions hosting repositories

Trigger: Need for universal access to large amounts of data

1.3.7.1.8. Use Case #oa_08_OpenAIRE: International search options (Hogenaar et al. 2011)

Goal: The ability to search for data across current political boundaries

Scope: Researchers want insight in the available data across political boundaries

Pre-conditions: Access for researchers to metadata from different countries

Success End Conditions: A search portal/ facility for researchers where data from across current political boundaries can be found

Failed End Condition: Researchers are not able to search data across current political boundaries

Primary actor: Provider of search facility

Trigger: A researcher wants to have knowledge of research data across current political boundaries through a search facility

1.3.7.1.9. Use Case #oa_09_OpenAIRE: Clear embargo regulations (Hogenaar et al. 2011)

Goal: Clearly defined embargo regulations to help stimulate open access publication of research data

Scope: Researchers wish to have clear regulations on data sharing and therefore the use of embargo

Pre-conditions: Clear knowledge of regulations regarding sharing according to (international) law; clear decisions on data sharing by data owners

Success End Conditions: Researchers are clearly informed of the options in regards to sharing of data and therefore more data could be published open access

Failed End Condition: Researchers are not sure of the possibilities regarding data sharing and therefore keep their data (longer) under restricted access

Primary actor: Data owner

Trigger: A researcher wants to have clear information on embargo regulations for research data

1.3.7.1.10. Use Case #oa_10_OpenAIRE: European regulation on personal data (Hogenaar et al. 2011)

Goal: A European regulation that allows handling of personal data for research purposes

Scope: Researchers want to use personal data for research but the regulation on using this type of data, now included in differing privacy laws, is not clear across Europe

Pre-conditions: There should be consensus in Europe on creating a European regulation regarding using personal data for research

Success End Conditions: There is a European regulation regarding personal data that makes it clear for researchers what and what cannot be done with personal data

Failed End Condition: Researchers only do research in a context where the regulation is clear for them

Primary actor: Europe

Trigger: A researcher wants a clear view on how research can be done on personal data in an international, European context

1.3.7.1.11. Use Case #oa_11_OpenAIRE: No copyright to facilitate free and open use of research material (Hogenaar et al. 2011)

Goal: Copyright free research data to make using research data possible and easier

Scope: Researchers are currently often limited or temporarily limited to the use of data due to the copyright restrictions

Pre-conditions: Current data and new data have to be made copyright free, juridical procedures need to have been executed; this could be done by applying a general exemption for research purposes

Success End Conditions: Researcher can access research data where no copyright applies

Failed End Condition: Researchers still have problems using material with copyright, or cannot use it entirely

Primary actor: Data owner

Trigger: A researcher wants to be able to use research data without copyright restrictions

1.3.7.1.12. Use Case #oa_12_OpenAIRE: Enable enrichment of data (Hogenaar et al. 2011)

Goal: Openness of data to enable enrichment by the research community

Scope: By letting the research community work on research data of others the quality can be substantially improved

Pre-conditions: The research data should be accessible by the research community

Success End Conditions: Research data can be enriched by the research community

Failed End Condition: Research data is not enriched by the research community and therefore might not be of the highest quality

Primary actor: Data owner

Trigger: A researcher wants to collaborate on the quality of data from a research project

1.3.7.1.13. Use Case #oa_13_OpenAIRE: Enhanced publications each using the same level of accessibility for all its components (Hogenaar et al. 2011)

Goal: Having researchers create enhanced publications and let the same level of accessibility, as open as possible, apply for all its components

Scope: Letting the same access level apply to all components, as open as possible, makes the use of an enhanced publication successful

Pre-conditions: The data should be stored and be accessible according to the chosen access level; the researcher should be able to create an enhanced publication

Success End Conditions: Researchers create enhanced publications and apply the same access level to its resources

Failed End Condition: Researchers may not be able to locate specific resources of a publication. Researchers cannot use the data because they cannot access all components

Primary actor: Researcher

Trigger: A researcher wants to be informed about a research project and at the same time about the location of the underlying resources; additionally a researcher wants to be able to make use of all components

1.3.7.1.14. Use Case #oa_14_OpenAIRE: Sustainable access (Hogenaar et al. 2011)

Goal: Sustainable access to objects in repositories

Scope: Sustainable access means objects can always be found via a link, regardless of its place on the web; this is relevant for citation, finding data and therefore in particular for enhanced publications

Pre-conditions: 1. Repositories implement a system through which each object receives a persistent identifier (PID) and can disseminate this; 2. objects in repositories are taken over by other repositories if the first can no longer execute their tasks; 3. there is always a resolver in place that facilitates creating a correct URL from a PID

Success End Conditions: Objects in repositories are sustainably accessible

Failed End Condition: Not all, or none, of the objects in repositories can be found via a link that a researcher used to point others to his/her resources

Primary actor: Repository

Trigger: A researcher wants to refer to his/her resources in a sustainable way

1.3.7.1.15. Use Case #oa_15_OpenAIRE: International search options (Hogenaar et al. 2011)

Goal: The ability to search for data across current political boundaries

Scope: Researchers want insight in the available data across political boundaries

Pre-conditions: Access for researchers to metadata from different countries

Success End Conditions: A search portal/ facility for researchers where data from across current political boundaries can be found

Failed End Condition: Researchers are not able to search data across current political boundaries

Primary actor: Provider of search facility

Trigger: A researcher wants to have knowledge of research data across current political boundaries through a search facility

1.3.7.1.16. Use Case #oa_16_ARIADNE: Data citation (Fernie 2014)

Goal: Providing a reference to data

Scope: System

Pre-conditions: The researcher first need to publish the data, or at very least, a description of the data

Success End Conditions: Establish a common method of data citation for adoption by partners as academic recognition is an important motivation for encouraging researchers to share access to their datasets

Failed End Condition: Data or highly sensitive data that is not available for ethical issues; draft outputs, or a highly confidential report

Primary actor: Researcher assign a persistent identifier (DOI) to the resource

Trigger: Storing the data to enable a stable access; assigning a DOI to the data; providing appropriate metadata to describe the data including citation information; publishing the metadata with a persistent identifier (DOI)

1.3.7.1.17. Use Case #oa_17_ARIADNE: Persistent identification (Fernie 2014)

Goal: Provide persistent identification

Scope: System

Pre-conditions: The system used should be capable to identify sub-set within collections

Success End Conditions: The researcher allocates a persistent identification of datasets ingested in the Portal-infrastructure

Failed End Condition: The system doesn't allow to assign a persistent identifier or doesn't persistently store the resource

Primary actor: Researcher assign a persistent identifier (DOI) to the resource

Trigger: Storing the data to enable a stable access; assigning a persistent identifier to the data

1.3.7.1.18. Use Case #oa_18_ARIADNE: Publishing licence (Fernie 2014)

Goal: CC-BY is recommended for open access

Scope: Guideline

Pre-conditions: Authors grant any third party the right to use the article freely as long as its original authors, citation details and publisher are identified

Success End Conditions: Publish open access articles under the terms of the Creative Commons Licence CC-BY which permits use, distribution and reproduction in any medium, provided the original work is properly cited

Failed End Condition: Authors don't allow third party a free reuse of the article

Primary actor: The article is published with a CC-BY licence

Trigger: Authors submit an article to an Open Access Journal

1.3.7.1.19. Use Case #oa_19_ARIADNE: Access (Fernie 2014)

Goal: A user wants to have access to resources.

Scope: Between the conditions of access to data, a type of access (controlled or open) has been selected

Pre-conditions: 1. Open access: the data is free for all to read and copy, but there may be conditions such as attribution of the data to its creator; 2. controlled access: access rights might depend on user identity and location

Success End Conditions: The user manages to have access to resources. If the access is controlled, the user has the rights on user identity and location

Failed End Condition: The user can't have access to the resources

Primary actor: User

Trigger: A user needs to access to resources

1.3.8. Narrative Use Case

The following use case was not transformed according to the Cockburn model, because it can be considered a best practice for open access and it could be a starting point for other WPs involved in the project.

A publisher wants to release a printed journal according to the principles of OA to encourage the dissemination of research material.

1.3.8.1. User Story

Nordic Wittgenstein Review

The journal is a specialized international journal, publishing texts in English. It is peer-reviewed (double-blind), and published by the Nordic Wittgenstein Society.

The journal used to be annual, but since 2014 it has been bi-annual.

The journal has a Nordic editorial board, appointed by the Nordic Wittgenstein Society board, and an international advisory board. Its sections include an Invited Paper, Submitted Articles, From the Archives, Interview, and Book Reviews. The theme of the journal is philosophy and other Ludwig Wittgenstein-related research.

For copyright, the online versions use a Creative Commons licence, CC-BY-NC-SA, Non-Commercial, ShareAlike, which allows the users online to share and adapt but not sell the

content forward; if adapted, it needs to be distributed under the same licence as the original.

Print publication

The journal was published in print and circulated via Ontos Verlag (small publisher, Issue 1/August 5, 2012) and De Gruyter (large publisher, Issue 2/August 28, 2013), following the purchase of Ontos by De Gruyter in May 2013.

It was sold as individual hard copies and print subscriptions to institutions and individuals by Ontos, later by DeGruyter, and Issue 2 also as electronic subscriptions in a bundle.

Online OA publication

NWR#1: The article PDFs of the first issue were published three months after print (Nov. 5, 2012) on the journal site (using the publishing platform OJS) www.nordicwittgensteinreview.com. Later, full text HTMLs were added.

NWR#2: The second issue was published in print Aug. 28, 2013 and for electronic subscription Aug. 20, 2013. Half of the article PDFs were made OA immediately upon print in the journal platform, and the rest three months later, Nov. 28, 2013. (The articles were also available OA on De Gruyter's site due to a mistake from October 2013.)

Full text HTML versions were added to the journal platform Dec. 18, 2013. The journal access and sales data were monitored during this time.

The journal also took part in an Open Review experiment, in which double-blind peer review was supplemented with a session of Open Review or Preview online of the submitted articles accepted for publication during one month (NWR #1 during April-May 2012, NWR#2 May 2013). During this time, many downloads of the preprints were recorded (after one month during the first preview, the PDFs had been downloaded on average 98 times each, ranging from 38 to 167 downloads per article, and during the second preview, half of the articles were on Open Review and these were downloaded on average 153 times).

The journal charged no publication fees or other author processing charges. The printed journal was offered on a subscription basis.

1.3.8.2. Goal

Develop sustainable Open Access (OA) business models

1.3.8.3. Scope

Promote OA academic publishing

1.3.8.4. Preconditions

1. **Funding:** In new publishing models, additional funding should be seen as “a natural necessity”.
2. **Awareness:** Publishers should inform their authors properly about Open Access and its benefits.
3. **Use of aggregators:** The online interface will require some input from the publisher. It is recommended to use an aggregator, which also takes care of proper indexing, such as the publication platform OAPEN Library (<http://www.oapen.org>).
4. **Interoperability:** The OA content should be easily interoperable with other repositories.
5. **Indexing:** Proper indexing is a way for a publisher to take care of the extra publicity or visibility that OA brings with it. Also, it is a way to ensure that the authors get the dissemination advantage that they expect.
6. **Distinction between customer segments:** Small publishers sell mostly to libraries. Downloaders are often individuals and they make up a different and wider customer segment, composed of end users. Publishers should note that the needs of these two segments differ and design products differently for them.
7. **Visibility:** Hybrid OA journals of all flavours need to become more visible and delayed OA needs to be branded or marketed as a viable option.

1.3.8.5. Success end condition

- **Strong dissemination advantage of OA publication:** a) Wider dissemination is the key to attracting quality authors. b) Interest in the material rises significantly after OA publication. c) The dissemination advantage for older open access printed publications is highly significant.
- **Sales advantage:** a) OA carries a very low risk of diminishing sales for print books. b) OA increases sales for old open access printed publications.
- **Registration:** this requirement is an advantage for knowing the customer segments and hence for marketing purposes.

1.3.8.6. Fail end condition

- Lack of awareness as an issue for not publishing OA: There is a tension between the researcher’s OA ideals and their own publication and community service (review) practices: although the researchers wish for more OA, they are not always aware of their own possibilities for publishing OA.

- The hybrid trap: the exclusion mechanism, generated by the restrictive conditions on OA formulated by the Budapest Open Access Initiative (BOAI). This prevents new hybrids from prospering and potentially sustainable models from being developed. Publishers should strive to avoid it.
- In the humanities and social sciences, it is often the case that hardly any funding is available for departments.
- Potential decrease in sales
- Loss of prestige
- High costs may arise from the configuration and maintenance of an Internet platform in which articles are published and kept, due to the layout of the material, the working hours required to coordinate the peer review process, the dissemination activities, marketing material etc.

1.3.8.7. Primary Actor

- Academic Book/ Journal publishers

1.3.8.8. Other Actor(s)

- Authors
- Readers
- Institutions (universities, libraries, research institutions, funding agencies)

1.3.8.9. Trigger

The results of research should be openly available online for everyone, free of charge. In this way, researchers and even the general public in different parts of the world will be equally situated with regard to access to research material.

1.3.8.10. Main Scenario

- The Journal is checked by peer-reviewers
- The journal is published in print
- PDF publication issue three months after print
- Later, full text publication after one year

2. Use Cases and Requirements on Standardization

Main authors: Petra Links and Annelies van Nispen (both KNAW / NIOD), Karolien Verbrugge (KNAW / NIOD, responsible for the final edition)

2.0. Introduction

For the PARTHENOS project, standardization is key to data sharing and reuse²² and the project has a mission in making a difference to those researchers who are not yet familiar with using standards. This section deals with the requirements of standardization expressed by the research communities involved in the project. Section 2.1 presents use cases of:

- researchers who do not use standards yet (or are at an early stage)
- researchers who have difficulty with implementing standards

The use cases are based on prior work done by infrastructural projects and the partners' institutions involved in WP2 Task 2: AA, CLARIN, CNR, CNRS, CSIC, FHP, FORTH, INRIA, KNAW-DANS, KNAW-NIOD, MIBACT-ICCU, OEAW, SISMELE and TCD.

The use cases presented in the next section serve the mission of PARTHENOS at large, and WP4 in particular. WP4 shall process the use cases as learning experiences and sand boxes. This supports to the claim that standards contribute to providing access and preserving data through time and space. The processed use cases will be incorporated in the deliverable of WP4.

Relevant reports, articles and deliverables have been collected in preparation of the use cases. In total twenty three documents have been gathered from eight projects or research infrastructures by the partners: ATHENA, CLARIN, DASISH, EHRI, Flarenet, IPE-RION CH, CENDARI and Meta-share.

2.1. Use cases

The format used to structure the content in the use cases was taken from the Cockburn simplified description form as presented at the PARTHENOS webinar 'How to write use cases' in September 2015.²³

²² *Grant Agreement - Number 654119 - PARTHENOS*, p. 134.

²³ PARTHENOS webinar: '[How to write use cases](#)', slides by Edi Marchetti (September 2015). See also: Cockburn (2000).

Use case elements taken from this method are: user story; goal; scope; preconditions; success end condition; primary actor; other actor(s); trigger; main scenario; extensions; variations.

The mandatory elements are: user story; goal; main scenario. The other elements were recommended.

Preconditions for every use case are expertise, time and budget. Because they are required for all research they are not mentioned in the separate use cases.

The initial use cases were written in October 2015. These were shared with WP4. In cooperation with WP4, it was decided to put extra effort in revising the use cases by adding information about the data creation process (PARTHENOS Vision) and the structure and nature of research data. The research phases of the PARTHENOS Vision were added in WP4. We also explicitly mention the use and purpose of standards for the scholarly process, as well as the primary actors and the goal of the use case.

The revised use cases were presented and discussed in the WP4 Standards Workshop held on the 3-5 December 2015. The last revisions were made on outlining the goal of standards in the use cases. Where standards were implicit, they were made explicit.

2.1.1. History

2.1.1.1. WW1 Historian and the trans-national/trans-institutional question of the development of the railways

Provided by: Trinity College Dublin / Cendari

Contributor(s): Jennifer Edmond / Francesca Morselli / Vicky Garnett

User Story

The researcher is a WW1 Historian, working on transport infrastructure history. She needs to analyse the alteration as well as the new construction of railway tracks in East Central Europe (starting with Lithuania and Poland) at the end of World War I.

Goal

Search in (or into) a unified environment across library, archive and textual data as well as upload data from other digital archives to that environment.

Scope

Find patterns of relevance to a transnational research questions from across a number of institutional collections of varying sorts.

Preconditions

- Data needs to follow a common standard - OR - tools need to exist to allow federation of 'like with like'
- Content Holding Institutions (CoHIs) need to provide machine readable access to their data - OR - an environment bringing these data together must be known to and usable by the researcher
- Researcher must have appropriate programming skills and familiarity with data formats such as RDF or XML - OR - tools to access and federate data must exist
- Institutional data must be as rich and complete as possible
- Researcher and her tools must be able to encompass all relevant languages of relevant material

Success End Condition

Researcher is able to access comparable data across countries and institutions and both search or upload the data herself to discover patterns in a 'like-for-like' manner

Fail End Protection

Not sure there is one: researcher won't be able to do this research (at least not efficiently and without a lot of travel budget)

Primary Actor

Modern Historical Researcher

Other Actor(s)

- CoHI collections and technical management
- Intermediaries (tool designers, infrastructure developers)
- Depends on the solution
- Other researchers with access to the Research Infrastructure

Trigger

Historical researcher requires comparable collections in multiple types of CoHI

Main Scenario

A WW1 Historian needs to analyse the alteration as well as the construction of new railway tracks in East Central Europe (starting with Lithuania and Poland) at the end of World War I. What she needs to have at the very beginning are maps of railway lines before the outbreak of World War I, maps of the construction of new railroad tracks under German occu-

pation, and maps of railroad construction plans of Poland and Lithuania after their respective declarations of independence and the setup of traffic/infrastructure ministries. After finding these maps on external Digital Humanities websites or directly within the system she is using, the challenge is to bring them in agreement regarding their scale so that she can create a map of track modifications and new constructions in the whole region.

She has reasonable programming skills and is familiar with common data formats such as RDF or XML. She also recognizes that many of her data sources will use different archival and library standards to structure their metadata: the large archives generally use EAD/ISAD(G), the large libraries MARC 21/MADS. But some smaller and some of the private (e.g. industrial) institutions, especially in rural Lithuania, may use Dublin Core only or, indeed, a custom standard. Some data may also be available in scholarly projects published online, which in a best case scenario may include full documents marked up in TEI; in a worst case scenario they may be only minimally described PDFs or other unstructured data (especially some of the maps). She also has experience of using tools to access, clean and federate data such as OpenRefine, MINT and some GIS tools. She plans to use these in order to rationalize the data to compare 'like with like' and therefore identify any patterns emerging in the data.

One of the key further elements in developing the basis for this analysis is to find timetables that make it possible to establish when these tracks were actually used, where the trains stopped, how long it took them to cross borders, etc., and information on what these trains actually transported – persons, cargo, soldiers? The challenge here is to find documents.

that enable her to compare not only data between the countries, but also debates and discussions on the development of the railroad network, which may be held in different kinds of archives and institutions (industrial archives, state libraries and archives, private collections, academic libraries, researcher projects, databases etc.). Rather than having to travel to the respective archives (which will be situated in at least three countries: Lithuania, Poland and Germany), she wants to use a tool that helps her locate the archives and allows her to bring together data on the relevant files in the archives. Thus, if she looks at departments of a Polish Ministry, she not only wants to be able to find and access metadata or collection descriptions at a minimally detailed level, she also wants the system to help her locate the respective departments of the Lithuanian ministry (e.g. Polish customs department – Lithuanian customs department; Polish national rail headquarters – Lithuanian national rail headquarters) and also the equivalents of the German occupation regime dur-

ing the war. Furthermore, she would like to be able to upload other data from other sources to create a wider field for comparison, perhaps on a collaborative basis with colleagues.

Requirements

As a researcher:

| I want to (something) | So that (benefit) | Test Case and / or Input / Output data |
|---|--|--|
| Find maps and descriptions of railway lines | I can find historical topographical data | I want to find out what railway lines existed and which were newly constructed |
| Compare accounts from similar sources 'like for like' (including map registrations) | I can bring descriptions into agreement and make queries across them | I want to combine several information sources to generate a single map and single database/finding aid |
| Find timetables of railway lines | I can see how railway tracks were used | I want to find out when railway tracks were used, where trains stopped and how they crossed borders |
| OCR scan timetables | I can work with the chronological data | |
| Integrate chronological data into a map | I can visualize the use and topographical differences of a network | I want to have a map that shows me which railway tracks were the most heavily used, where the main railway hubs were, etc. |
| Find equivalent institutions in different states | I can find comparable data | I want to find the Lithuanian equivalent to Polish transportation institutions (ministry, railway, customs, etc.) |

Extensions

- the researcher could restrict herself to only some sources, so as to have a more unified approach

- the researcher could resign herself to travelling to these archives, so as to mitigate through interaction with the local experts her reliance on well described data to support discoverability
- the researcher could change her topic to focus on only some countries/institutions

2.1.1.2. Collection Holding Institution publishes data on the EHRI portal

Provided by: KNAW / NIOD / EHRI

Contributor(s): Petra Links / Annelies van Nispen

User Story

EHRI identifies a Collection Holding Institution (CoHI) with Holocaust related sources. The data is considered as relevant for Holocaust scholars. The relevant archives and collections have been created during WWII by persons, organisations or companies, or have been created after the war, for instance by survivors. The sources are usually kept in paper format, and in some cases they have been digitized, in rare instances as Optical Character Recognition (OCR). EHRI aims to integrate descriptions of the identified sources into its portal. EHRI contacts the CoHI and surveys the CoHI on the opportunities for data integration.

For this use case, digital metadata and/or representations of the content itself are not available. The CoHI has limited budget and (technical) expertise available. EHRI invests time and expertise to support the archive to make digital descriptions of the Holocaust related sources and the CoHI is willing to invest in making digital collection descriptions. EHRI and the CoHI make a plan of action together. EHRI wants the CoHI to make standardised descriptions. EHRI follows international archival standards:

- Encoded Archival Context – Corporate Bodies, Persons and Families (EAC-CPF)
- Encoded Archival Description (EAD)
- Encoded Archival Guide (EAG)
- General International Standard Archival Description (ISAD(G))
- International Archival Authority Record for Corporate Bodies, Persons, and Families (ISAAR(CPF))
- International Standard for Describing Institutions with Archival Holdings (ISDIAH)

The CoHI makes the descriptions and exports it to EHRI, who publishes the data on the EHRI portal.

Alternative story

The CoHI is not willing to invest in the digitization of collection descriptions itself so instead EHRI and the CoHI work together to make a plan of action. EHRI copies (non)digital

metadata from whichever format they are in or EHRI writes a high level collection description from scratch. For this purpose, guidelines are available based on and structured according to the standards mentioned above. EHRI publishes these collection descriptions on the EHRI portal.

Goal

EHRI presents the metadata of Holocaust-related sources of this institution on the portal.

Scope

Make metadata of Holocaust-related sources accessible via the EHRI portal.

Preconditions

- Staff with expertise in describing archival collections
- Guidelines to describe archival collections in a standardised way
- Tools that support describing archival collection in a user friendly way
- Export facilities at the CoHI to deliver descriptions to EHRI

Success End Condition

The metadata is integrated in the EHRI portal in a sustainable manner.

Fail End Protection

A disclaimer is available on the metadata to warn users of the portal that the data might be outdated (delivered only once).

Primary Actor

Collection Holding Institution (CoHI)

Other Actor(s)

- EHRI staff to identify the CoHI and to make a description of the CoHI
- EHRI technical staff
- Possible subcontractors

Trigger

CoHI wants to disseminate its Holocaust related collections with the research community, via the EHRI portal.

Main Scenario

- 1) CoHI contacts EHRI
- 2) The CoHI is willing to cooperate with EHRI
- 3) EHRI surveys the collection of the CoHI
- 4) EHRI and the CoHI make a plan of action (amount of work, level of description, selection of tools, person hours, planning)
- 5) CoHI makes descriptions according to EAD
- 6) CoHI exports the collection descriptions from its Collection Management System
- 7) EHRI publishes the descriptions on the EHRI portal
- 8) EHRI writes a description of the CoHI according to ISDIAH and publishes it on the EHRIportal

Extensions

- 2a) The CoHI is not willing to cooperate with EHRI
- 2b) EHRI copies (non)digital metadata or writes a high level collection description from scratch
- 4a) The CoHI doesn't have a collection management system available for making descriptions and installs one
- 7a) It is not possible to publish the descriptions on the EHRI portal

2.1.1.3. Holocaust Researcher investigates person information and networks

Provided by: KNAW / NIOD / EHRI

Contributor(s): Petra Links / Annelies van Nispen

User Story

Within the framework of EHRI a Holocaust researcher aims to investigate the networks in which European Jews operated during their persecution in the Second World War through prosopography. A prosopography can be defined as an investigation of a historical group linked by a common factor based on the connections between individual members of this group. The leading question is the way these members operated within and upon the social, political, legal, economic, and intellectual institutions of their time. Through a prosopography the researcher analyses patterns of activities and interrelationships within a historical group. The prosopographical approach as proposed deals with large quantities of archival source materials and involves mapping out and analysing the various networks represented in those sources by using computational techniques (Natural Language Processing). Together with a data manager / information specialist the researcher sets up a

“Linked Data” information model to capture relationships between entities. The information about the personal entities is structured according to the standard Encoded Archival Context – Corporate Bodies, Persons, and Families (EAC-CPF). An open-source toolkit is used for the creation and enrichment of prosopographical resources, integrating text mining tools and services to automatically tag and disambiguate the mentions of known entities, as well as to discover new entities that need to be added to the knowledge base.

The researcher gathers the archival sources from several collection holding institutions (CoHIs). He identifies this material through the EHRI portal and requests the CoHIs to digitize the materials and to make the content available in a machine readable format, e.g. alto-xml. The researcher analyses the prosopography and answers his research question.

Goal

Answer research questions that investigate the networks in which European Jews operated during their persecution in the Second World War.

Preconditions

- Archives with network information
- Suitable tools

Primary Actor

Holocaust researcher

Other Actor(s)

- CoHIs
- Data manager / information specialist

Trigger

Research interest

Main Scenario

- 1) Researcher defines research question
- 2) Researcher selects relevant sources
- 3) CoHI provides sources in a format requested by the researcher (e.g. alto-xml)
- 4) Data manager / information specialists extracts personal (EAC-CPF) and possible network information from data, using text mining tools
- 5) Researcher identifies and captures persons & relationships between entities

- 6) Data manager / information specialists represents this in a tagged network or graph structure
- 7) Researcher analyses the prosopography

Extensions

- 3b) CoHI digitizes the source and processes it to requested format

2.1.1.4. Historian wants to publish his research data and make it reusable with the DARIAH-DE repository

Provided by: FHP / DARIAH-DE

Contributor(s): Jenny Oltersdorf / Juliane Stiller

User Story

A historian wants to publish the research data she has gathered and used for publication of a peer-reviewed output so that other researchers can verify their results and reuse her data. She wants to publish the data in the DARIAH-DE repository.

Research data can be accessed via an API, and are arranged with EPIC-PIDs and therefore can be reused by other tools and services like the DARIAH-DE Collection Registry²⁴.

The DARIAH-DE Generic Search indexes the collections of the DARIAH-DE Collection Registry and enables user-friendly access.

Goal

Researcher wants to publish his/her research data, which is in the form of text files and spreadsheets, and make it reusable.

Scope

The research data is comprised of different file formats.

Preconditions

- Digital research data and preliminary metadata describing the data
- Standards for describing administrative and technical aspects of metadata

Success End Condition

- Publication of the research data and indexing of the data for further reuse.
- Creation of a persistent identifier for the research data collection and its objects.

²⁴ The DARIAH-DE Collection Registry includes information on repositories and metadata of collections.

Fail End Protection

The published data cannot be retrieved or is lost.

Primary Actor

Researcher in History

Other Actor(s)

Developers of the DARIAH-DE repository

Trigger

Researcher wants to publish her data or needs to publish her research data as a prerequisite for getting published in a journal.

Main Scenario

- 1) user selects her collection via publish web-interface of repository
- 2) collection objects including metadata go via the API to internal storage
- 3) user generates metadata for each object of the collection
- 4) user generates metadata for the collection itself
- 5) user determines legal status e.g. CC-0 license

Extensions

Linkage of research data with publication

2.1.1.5. Historian wants to track the dissemination of a given author's works during the Medieval and Early Modern period

Provided by: SISMEL / CENDARI

Contributor(s): *Emiliano Degl'Innocenti / Roberta Giacomi*

User Story

Within the history disciplines community, scholars are interested in the accessibility of research data on authors, sources (i.e. manuscripts and printed books) and transmitted works. Other related information, coming from repertories and hand lists, authority lists and bibliographies are important as well to provide additional context and are to be integrated. When dealing with multilingual contents, access to both Latin and vernacular resources is required. The researcher is interested in tracking over space and time the dissemination of a given text, e.g.: Donatus Ars minor (a Medieval condensation of the late Roman school-

book, in which a series of dialogues conveyed the rudiments of the language) in the Medieval and Early Modern era.

More generally, the user goal is to investigate the spread of literacy in Early Modern Western European society, since *Ars minor* was quite possibly the first book printed with moveable type both in Germany and in Italy.

Unfortunately, the original editions have been lost but the researcher can compensate for the loss of evidence today with the use of documentary material made available by focused initiatives²⁵ and other scholarly projects and databases.

Goal

Address the question of the spread of literacy in early modern European society using a combination of digital resources (i.e. metadata, descriptions etc.), based on different standards (i.e. DC, XML-TEI, Custom profiles, etc.)

Scope

Access information stored in catalogues of manuscripts held by contemporary libraries; assess what was available during the Medieval and Renaissance period by accessing information in catalogues of Medieval libraries; access related primary sources reproductions and descriptions; access related secondary literature.

Preconditions

- Named Entity Recognition (NER) service to extract relevant entities (i.e.: names of persons and places, titles of works etc.) [(N)ERD 2016]
- Reference tools for the disambiguation of:
 - names of persons and places
 - documents (i.e. manuscripts shelf marks)
 - titles of texts/works
- Access to a LOD web of authors, works, documents and related information (i.e. available information about origins and provenances of the documents)
- Tool to perform searches across multiple scholarly resources
- Tool to display the results on a map and / or timeline
- Knowledge on the structure of the involved databases and resources

²⁵ Like the 15c BOOKTRADE Project <<http://www.modlangs.ox.ac.uk/research/15cBooktrade/>>. It will make an edition of the only surviving bookseller's ledger from late 15th century Venice available for scholarship: it contains detailed information on the sale, with their price, of 25,000 printed books over a period of just under four years. Donatus' *Ars minor*, together with other texts for primary education, are the most sold, totaling around 1652 copies. This evidence, however, has to be compared against the contemporary, and previous, manuscript production of this work, both in terms of quantity, of geographical and chronological spread, and of their users (last access: 4.12.2015).

Success End Condition

A dossier with all the relevant resources related to the textual tradition of Ars Minor, including primary sources and secondary literature, is produced. It could be possibly exported and reused. A map and / or timeline displaying the information is available

Fail End Protection

Information about sources and / or secondary literature is not accessible. User has to perform many different searches over a number of dispersed resources.

Primary Actor

Researchers working in the history disciplines, acting as data consumers.

Other Actor(s)

- Researchers and institutions producing data on authors, texts, sources etc. (Typology: data providers)
- Holding institutions preserving sources (Typology: GLAMs, holding institutions)
- D/H community involved on the same field (Typology: standards developers)

Trigger

A researcher is interested in tracking the textual tradition of a given text or the transmission of a given manuscript.

Main Scenario

- 1) Survey all extant editions of Donatus, Ars minor in the Incunabula Short-Title Catalogue (ISTC)²⁶ and in TEXT-inc²⁷
- 2) Assess the 15th and 16th century use of these editions by discovering who were the users of the surviving copies in Material Evidence in Incunabula (MEI)²⁸
- 3) Establish how many Medieval and Renaissance manuscripts of this work survive today in our libraries using the meta-opac CERL Portal²⁹ to access a wide number of electronic catalogues of manuscripts.
- 4) Assess the presence of this work in catalogues of Medieval libraries in Europe, to understand the popularity and circulation of this work in the Medieval and Early

²⁶ <http://www.bl.uk/catalogues/istc/> (last access: 8.08.2016)

²⁷ TEXT-inc. A corpus of texts printed in the 15th century <<http://textinc.bodleian.ox.ac.uk/>> (last access: 8.08.2016)

²⁸ http://data.cerl.org/mei/_search (last access: 8.08.2016)

²⁹ <http://cerl.epc.uu.se/sportal/> (last access: 8.08.2016)

Modern period by using Medieval Libraries of Great Britain (MLGB3)³⁰, Biblissima³¹ and TRAME³² tools.

- 5) Ensure that the CERL Thesaurus is running at the back of the above listed tools to assure inclusiveness of data
- 6) Linking out to secondary literature on this work using TRAME and Biblissima tools

2.1.2. Language Related Studies

2.1.2.1. Natural Language Processing Expert wants to test her tool for semantic annotation on an available digital edition of historical texts

Provided by: CNR / ILC / CLARIN / DARIAH

Contributor(s): *Francesca Frontini / Monica Monachini*

User Story

Within the Language Technologies (LT) community, strong interest is building up in the potential for testing text analysis tools on corpora other than newspaper articles. Using CLARIN/DARIAH resource repositories, a language technology provider identifies a set of corpora that a particular community of scholars have made available. It may be a philologically curated electronic edition of a historical text, for instance the Nuova Cronica, a history of Florence by the medieval merchant Giovanni Villani. The tool the expert wants to test performs some type of semantic annotation, for example Named Entity recognition, in particular of persons and places. This could be done via linking to DBpedia.

LT experts would like to test their tools on this kind of data, but unfortunately they face a series of issues concerning input and output formats. More specifically, it is often the case that the tool developed by LT experts only takes plain text as input, whereas an electronic edition is – in the best case scenario – encoded in TEI/XML, or – in the worst case scenario – a HTML page. As a consequence, some code needs to be written in order to extract plain text from the TEI/XML or HTML. Even in the best scenario this may be complicated, as the details of the structure of the TEI schema are not well known among a wider community of LT experts.

Moreover, it is often the case that LT experts would like to reinject the automatic annotation in the original TEI, so as to send it back to the editors for validation, as they too might find it useful to have an enriched version of their text. But the tool only outputs data in a plain, one token per line, tab separated format that is commonly used by many LT applica-

³⁰ <http://mlgb3.bodleian.ox.ac.uk/> (last access: 8.08.2016)

³¹ <http://biblissima-condorcet.fr> (last access: 8.08.2016)

³² <http://git-trame.fefonlus.it> (last access: 8.08.2016)

tions. Building a wrapper that converts this into the right format is costly and may require collaboration with the editors of the TEI text.

Goal

An LT expert wants to test her Named Entity tagger on a digital edition of a historical text that has been made available online.

Scope

Semantic annotation of textual data.

Preconditions

- A language processing tool able to read text and find mentions of people and places, and referencing them with a unique link to a DBpedia URI
- The LT expert has the right permissions to access the resource
- An electronic edition in the language and of the type/genre required by the NLP expert
- Expertise in describing language technologies and natural language processing, but also in the semantics of TEI documents
- Documentation on the structure of the digital edition

Success End Condition

A TEI document is produced with enriched DBpedia links for mentions of person and place

Fail End Protection

Enrichment of the document cannot be produced; plain text may be extracted from TEI and enriched in the native format of the NER system.

Primary Actor

LT expert (Typology: a researcher that needs standards in order to achieve his / her research)

Other Actor(s)

- Philologists that produced the electronic edition of the text to be processed (Typology: collection/content holding institution)
- TEI community (Typology: institution/consortium involved in standards development)

Trigger

An LT expert finds the existence of an interesting corpus while browsing CLARIN / DARIAH or PARTHENOS repository

Main Scenario

- 1) LT expert sees the digital edition on a repository
- 2) LT expert is granted access to the full text of the digital edition
- 3) LT expert extracts plain text from the digital edition
- 4) LT expert runs tool on text
- 5) LT expert re-injects the results of her tool into the digital edition
- 6) LT expert contacts editors of the digital edition asking them to validate results
- 7) Editors validate and correct results
- 8) Editors provide LT expert with feedback on resource
- 9) Editors publish an enriched (manually revised) version of text with links to people and places
- 10) LT expert uses feedback to improve system

Extensions

- 2a) LT expert has no access to resource
- 3a) LT expert has issues in understanding format
- 3b) LT expert contacts editors
- 3c) Editors provide LT expert with documentation/help
- 4a) Tool fails due to encoding/linguistic problems
- 5a) Problems converting back from TSV to markup annotation emerge
- 7a) LT experts are not willing to collaborate
- 10) Philologist's feedback is provided in a format that is not usable for either training or testing the tool

2.1.2.2. Create annotated digital edition

Provided by: OEAW / CLARIN / DARIAH

Contributor(s): *Klaus Illmayer / Vanessa Hanneschläger*

User Story

Researchers from the domain of language-related studies (LRS) are creating a digital edition of texts, both for publishing and for preparing data for further analysis.

The source of the edition comprises heterogeneous material. It is printed (e.g. the scope of the edition is only published material) but it could also be a combination of handwritten manuscript, notes, letters, printed documents and so on (e.g. the estate of a poet).

LRS researchers (LRSRs) create a digital edition in different ways, depending on the level of experience regarding XML technology. It is often collaborative work. To date, the main purpose in many edition projects, is the publishing in print. For digital analysis and digital publishing an annotated version has to be provided. As there are no standards on tools to use in this area to date, this use case highlights best practice.

The annotation of the digital edition should cover interdisciplinary reusability, integration with controlled vocabularies (or generating domain specific vocabularies) and availability of (meta)data for visualization, statistical analysis and further processing.

LRSRs annotate in different ways; the vocabulary used is often language dependent. Use of XML TEI P5 is recommended and indeed this is a de facto standard in the scholarly community for creating a digital edition. However, there are different strategies for annotation and handling of the source material and there are a lot of variations in presenting a digital edition. The flexibility of XML TEI comes at the cost of a broad variety of different approaches on how to efficiently create a digital edition. Taking this into consideration, we recommend best practices for the whole creation process as shown in the main success scenario of this use case (e.g. highlighting preferred vocabularies and tools for NER, and pointing out which data to annotate and the data type/structure of an annotation). It is also necessary to discuss where it is useful to define standards in the process, mainly in the field of data exchange. The primary actors in this use case are researchers working in a team. They need standards/best practices in order to support the whole process of creating and publishing a digital edition.

Goal

Create an annotated digital edition of texts in XML TEI P5

Scope

LRSRs active in edition philology with experience in the creation of digital editions

Preconditions

- Transcriptions of the edited texts are digitally available
- Legal situation of processed texts is clear (copyright)
- Availability of relevant data sets for enrichment/semantic annotation

- Knowledge of annotating data in XML TEI P5

Success End Condition

Publication of the annotated digital edition

Failed End Condition

There is no digital edition or digital edition is not annotated

Primary Actor

LRSRs (= Team)

Trigger

Obtaining digital texts for the edition

Main Success Scenario

- 1) Prepare texts, sort them, compile them and integrate them into raw XML TEI P5
- 2) Tokenization and lemmatization of the texts
- 3) Perform NER (named entity recognition) on the texts
- 4) Enrich texts with edition specific annotations
- 5) Develop mode of presentation/layout
- 6) Publish edition

(Sub)Variations

- 1a) Transcriptions of texts are not made in XML TEI P5
 - 1a.1) Use-style sheet information for later mapping to XML TEI P5
 - 1a.2) Perform mapping

Extensions

- 4) If facsimile is available: connect pictures to text
- 6) Publish data

2.1.2.3. Build a corpus of linguistic data for analysis

Provided by: OEAW / CLARIN / DARIAH
Contributor(s): *Klaus Illmayer*

User Story

Researcher from the domain of language-related studies (LRS) is looking for data on a formulated research question. LRS researcher (LRSR) has starting points for the search

(e.g. keywords, domain, type of resource, language) and a concrete idea of useful data sets (e.g. required type of annotation, minimal size).

The LRSR wants to build up a corpus so she/he tries to get as much data as possible from different sources. The data is annotated (automatically by standard NLP-tools) based on their research question. Finally, the enriched corpus is analysed (computation of statistical information, metrics) and aggregated results are available in a human readable format (ideally visualized) for detailed inspection/exploration.

This user story is based on best practices. There is a lack of recommendations for tools and the usage of standards in parts of this use case, especially for the annotation and for the visualization. A lot of tools are available, but there are many different approaches for how to choose and combine them. The archiving of the research data and the actions taken to analyse the data is also open for discussion. As annotation and analysis depends on the research question and on the field of interest of the researcher, it would probably need more

'best practice' examples than standards. One open question - which is not covered in this use case - is the setting of standards for providing and presenting the research results so that interoperability is guaranteed.

The primary actor/actress in this use case is a researcher with experience in the field of linguistic data analysis. She/he needs standards and best practice examples in all steps of the main success scenario. For non-experienced researchers there is the need for easy-going applications and how-to manuals.

Goal

Analyse a corpus

Scope

Researchers mainly from LRS (but could be also from other research communities) working with annotated corpora.

Preconditions

- Research question
- Idea of useful data sets
- Availability of appropriate data sets
- Information about search platforms for repositories or/and already gathered data sets
- Knowledge on annotating data in XML TEI P5 and a concept for annotating the research data

- Availability of NLP-tools for given data and task

Success End Condition

Corpus annotated and analysed to cover the research question

LRSRs (need standards according to which to organize the corpus they are building)

Trigger

LRSRs start search for data based on a formulated research question

Main Scenario

- 1) LRSRs receive machine readable data from repositories (ideally in XML TEI P5)
- 2) LRSRs compile different data to a corpus (usually on a local environment)
- 3) LRSRs annotate manually data in XML editor
- 4) LRSRs perform analysis either manually or with the help of tools on annotated data

Variations

- 1a) Data provider does not grant access to repository
- 1a.) Obtain access or get data from provider in another way e.g. via email after declaration of consent
- 1b) Repository does not deliver data in XML
- 1b.1) Convert data in XML

Extensions

- 1') Data gathered offline from a data provider or as a result of a project
- 3') LRSRs use automatic pre-processing for annotation
- 3'.1) LRSRs need to post process the automatic annotation
- 3'.2) LRSRs use XML editor, script or specialized tool supporting batch processing to correct annotations
- 4') If data policy allows it, put new compiled corpus and analysis results into a repository
- 4'') Prepare data and results of analysis for online presentation

2.1.2.4. Interoperability in literature using the TEI

Provided by: CNRS / Huma-Num

Contributor(s): *Stéphane Pouyllau / Adeline Joffres*

User Story

The Huma-Num's Consortium "Authors of Corpora for the Humanities: Computerization, Edit, Search" (CAHIER) is a cross disciplinary consortium. It aims to bring together the various existing or planned initiatives in France in the fields of "Authors' Corpora". They come from literature, philosophy or themes related to a school or practice to provide coordination, share experience and promote access to data.

In that context, the consortium members had been thinking about building a unique core format based on TEI (Text Encoding Initiative) to describe digital objects (above all corpora) derived from various sources and different formats. They conducted a "grand dialogue" on metadata and data and started working on the project.

The second goal is to build a tool in order to publish and share data shaped in this format in an interoperable way. The WEB-OAI tool provides a virtual research environment in order to describe all the literature's objects considered in CAHIER's consortium in a normalized way and give (human) access to the catalogue. Another feature of WEB-OAI is to publish a TEI header's normalized metadata through an OAI-PMH repository in order to be harvested with rich metadata vocabulary (dcterms)

In order to achieve these goals, CAHIER organizes several workshops and summer schools for the community involved.

Goal

- To define a common interoperable format
- To build an open publishing tool and share data
- National and international coordination within TEI community

Scope

Building a unique core format based on TEI (Text Encoding Initiative) to bring together the various existing or planned initiatives in France in the fields of "Authors' Corpora".

Success End Condition

Finding a suitable data description and doing one recommendation for all the objects

Fail End Protection

Not enough metadata for processing to interoperability

Primary Actor

Huma-Num CAHIER Consortium and its partners (see: <http://cahier.hypotheses.org/partenaires>)

Other Actor(s)

TEI production line of the Centre for Open Electronic Publishing (CLEO – France) and Caen

University Press (PUC – France)

Trigger

- Need for diffusion in and by common catalogues
- Giving access to a normalized corpora
- Repository to be harvested using OAI-PMH protocol (ISIDORE, Gallica, Europeana etc.)

Main Scenario

- National coordination and coordination within research communities
- International coordination with European structures and others
- “Toolify” (develop specific tools) for literature

Extensions

- Work on uses for research (and not archiving) of TEI and EAD’s norms
- Promote good practices of describing resources
- Prepare the evolution toward semantic web technologies

2.1.2.5. Linking original text in literature studies to commentary, translations and external sources

Provided by: CLARIN / UCPH

Contributor(s): Lene Offersgaard / Claus Povlsen

User Story

Researchers working with Latin and Greek texts at UCPH need formats for linking information in commentary, translations and other sources to marked up versions of the original texts. This linking of information can facilitate publishing texts with commentary in two major uses: in a simple reading system that can easily display needed and interesting information based on the user’s reading skills; and in a more advanced system that can support research, development of new commentaries to students, and other material. The ability to link in a standardised way should also enable researchers to easily extend the in-

formation in the commentary and the linking to other resources in collaboration with other researchers and students. This interest is not limited to researchers in Latin and Greek, but can be used in studies for other languages as well. The primary challenge is that current standards are available on different sub areas of this setup, but a single researcher does not have the time resources, the knowledge of many standards, and the overview of how to combine the right standards and formats when creating a commentary or a translation of a source text.

Data creation is mainly done by the researchers and teachers. Some write commentary for students and some focus on commentary to share with researchers. Translations can be created in modern language to make available an “easy/modern” version of old texts. Furthermore, data creation can also be seen as the linking of existing resources to each other, and linking sections, details or words in one text to another. In this view data creation is both the link and the supplementary information (metadata) describing the link.

Goal

A defined set of formats for texts, translations and commentary that enables linking on different levels. Levels of linking can be:

- a. Attaching a link to a specific word, with a specific note in the commentary
- b. Linking a sentence in the source text to another source text that has the same sentence or cites the first source text
- c. Linking a section of the source text to a section in a translation of the text
- d. Linking a source text to e.g. a translation of the text.

Format and mechanism to link information among the different texts and to other external resources should be tested in a web-application and documented.

As TEIP5 is a commonly used format for annotation of text, reuse of TEIP5 where possible is preferable. However, TEIP5 has some limitations in specifications of linking, such as the need to handle alignments of the types 1:n, n:1, 0:1, and 1:0. This is not handled well by TEIP5. Another thing to be aware of is that TEIP5 allows a large freedom in annotations of text, and researchers could be guided by further standardisation of the use of TEIP5.

Scope

Researchers studying texts with an interest in sharing commentary or translations.

Preconditions

- A set of texts with commentary and translations that can be used as test material without copyright issues.
- Knowledge of TEIP5, Linked Open Data, and other formats and standards that are relevant.
- Descriptions and examples of the needed linking
- Access to digital resources e.g. dictionaries or other resources that can test the linking mechanism.

Success End Condition

A simple reading system publishes texts and translations for students, and a more advanced system enables researchers to share new commentary and links to resources in a dynamic way. Sub-products are formats for commentary, reference system (links) to original texts converted to a documented format, formats for aligning source and translations. Examples of a usable linking format to external resources.

Primary Actor

The primary actors are researchers that need standards in order to express his / her research (e.g. commentary) or need formats and a standardised way to create and annotate links.

Trigger

Teachers in translation studies want to use a digital platform in teaching. Researchers want to collaborate on commentary of texts and linking to other sources.

Main Scenario

- 1) Convert texts, translations and commentary files to TEIP5 format, using a specification of which functionality to use from TEIP5
- 2) Extend format to handle linking, based on examples of how linking can be made. Standards are also important here, but linking cannot be done satisfactorily by using TEIP5; the format has to be changed/extended.
- 3) Enable linking to other resources/applications with, for example, an online dictionary with examples of how to do it.
- 4) Create reading web application
- 5) Upload texts, translations and commentary files in reading application
- 6) Test upload
- 7) Test linking is working – hopefully in an automatic way

- 8) Prepare documentation including examples of use of standards
- 9) Publish the resources online with a URL for teaching

Variations

- 4a) Facilities for administration of copyrights has to be included

Extensions

- 1a) Researchers want to collaborate in creating commentary

2.1.2.6. Sustainability and improved viewing of Assyrian text resources

Provided by: CLARIN / UCPH

Contributor(s): Lene Offersgaard / Claus Povlsen

User Story

The user requirements consist of more elements. In order to secure and procure the ancient texts, the data must be represented and embedded in a sustainable format.

Secondly, the user wants to make queries for relevant text collections by exploiting the metadata assigned to the text collections.

Finally, the user needs to be able to view photos of the original clay tablets, their transliterations and translations in English in the same window. This parallel viewing implies manually annotated alignment between the transliterations and their translations. Even though the Assyrian language operates within a sentence concept, the transliterated sentences are not marked up with punctuation information, meaning that sentence alignment requires manual work.

Scope

An existing repository with backend and frontend functionality

Preconditions

- Copyright license to establish public access to the data
- the data is available in a digital format
- the data texts are manually aligned.

Success End Condition

The text data sets are stored in sustainable format and on a server that is maintained in a long run perspective. The users can view the three representations of the data in parallel.

Primary Actor

Researcher from the field of Assyriology (who wants to share his/her data and, at the same time, wants to ensure that the data is kept sustainable)

Trigger

Display of the three representations of the text data

Main Scenario

- 1) The user makes a query for the collection of texts that he/she wants to use in his/her research.
- 2) The user triggers a view of the results of the query.
- 3) The search results for text data are displayed as clay tablets, their transliterations and translations in parallel.
- 4) The users can scroll through the text data preserving the parallel viewing of the three representations of the text.

Extensions

- 2a) The users are offered the possibility of making queries directly in the transliterated version of the text data.
- 3a) The results of the queries are shown as parallel representations.

2.1.3. Archaeology, Heritage and Applied Disciplines

2.1.3.1. Conservation scientist wants to publish information about experimental conditions for Raman analysis of wall painting fragments and report in particular proper experimental measurement conditions for safely detecting and identifying certain types of pigments

Provided by: FORTH / IPERION CH

Contributor(s): *Panayiotis Siozos / Demetrios Anglos*

User Story

A conservation scientist wants to perform Raman analysis on a series of wall painting fragments. She/he wants to define the most suitable experimental conditions for analysis (laser wavelength, laser power etc.). The conservation scientist is searching for data and analysis guidelines in digital resources and utilizes the information collected and procedures proposed in order to perform the experimental measurements.

However, she/he discovers that the material is undergoing weak discolouration when the laser intensity exceeds a certain threshold. She/he wants to report this finding as soon as

possible in order to advise other users on the issue of safe limits concerning irradiation of sensitive paint materials during Raman analysis. Thus, she/he sends a brief report to the authors of the above digital resources in order to inform them about the findings. The authors update the resources and include the finding.

Goal

A conservation scientist wants to make his / her observation of weak discolouration available when the laser intensity exceeds a certain threshold during Raman analysis of wall painting fragments.

Scope

To report and to update proper experimental measurement conditions for safely detecting and identifying certain types of pigments using Raman spectroscopy.

Preconditions

- The conservation scientist is able to find any information about the experimental conditions in digital resources
- The conservation scientist has access to the specific digital resource

Success End Condition

The digital resource is updated quickly and accurately.

Fail End Protection

A disclaimer is available on the metadata to warn users of the digital resource that the data might be outdated.

Primary Actor

Conservation scientist

Other Actor(s)

Authors of the digital resource

Trigger

Conservation scientist detects weak discolouration in the wall painting fragment after Raman analysis

Main Scenario

- 1) The conservation scientist performs research in digital resources for appropriate experimental measurement conditions
- 2) The conservation scientist collects the experimental conditions from a digital library
- 3) The conservation scientist applies the experimental conditions to the analysis of the material
- 4) The conservation scientist detects weak discolouration on the wall painting fragment
- 5) The conservation scientist prepares a report (document, graphs, images)
- 6) The conservation scientist uploads the report by using the digital library platform
- 7) The platform informs the authors about the uploaded information
- 8) The authors evaluate the reported findings
- 9) The authors confirm the reported findings
- 10) The authors update the information of the digital library

Extensions

- 5) The conservation scientist is not willing to report the finding
- 6a) The authors are not willing to update the digital resource
- 6b) There is no available procedure to update the digital library

2.1.3.2. Researcher using lasers in conservation/restoration identifies the necessity of standardised reports of the laser application conditions and the evaluation of the obtained results

Provided by: CSIC / IPERION CH
Contributor(s): *Marta Castillejo / Esther Carrasco*

User Story

Laser cleaning in Cultural Heritage (CH) is an activity that proceeds without standards at present. A researcher in the field has been collecting documents about procedures that use lasers for conservation and restoration of artworks and heritage objects and substrates. The researcher has difficulties in comparing results from different published sources due to the lack of specified information about the selected conditions and parameters. He/she finds the necessity of developing guidelines to document all the relevant data related to the laser- material interaction and the analysis of the produced effects, including cleaning quality and side effects.

Data creation process and structure/nature of the research data

Laser cleaning in CH requires a procedure that can be described by three successive steps:

- a) The physical and chemical characterization of the sample/artefact, before the laser cleaning process. A full characterization requires several complementary microscopic and spectroscopic techniques. This analysis determines the specific treatment to be performed on the surface sample.
- b) The laser cleaning itself. If the experimental setup has implemented a suitable analytical technique, the cleaning can be in situ monitored.
- c) The post cleaning sample/artefact characterization, equivalent to the one in a).

This analysis allows the user to evaluate the cleaning quality and to determine the presence of side effects. Depending on this evaluation, b) can be performed again with new adjusted conditions.

The process of data generation follows the steps a), b) and c). The types of acquired data are (usually 2D) digital images and spectra (stored as or converted to ASCII format), which are processed afterwards and displayed like graphs and images (image file formats). The measurement conditions for each employed technique must be reported, but these data are not necessarily included in the acquisition data files. They are annotated (in laboratory books, spreadsheets or similar files) and transferred to the corresponding report afterwards. The in situ monitoring (if done) generates data of the same nature than those described in a) and c). The employed experimental setup is documented with a diagram and/or pictures (digital images). The laser characteristics and the laser cleaning parameters employed (laser wavelength, laser energy, fluency, laser pulse duration, repetition rate and the number of pulses) need to be annotated (in laboratory books, spreadsheets or similar files). All the previous data from b) are included in the aforementioned electronic report with the required descriptions and explanations.

Goal of standards

The standard would indicate a methodology to document the kind, extent, and objectives of the laser cleaning, the employed laser cleaning parameters and the obtained results, including cleaning quality as well as side effects or induced damage. The standard would contain guidelines to report the aforementioned required steps in the employment of lasers for conservation/restoration, in order to allow the comparison of results with different treatments performed in different studies.

Existing standards

In 2016 a standard is approved: EN 16782:2016 Conservation of cultural heritage - Cleaning of porous inorganic materials - Laser cleaning techniques for cultural heritage. It provides the fundamental requirements of the laser parameters and guidelines for the choice

of the laser operational parameters, in order to optimize the cleaning procedure of porous inorganic materials.³³

Goal

The researcher obtains a collection of relevant parameters related to the conditions of the laser material interaction in the CH conservation/restoration area. The researcher gets guidelines to assess the physicochemical effects of the laser irradiation and the undesired side effects or collateral induced damage and their systematic documentation. The researcher produces a report, which includes all the relevant information.

Scope

A standardised report on how to document/report the conditions of laser cleaning applied to

CH in a systematic way is published on the IPERION CH and/or PARTHENOS portals.

Preconditions

- Expertise in laser characterization, cleaning of artworks and heritage objects and substrates of CH
- Documentation about standards in materials analysis in CH and treatment by lasers
- Guidelines to describe procedures in a standardised way
- IPERION CH / PARTHENOS infrastructures to publish documents on guidelines and standards

Success End Condition

- A guideline to create reports on laser cleaning is produced, where relevant parameters of laser conditions are identified and listed in a standardised way and assessment of laser effects is systematically documented. This guideline is published in the IPERION CH / PARTHENOS portals in a sustainable manner.
- A standardised report of laser cleaning prepared by the researcher is published in the IPERION CH / PARTHENOS portals in a sustainable manner.

Fail End Protection

A disclaimer is available on the metadata to warn users of the portals that the related data might be outdated (delivered only once).

³³

https://standards.cen.eu/dyn/www/f?p=204:110:0::::FSP_PROJECT,FSP_ORG_ID:38496,411453&cs=148F3E52FD2C8CF54836A9D4470681779

Primary Actor

Researchers in the field of laser cleaning in conservation/restoration of CH.

Other Actor(s)

- Other researchers employing laser cleaning for restoration of artefacts.
- Researchers working in the development of laser equipment for restoration.
- CH conservation and Heritage Science communities interested in the application of laser for conservation/restoration.
- IPERION CH / PARTHENOS portal.

Trigger

Researcher finds the existence of information of interest by browsing IPERION CH and PARTHENOS websites and repositories.

Main Scenario

- 1) Researcher contacts IPERION CH due to the necessity of characterizing a CH object and documenting the materials composition of the artwork before restoration.
- 2) IPERION CH analyses the necessities of the researcher in order to evaluate the viability of the collaboration.
- 3) IPERION CH guides the researcher and eventually supplies the access to the necessary equipment, archives or infrastructures.
- 4) The researcher and IPERION CH / PARTHENOS investigate possibilities to generate guidelines and reports on a standardised way.
- 5) IPERION CH / PARTHENOS offers tools to publish the report in the standardised format.
- 6) The researcher publishes his/her standardised report on laser cleaning.

Extensions

- 3) The researcher and IPERION CH do not have access to the required resources.
- 4) IPERION CH / PARTHENOS community rejects standardised reports since the process of defining laser cleaning conditions and the evaluation of results are considered too complex.
- 5a) IPERION CH / PARTHENOS has no tools to publish standardised reports and guidelines.
- 5b) The researcher has issues in understanding format to develop standardised reports.

2.1.3.3. A dataset for the products used in conservation treatments, in order to share information about their application parameters, their effectiveness and their durability in time, related to the type of material and its state of conservation

Provided by: CNR / ICVBC / TeCon@BC

Contributor(s): *Rachele Manganelli Del Fà / Marco Realini*

User Story

Among the goals pursued by research institutions and professionals involved in the protection and conservation of Cultural Heritage (CH), there is the efficacy and the durability in time of the conservation treatments, carried out for buildings or objects of artistic and historical importance.

In literature is possible to find many examples to verify the efficacy of the treatments carried out on building and artefacts or laboratory tests on the resistance of products subjected to accelerated aging cycles, but no tools exist that are able to "correlate" the performance of protective or reinforce treatments and their durability in time, with different types of materials or substrates, decay and climatic conditions to which they are subjected.

In other words, you can't get information in a direct and simple way, but still based on scientific data, on the most appropriate products for a particular material, exposed to specific environmental conditions.

Data creation process/data types

First of all, it is necessary to identify all the parameters which appear to be most significant for the description of the state of conservation of the constituent material and the environmental conditions to which it is exposed. It should always be remembered that the descriptive parameters of conservation status should be measured through simple, non-destructive or micro-invasive methodologies, from which we choose the most relevant for the type of material. For example, for a stone material the capacity of absorption of water and its mechanical characteristics are very important, for metals it is important to define the possible presence of alterations of the alloy and active corrosion phenomena. Similarly for the definition of the environmental parameters that most affect the effectiveness and durability of the products, the data related to thermo-hygrometric variations, the concentration of pollutants, the amount of rain and solar radiation are to be considered more meaningful; obviously data directly acquired by sensors placed on site are to be preferred.

Keeping in mind that the methodology described below should not to be considered a standard but a "good practice", and that the individual cases may also be addressed in a different way, we can summarize the most significant phases in the following way:

STEP ONE: *physical, chemical, mineralogical and petrographic characterization of materials and definition of its state of conservation* by taking samples and using several analytical techniques. All the analyses are closely related to the type of material.

Therefore, we present below an overview of the most common techniques, the reasons of their use and what kind of output they return. **Very important: not intended to be exhaustive.**

Just some of the available techniques are cited here. It is not possible to standardise the steps to characterize the materials and their state of conservation, because of this depends on the type of material and on its state of conservation. On the contrary, all the types of output data are well represented: numerical data, spectra, digital 2D images and written texts.

PORTABLE FIBRE OPTIC REFLECTANCE SPECTROSCOPY (FORS)

Description: The application of FORS is mainly addressed to the identification of pigments, dyes, and to the detection of the chromatic coordinates and their variations. It's a totally non-invasive technique and thanks to the portable equipment, allows in situ measurements.

The identification of the pigments is carried out for comparison with reference spectra from mock-up paintings.

Data Types: spectra that can be easily exported in ASCII format.

PORTABLE X-RAY FLUORESCENCE SPECTROSCOPY (XRF)

Description: X-ray fluorescence spectroscopy is a non-destructive technique that allows the researcher to obtain the elemental composition of the materials through the study of the radiation of the secondary X fluorescence. It can be applied in situ and can be used for the characterization of inorganic material such as metals, alloys, ceramics, pigments, corrosion products, etc.

Data Types: spectra that can be easily exported in ASCII format.

FTIR AND RAMAN BENCHTOP AND PORTABLE SPECTROSCOPY

Description: Infrared and Raman spectroscopy allows the chemical characterization of both organic (varnishes, coatings, adhesives, binding media, etc.) and inorganic materials (pigments, corrosion products, salts, etc.).

Data Types: spectra that can be easily exported in ASCII format.

BENCHTOP X-RAY DIFFRACTOMETRY

Description: The technique provides the mineralogical identification of crystalline phases; it is used for archaeometric studies and for the assessment of the conservation state of

natural and artificial stone materials (mortar, ceramics, etc.), for the study of pigments and their alteration, for the study of the alteration products of metals and of the crystalline phases in the glass. Cross-sections, thin sections and micro samples can be analysed by the micro diffraction system (spot 100 µm).

Data Types: spectra that can be easily exported in ASCII format.

BENCHTOP POLARIZED LIGHT AND REFLECTED LIGHT MICROSCOPY

Description: the techniques allow the researcher to analyse thin cross-sections providing the petrographic and stratigraphic characterization of substrate, finishing layer and decay.

Data type: digital 2D image also features an additional relationship with the interpretation of the section (written text).

BENCHTOP MERCURY INTRUSION POROSIMETRY

Description: This is useful for the microstructural characterization of all non-metallic porous materials, for an assessment of their degradation.

Data types: numerical data and graphics that can be easily exported in ASCII format.

STEP TWO: *evaluation of the environmental parameters that have most influence the degradation of constitutive material* (e.g. thermo-hygrometric variations, concentration of pollutants, rain, wind, solar radiation). In general this kind of data derive from sensors for detecting environmental directly placed on site.

When it is not possible to have sensors on site, it would be desirable to derive the information about environmental parameters directly from other databases available online.

Data type: numerical data and graphics that can be easily exported in ASCII format.

STEP THREE: *determine the characteristics of the surface and of the material before treatment (Time Zero)*. To assess possible modifications of the surface due to the treatment, it is essential to know its starting conditions (Time Zero – T₀). The tests must define the characteristics of the constituent materials such as colour, water absorption, resistance, and so on.

Below we give an overview of the most common techniques, the reasons of their use and what kind of output they return. **Very important: not intended to be exhaustive.**

PORTABLE OPTICAL MICROSCOPY

Description: the technique allows the observation of the surface at high magnifications, to assess and document in detail its characteristics. Furthermore, the technique is a valid support during diagnostic campaigns to select the best analytical approach.

Data types: digital 2D image also features additional relationships with the interpretation of the images (written text).

PORTABLE COLOURIMETRY/SPECTROPHOTOMETRY

Description: the technique provides an objective and reproducible definition of the colour of a surface, converting the colours perceived by the human eye into a numerical code (3 numbers, called tristimulus values). It is applied for very important and different aims such as monitoring possible variations of the colour of surfaces due to treatments, to deterioration or to following interventions, making comparison between differently treated or differently exposed surfaces.

Data Types: numerical data easily exportable in a spreadsheet

BENCHTOP MERCURY INTRUSION POROSIMETRY

Description: This is useful for checking the effects of conservative treatments.

Data types: numerical data and graphics that can be easily exported in ASCII format.

PORTABLE CONTACT SPONGE METHOD

Description: Contact sponge method is easy usable in situ, unlike other water absorption methods. Water absorption measurements are very widespread because they are useful for the evaluation of the conservation state of a surface and of the performance of conservative treatments, in particular protective ones.

Data type: numerical data easily exportable in a spreadsheet.

PORTABLE PEELING TEST DEVICE

Description: An adhesive tape, previously weighed, is applied with a regular pressure on the surface, and subsequently removed with a controlled speed; during the tape removal the instrument's control software acquires the values of breakout forces. After the removal the tape is weighed again.

These operations can be repeated even after a consolidating treatment in order to compare the data obtained before and after treatment.

Data type: graphics that can be easily exported in ASCII format.

PORTABLE ULTRA-CLOSE RANGE PHOTOGRAMMETRY

Description: The technique is used to assess the state of conservation and monitoring of restoration projects in the field of Cultural Heritage. Generating a 3D model of the studied area, is possible to estimate the surface modification, and generate roughness profiles. Moreover the 3D models, generated at different times, can be overlapped in order to study

the surface modification. This can be easily used in situ, on painted surfaces, fresco, stone and wood material.

Data type: 3D model and roughness profile, both easily exportable in ASCII format.

STEP FOUR: *choosing the best product and treatment methodology.* In order to accomplish this task, data from Step One and Step Two have to be available and analysed. For the description of the final choice it is very important to report: the name of the product, the technical sheet, the solvent and the concentration, the method (e.g. brush, poultice) and the time of application.

This information may be described through text boxes.

STEP FIVE: *Monitoring the behaviour of the applied product to determine its durability and its effectiveness.*

The tests carried out at T_0 (untreated surface) will be repeated after the application of the product with a scheduled frequency ($T_1, T_2, \dots T_n$), or after ageing cycles if tests are carried out on specimens in a laboratory.

Conclusion

Finally, in general the types of data are numerical data, spectra, graphs and digital 2D image, even if sometimes is possible to generate a 3D model of the surface (in this case data can be exported in ASCII format or as a .xyz file).

The measurement condition, environmental parameters, and any other information may be recorded in a report that may contain digital images.

Goal

Availability of a dataset containing information about a large series of treatments carried out on site and in laboratory, from which the benefits provided by various products in different situations can be deduced. A system that is able to describe the investigations performed, the obtained results, and the conservation history (present and past works) for each material.

Collect a large series of data about conservation treatments, in order to advise planners and restorers to identify the most suitable products for the conservation of artefacts under certain conditions (environment, decay), and choose the best treatment methodology (application technique, application time, concentration).

Preconditions

- Identified large series of products and their best method of application, in order to obtain the best results, by laboratory and on site tests.
- Identified most significant parameters for the description of the state of conservation of the constituent material and the environmental conditions to which it is exposed.
- Identified large series of materials, degradation processes and their evolution in time for the works exposed outdoors.

Success End Condition

- The most suitable product is chosen, considering many parameters and criteria and taking advantage of previous tests and conservation yards.
- Researchers develop a common report that includes the most relevant information necessary for the evaluation of treatments.

Fail End Protection

The choice of product is based on a restricted number of “criteria” that differ from task to task.

Primary Actor

- Researchers who needs information about a treatment in order to know the state of art;
- Researchers who want to share their results in order to amplify the result of a specific research;
- Researchers who want to compare their results;
- SME working in the development of products for conservation that wants to test its products and looking for a good practice guidelines.

Other Actor(s)

- SME working in the development of products for conservation;
- Professionals in the sector of CH;
- Other researchers in the field of CH.

Trigger

Conservation work.

Main Scenario

- 1) The user contacts PARTHENOS to access existing documentation or enter new data.
- 2) PARTHENOS evaluates a possible collaboration.

- 3) Provides access to view existing data.
- 4) Guides the data input through a wizard and standardised procedures.
- 5) The users publish their data or allow access to previously published data

Extensions

- 4) The format of the data entered is not usable within the platform

2.1.3.4. DYAS contacts CoHI and asks to integrate metadata of its digital collection into the “Humanities Resources Registries” portal

Provided by: AA / DARIAH-GR

Contributor(s): *Christos Chatzimichail*

User Story

Framework

The Greek Research Infrastructure Network for the Humanities - DYAS - is a network of Greek academic institutions, universities and research centres, which was established in order to contribute to the development of research in the Humanities using information technologies. Four partners of DYAS network, Academy of Athens (AA - project coordinator), Athena Research Centre-Digital Curation Unit (DCU - system developer), National and Kapodistrian University of Athens-Faculty of History and Archaeology (UOA) and Athens School of Fine Arts (ASFA), developed within DARIAH-GR the “Humanities Resources Registries” portal, which consists of the Organisations Registry, the Collections Registry, the Persons Registry and the Metadata-Standards Registry. The user, an arts and Humanities researcher or scholar, can use the registries to access information on Greek institutions or individuals and the collections, both analogue and digital, that they own or manage. The tool takes advantage of existing expertise and available digital resources to both improve the quality of users’ research and for educational purposes. The content of the digital tool is being continuously enriched and updated with the aim of enhancing the visibility of Greek analogue and digital content and to provide increased access to scholarly content.

Data creation process and structure/nature of the research data

AA identifies and contacts an organisation (or an individual) that manages/owns a digital collection, asking to integrate descriptive metadata about the organisation, the staff, the collection and the metadata schema used into the “Humanities Resources Registries” portal. The organisation agrees to share information of its structure (Organisations and Persons Registries) and its collection (Collections Registry), but it appears that it does not use

a standardised way to describe that collection (Metadata-Standards Registry). AA asks if it can receive a collection's metadata by filling in a form based on Dublin Core (DCCAP) that contains the relevant fields for the Registry (dcterms: title, language, subject, abstract, rights, format, creator, provenance, spatial, temporal, etc.). The institution emails back the form filled with the available information. AA evaluates the incoming information and finally integrates metadata into the Registry.

Issues

The main issue emerging during this process is that collection holding institutions, especially the smaller ones, tend to face serious problems in describing their digital content in a standardised format. They either make a very basic description lacking essential elements, or, more often, they do not make such a description at all – in most of these cases they turn out to be totally unfamiliar even with the term “metadata standard/schema”. In addition, it is not unusual at all for the contact between the collection institution and the developer of the digital content (IT company) to have been lost, and - as a result - the curator/manager of the digital collection and contact person of the institution is unable to provide standardised metadata, even if it exists. Therefore, unlike the other three registries in which data is continuously enriching, the content of the Metadata-Standards Registry remains deficient and unsatisfactory.

Requirements

Although there are various metadata schemas, still there is a need, at least locally, to develop and appropriately disseminate overall and clear methodological guidelines on why and how data should be structured and presented.

Goal

Collection Holding Institutions (CoHI) and individual owners understand the importance and the methodology of metadata standardisation and start using standards compatible with DYAS Registry to describe their digital collections. Preferably used standards:

- DCAP - Dublin Core Application Profile
- DCAT - Data Catalog Vocabulary (metadata registry standard)
- ESE
- EDM - Europeana Data Model
- DC - Dublin Core
- ARIADNE - a virtual research infrastructure for archaeology

Scope

- Make metadata schemas easily integrated into DYAS Humanities Resources Registries.
- Enrich Metadata-Standards Registry.

Preconditions

- DYAS/AA specify the preferable metadata schemata
- Describe digital collections in a standardised way
- Contact between collection manager and developer

Success End Condition

- Metadata integration into Humanities Resources Registries
- Inserting content process into the Registry becomes simpler and not as time-consuming
- Use of standardised descriptive metadata increases sustainability of digital collections
- Publishing metadata into DYAS Registry increases visibility of collections

Fail End Protection

Disclaimer stating that the institution/individual has clear legal rights over the collection

Primary Actor

DYAS/AA

Other Actor(s)

- CoHI/individual
- Digital collection developer

Trigger

- DYAS/AA wants to integrate metadata of a digital collection into the Humanities Resources Registries portal
- Institution wants to publish into the Registry descriptive metadata of its digital collection.

Main Scenario

- 1) AA contacts Institution
- 2) Institution agrees to share information about the organisation, the staff, the collection and the used metadata schema

- 3) AA sends a DCCAP based form with guidelines about content and format
- 4) Institution provides all the available information in a standardised format
- 5) AA evaluates the information and, if necessary, further supports institution
- 6) Collection metadata is published into DYAS Registry

Extensions

- 1) Institution does not want to publish description of its collection into DYAS Registry
- 2) Institution has unclear or no legal rights over the collection
- 3) Collection manager fails to re-establish contact with the collection developer
- 4) The provided description does not meet the basic criteria/guidelines and AA does not integrate data into Registry

2.1.3.5. Private Foundation wants to publish the digital collections of its library and museum in the online Public Access Catalogue and in Internet Culturale and CulturalItalia

Provided by: MIBACT-ICCU / ARIADNE

Contributor(s): Sara di Giorgio / Antonio Davide Madonna

User Story

A Private Foundation (AF) founded by a philanthropist in the late 1800s, has digitized its vast private collections related to art works, historic books and manuscripts and wants to share it in the Online Public Access Catalogue of the National Bibliographic Service (OPAC SBN) in Internet Culturale, the digital library of the Italian Libraries, and with CulturalItalia, the Italian National Aggregator.

AF contacts ICCU (who manages the OPAC SBN), Internet Culturale and CulturalItalia to ensure clear instruction on standards and guidelines. The digital cultural heritage sector is very well normalized for cataloguing physical objects and publishing and making the related digital collections interoperable in the main National and European portals.

ICCU offers technical support to cultural institutions, both public and private, during the process of sharing their digital collection in the main National portals. Moreover, CulturalItalia is interoperable with Europeana and other international initiatives like ARIADNE and if the content provider agrees it can also share its collections within those European Portals. AF will have its catalogue and digital collections available on its own website and on other National portals.

The museum's objects will be hosted and published by MuseiD-Italia, a digital library for museums, integrated in the CulturalItalia portal.

Standards, guidelines and tools are available on line (mainly in Italian):

Bibliographic resources

- Catalographic rules for libraries:
 - [Reicat³⁴](#) Catalographic National Code
 - Rules for a uniform title of music materials
 - [Guideline for cataloguing modern material in the National Librarian System³⁵](#)
 - Semantic cataloguing: the [Thesaurus³⁶](#) of the new subject of the National Central Library of Florence in SKOS format
 - See more guidelines here:
<http://www.iccu.sbn.it/openccms/openccms/en/main/standard/37>
- Descriptive standards and conceptual models
 - [ISBD³⁸](#): International Standard Bibliographic Description. Consolidated edition
 - [Functional requirements for Authority Data³⁹](#). A conceptual model
 - FRBR
- Metadata for libraries objects
 - [Dublin Core Metadata Element Set⁴⁰](#)
 - [Dublin Core Mapping / UNIMARC⁴¹](#)
 - [MAG standard⁴²](#) Metadata for digital bibliographic objects
 - [MAG User manual⁴³](#): html – pdf
- Digitisation Guidelines
 - [Guidelines for digitisation projects relating to photographic material⁴⁴](#)
 - [Guidelines for the digitisation of maps⁴⁵](#)
 - [Guidelines for the digitisation of proclamations, broadsides and single-sheet publications⁴⁶](#)
 - [Technical guidelines for the creation of digital cultural contents⁴⁷](#)
- Technical documentation about CulturalItalia (Application profile, standard, mappings, tools and guidelines)

³⁴ <http://www.iccu.sbn.it/openccms/export/sites/iccu/documenti/2015/REICAT-giugno2009.pdf>

³⁵

http://www.iccu.sbn.it/openccms/export/sites/iccu/documenti/2012/Guida_con_esempi/GUIDA_SBN_giugno.2012.pdf

³⁶ <http://thes.bncf.firenze.sbn.it/>

³⁷ <http://www.iccu.sbn.it/openccms/openccms/en/main/standard/>

³⁸ http://www.ifla.org/files/assets/cataloguing/isbd/isbd-cons_20110321.pdf

³⁹ http://www.ifla.org/files/assets/cataloguing/frad/frad_2013.pdf

⁴⁰ <http://dublincore.org/documents/dces/>

⁴¹ http://www.ukoln.ac.uk/metadata/interoperability/dc_unimarc.html

⁴² http://www.iccu.sbn.it/openccms/openccms/it/main/standard/metadati/pagina_267.html

⁴³ <http://www.iccu.sbn.it/openccms/openccms/documenti/manuale.html>

⁴⁴ http://www.iccu.sbn.it/openccms/export/sites/iccu/documenti/Linee_guida_fotografie.pdf

⁴⁵ http://www.iccu.sbn.it/openccms/export/sites/iccu/documenti/linee_guida_digit_cartografia_05_2006.pdf

⁴⁶ http://www.iccu.sbn.it/openccms/export/sites/iccu/documenti/linee_guida_bandi_sett.2006.pdf

⁴⁷ <http://www.minervaeurope.org/interoperability/technicalguidelines.htm>

Museum resources

- Catalographic rules for museums objects
 - Standards for cataloguing archaeological, architectural, artistic, ethno-anthropological, technical scientific and natural heritage objects by ICCD
- Metadata for digital museums collections for MuseiD-Italia / Culturaitalia
 - [MuseiD-Italia Application profile based on METS](#)⁴⁸
 - [Application for generating metadata automatically](#)⁴⁹ (MUSEI-METS)
 - Musei-Mets [Handbook](#)⁵⁰
 - [Validator](#)⁵¹
- MuseiD-Italia is interoperable with Culturaitalia via OAI-PMH
- Culturaitalia is interoperable with Europeana via OAI-PMH. The AF digital collections related to historic library will be available.

Goal

To make CH digital collections in the main National and International portals available to allow access to digital cultural heritage.

Scope

Make normalized and interoperable metadata of digital collections for a wider spread among different platforms.

Preconditions

- Cataloguing activity
- Contact between collection manager and developer

Success End Condition

- Librarian digital collections integrated into Internet Culturale; museum digital collections integrated into MuseiD-Italia; both indexed into Culturaitalia
- Content ingestion process into the portals becomes simpler and more efficient.

48

http://www.culturaitalia.it/opencms/opencms/attachments/museiditalia/profiles/mets/MuseiDItalia_METS_profile.html

49

http://www.culturaitalia.it/opencms/export/sites/culturaitalia/attachments/museiditalia/profiles/mets/metamets_empty.zip

50

http://www.culturaitalia.it/opencms/export/sites/culturaitalia/attachments/museiditalia/profiles/mets/MUSEIMETS_MANUALE_UTENTE_v_1_0.pdf

51

<http://www.culturaitalia.it/opencms/export/sites/culturaitalia/attachments/museiditalia/profiles/mets/MDIValidator.zip>

Fail End Protection

Disclaimer stating that the institution has clear legal rights over the collection.

Primary Actor

- AF
- ICCU

Other Actor(s)

- ICCD
- Digital collection developer

Trigger

Institution wants to catalogue its cultural heritage and to publish the digital collections online, within the main National portals.

Main Scenario

Once AF gets all the standards, it will proceed:

- 1) AF contacts ICCU
- 2) ICCU sends the information about standards for digitisation, cataloguing, licenses and the workflow for making the digital collections among the different National and European Systems interoperable
- 3) Institutions agree to share metadata and digital collections within an agreement where the licenses are indicated (generally we propose the rights statements of the Europeana Data Model⁵², which express the copyright status of a work, as well as information about how the content provider access and reuse objects)
- 4) Catalogue the books and other bibliographic materials in the Central/Local Catalogue
- 5) Digitize the books and preparing the metadata for the digital collections of the libraries and send them to ICCU for uploading into Mag Teca, the digital library integrated in Internet Culturale
- 6) Internet Culturale makes the digital collections interoperable via OAI-PMH to CulturalItalia; CulturalItalia uses the OAI-PMH protocol for distributing the metadata to Europeana and other portals. CulturalItalia also makes the data available through a SPARQL end point.
- 7) Catalogue the museum objects such as paintings, sculptures and fine furniture, following the ICCD standards

⁵² <http://pro.europeana.eu/share-your-data/rights-statement-guidelines/available-rights-statements>

- 8) Prepare the METS files of the museum's digital collections using the MuseiD-Italia application and uploading it to the MuseiD-Italia digital library. The system then sends the files via OAI-PMH protocol to CulturalItalia.
- 9) All the digital collections can be integrated into the AF website.
- 10) Metadata is available in the CulturalItalia OAI-PMH provider in DC, PICO (CulturalItalia application profile format); EDM (Europeana Data Model) and CIDOC-CRM. Data is also available through a SPARQL end point.

Extensions

- 1) Institution does not want to publish a description of its collection into CulturalItalia
- 2) Institution has unclear or no legal rights over the collection
- 3) Collection manager fails to establish contact with the collection developer
- 4) The provided description does not meet the basic criteria/guidelines

2.1.3.6. Working on 3D formats for archiving and on common metadata

Provided by: CNRS / 3D Huma-Num consortium

Contributor(s): *Stéphane Pouyllau / Adeline Joffres*

User Story

Today, the digital model has become indispensable for scientific restitution. However, faced with the development of 3D technology, it is now important to assist with the integration of these tools and support new uses that enables “des sciences humaines et sociales” (SHS) community to produce an exponential amount of numerical models.

In that context, 3D Huma-Num consortium is working on the virtual representation of missing environments, experimentation and safeguarding the industrial and technical heritage; the acquisition and the spatial and temporal modelling; and the development of simulation tools in architecture and heritage. Created in 2014, Huma-Num 3D consortium now comprises nine French research centres. To fulfil its mission, this consortium focuses on choosing and developing open source tools.

Currently, numerous 3D file formats are used to represent 3D objects. Some formats are linked with hardware (such as 3D scanners) and are not open format. These formats are propriety, understandable only by specific software and are therefore not suitable for long-term preservation.

3D Huma-Num consortium participates in the selection of different formats to prepare 3D data for long-term preservation. The second stage is to develop tools to create, control and manipulate these selected formats.

There is a need in the 3D users' community to define a common format of metadata. The Huma-Num consortium participates in the definition and evolution of CARARE 2 metadata schema developed by CARERE (Connecting ARchaeology and ARchitecture in European; www.carare.eu). The researchers and engineers of this consortium enrich it to work more closely to their needs, which implies a pooling of work even though they do not necessarily handle the same objects in 3D. This is all done within the common core of CARARE. Another important part of this 3D consortium work is to stabilize reference vocabularies in order to build a consistent description of 3D objects in relation with spatialized data and knowledge.

In order to achieve these goals, the 3D consortium will organize several workshops and summer schools for the community involved.

Goal

Concerning the archiving of 3D objects, the aim of the 3D/Huma-Num consortium is to agree on a "pivot" format with "acceptable" information losses. After size(s) is selected, the development of tools for handling / verification of formats will be determined.

As far as 3D object metadata is concerned, the consortium aims to build an interoperable metadata format for 3D objects - from parameters for the size being defined in the CARARE project, and also used by Europeana - and make it usable for the long-term preservation process.

Scope

Mainly researchers in archaeology, history, architecture and geography; but eventually all users of 3D objects (museology?).

Preconditions

- To have in mind the multitude of 3D formats
- Community meetings
- International cooperation

Success End Condition

To find an acceptable and suitable format for all 3D models over pursued goals.

Fail End Protection

Not to find a common software or open format, independent of software vendors.

Primary Actor

3D consortium of Huma-Num (CNRS).

Other Actor(s)

Huma-Num and all the consortium's partners, researchers in archaeology, history, geography and architecture. For archiving : CINES (Informatic Centre of Higher Education – France).

Trigger

Need for long term and content preservation.

Main Scenario

- 1) National coordination and coordination within research communities
- 2) International coordination with European structures and others (MITI...)
- 3) Propose standards for metadata description of 3D objects
- 4) Develop tools to check 3D files expressed in a normalized format

Extensions

Survival Kit to archive 3D objects from SSH and CH Guidelines, good practices guides' edition

Participate in evolution of CARARE format for Europeana

2.1.3.7. Collection holding institution wants to have a standard that makes cross search through cultural periods across Europe possible

Provided by: KNAW / DANS

Contributor(s): *Emilie Kraaikamp / Hella Hollander*

User Story

A collection holding institution (CoHI) has data with temporal metadata on cultural periods and wants to display this in a meaningful way, ensuring that the temporal metadata can be successfully interpreted and is standardised. Period concepts are entangled with space (they are different from place to place, as they are from scholar to scholar) therefore a shared reference point is used, PeriodO, <http://perio.do/>. PeriodO provides a gazetteer of scholarly definitions of historical, art-historical, and archaeological periods. Definitions about the geo-spatial area and absolute dates are assigned to period terms, which makes cross-referencing through cultural periods across Europe possible. Furthermore, when used by other CoHIs and aggregators, all data with the same time-range can be connect-

ed through these definitions. Using PeriodO for temporal metadata on cultural periods contributes to a more effective workflow for researchers searching and finding data on cultural periods.

Temporal metadata on cultural periods can relate to various disciplines, but is often used for archaeological data. Data from archaeological research is increasingly digital-born and can consist of databases, tabular data, location data, maps and digital photographs. Data from archaeological excavations are often gathered using digital measuring and mapping equipment. During research, data are automatically and/or manually entered into tables, spreadsheets and databases. Digital materials receive ideally a structured metadata for each individual file. When data of archaeological excavations is stored at the CoHI the researcher also provides metadata including the information about the cultural periods.

Step one: the researcher maps his / her temporal metadata to PeriodO or uses the current classification of the collection holding institution.

Step two: the collection holding institution provides international aggregators like ARIADNE and EUROPEANA with the temporal metadata referring to PeriodO by providing its own temporal classification to PeriodO.

Step three: international aggregators like ARIADNE and EUROPEANA can disseminate the temporal metadata in a meaningful and uniform way.

Step four: researchers can find data by cross-referencing cultural periods across Europe.

Goal

A CoHI can link its temporal metadata on cultural periods with data from other collection holding institutions across Europe, enabling researchers to search across data on cultural periods throughout Europe.

Scope

Make temporal metadata of collection holdings available using a predefined standard.

Preconditions

- Guidelines for the collection holding institution on how to translate their temporal metadata to the gazetteer of period definitions for linking and visualizing data (PeriodO).
- Guidelines for the researcher on how to provide temporal metadata in a standardised way.
- Collection holding institution has organized metadata.

Success End Condition

Collection holding institutions within the fields of history, language studies and related fields across the digital Humanities all use PeriodO for the exchange of their temporal metadata.

Primary Actor

Collection Holding Institution (that archives data and temporal metadata and provides international aggregators with its temporal metadata)

Other Actor(s)

- International aggregator(s) (that can harvest and disseminate temporal metadata)
- Researcher (who wants to reuse data with temporal metadata)

Trigger

Collection holding institution has data with temporal metadata and wants to display this in a meaningful way.

Main Scenario

- 1) The researcher maps his / her temporal metadata to PeriodO or uses the current classification of the collection holding institution.
- 2) The collection holding institution provides international aggregators like ARIADNE and EUROPEANA with the temporal metadata referring to PeriodO by providing its own temporal classification to PeriodO.
- 3) International aggregators like ARIADNE and EUROPEANA can disseminate the temporal metadata in a meaningful and uniform way.
- 4) Researchers can find data by searching across cultural periods across Europe.

Extensions

- 2) CoHI is not willing to provide its own temporal classification to PeriodO

2.1.4. Social Sciences

2.1.4.1. Platform for inventorying and archiving field surveys in political science (political sociology)

Provided by: CNRS / archiPolis Huma-Num consortium

Contributor(s): *Stéphane Pouyllau / Adeline Joffres*

User Story

The archiPolis Huma-Num consortium was labelled as such in 2012. The main mission of this consortium is to develop a collective strategy for inventorying, collecting, preserving - through digitisation in the case of older surveys which have not been entered in digital format - and defining common metadata for field surveys conducted by political scientists, sociologists and other social scientists interested in the political subject. This is to make these investigations intelligible through a documentation and commissioning-consistent context. The objective is indeed to avoid depletion or even abuse of the research work, that could lead to the conservation and availability of the original research data out of context.

The consortium is also on a mission to lead the discussion on the objectives, risks and limitations of archiving, in order to convince as many teams and colleagues as possible to participate. It will have appropriate structures to organize the collaborative production and distribution of good practice guidelines to encourage and facilitate archiving of investigations. This is to extend and specify existing initiatives at national and international level and adapt them to specific research practices to the relevant scientific communities.

Finally, the consortium sets up awareness for the conservation of past, present and future surveys (pre-classification, preventive conservation, format natively digital documents, etc.) among the scientific community at large, that is, beyond teams and researchers from professional bodies, training institutions (such as graduate schools), but also research funding bodies.

In 2014, a partnership with the French BeQuali portal was finalized in order to share the whole data collected in a unique digital platform based on the inventory grid (metadata) that archiPolis structured and worked on. This work reveals a huge pedagogical discourse to the researchers, the students, and academic archivists and librarians.

Goal

- To agree on a single and normalized metadata formulary/ inventory grid/layout
- Convincing and training “users” to join their data of field surveys for the purposes of sharing, preserving, archiving and reusing it

Scope

Political scientists, archivist and digital curators

Preconditions

- Community meetings
- International cooperation

Primary Actor

archiPolis Huma-Num consortium (<http://archipolis.hypotheses.org/> ; <http://www.humanum.fr/consortiums#ARCHIPOLIS>)

Other Actor(s)

Huma-Num and all the archiPolis consortium's partners; researchers in political science as well as ethnology, sociology, and anthropology.

Trigger

- Work in progress since 2012
- Partnership with BeQuali consortium (centre for social-political data)

Main Scenario

- 1) National coordination and coordination within research communities
- 2) Build a common approach to describe qualitative surveys with metadata
- 3) Create a catalogue of interesting surveys
- 4) Make the DDI (international standard for describing statistical and social science data) format more adequate for these types of surveys

Extensions

Share the catalogue of surveys (e.g. SHARE project) at European level

2.1.4.2. A researcher wants to share and use social science data in an effective way through a collection holding institution that has implemented the DDI standard

Provided by: KNAW / DANS / ARIADNE
Contributor(s): *Emilie Kraaikamp*

User Story

Researchers in Social Sciences often create data in statistical data programmes such as SPSS (Statistical Package for the Social Sciences) and STATA (Data Analysis and Statistical Software). A research project is usually comprised of one or more data files accompanied by a codebook and supplementary documentation such as a questionnaire. These latter files are typically PDF files.

A researcher in Social Sciences wants to share and use data in an effective way. Therefore she / he uses the services of a collection holding institution that implemented the DDI standard (DDI Lifecycle 3.2 <http://www.ddialliance.org/Specification/>). By using the DDI standard, data from this institution is comprehensively described in a structured manner enabling effective discovery, analysis and sharing. To offer this to the researcher, the collection holding institution has the following in place for social science data:

- Standardised research metadata according to the DDI standard.
- Standardised, well-documented data according to the DDI standard.
- Structured and interactive DDI codebooks, enabling researchers to navigate through a collection.
- Data catalogues based on the DDI standard for searching at both the study and variable levels to enable researchers to discover data of interest.

If data is not structured and presented in a standardised way, using and sharing data is more difficult. When the DDI standard is used for its documentation and metadata, it can be found more easily, interpreted, analysed and combined, because it is structured and described in the same way. Furthermore, because a researcher can search through the content of different studies directly, discovery and sharing of data can be done much more easily and faster.

Goal

A researcher can share and use Social Sciences data effectively via a collection holding institution that has implemented the DDI standard.

Scope

Researchers in Social Sciences documenting and archiving their data, and a collection holding institution (or multiple CoHIs), archiving and disseminating their data according to the DDI standard.

Preconditions

- The Social Sciences researcher is willing and able to provide documentation according to the DDI standard
- The content holding institution is willing and able to disseminate DDI documentation for Social Sciences data
- Guidelines for the Social Sciences researcher on providing documentation according to the DDI standard

- Guidelines for the content holding institutions on disseminating the documentation according to the DDI standard

Success End Condition

Through a content holding institution the researcher can share and use Social Sciences data described according to the DDI standard.

Fail End Protection

Collection holding institution only accepts Social Sciences data described according to the DDI format but this cannot be provided.

A Social Sciences researcher has his / her data described according to the DDI standard but the collection holding institute cannot disseminate this description.

Primary Actor

The content holding institution: archiving and disseminating the data

Other Actor(s)

The researcher: producing and using Social Sciences data

Trigger

A researcher in the Social Sciences wishes to share and use Social Sciences data in an effective way.

Main Scenario

- 1) The collection holding institution sets requirements regarding archiving Social Sciences data; the use of the DDI standard is implemented for Social Sciences data.
- 2) A Social Sciences researcher creates a comprehensive description of his / her data using the DDI standard.
- 3) A Social Sciences researcher archives his / her data at a collection holding institutions and provides the archive with the description according to the DDI standard.
- 4) The collection holding institution archives and disseminates the data and the data description according to the DDI standard.
- 5) A Social Sciences researcher can use the archived data and the DDI description.

Extensions

1a) The collection holding institution is not yet able to have the DDI standard implemented.

1a1) The collection holding institution works on the implementation first.

- 2a) The researcher is not able to create the required descriptions
- 2a1) The researcher looks into other ways to create the required description; the collection holding institution assists in creating the required description
- 3a) The collection holding institution finds the description insufficient.
- 3a1) The researcher is asked to change the description.

2.2. Requirements

The following table shows the most obvious requirements extracted from the use cases. The authors have summarized the requirements in a short description. Please note that this list should not be considered as a complete overview of the user requirements of the research communities in regard to standardization.

The identified requirements relate to different parts of the data lifecycle, i.e.:

- Standards in ingest phase
- Source material
- Research data
- Output standards
- Reference standards
- Enrichment standards
- Data exchange standards

| UC # | Use Case | Actor | Extracted Requirement(s) |
|------|--|---------------------------------------|---|
| 1 | WW1 Historian and the trans-national/trans-institutional question of the development of the railways | Modern Historical Researcher | See user requirements listed in main scenario of user story. |
| 2 | Collection Holding Institution publishes data on the EHRI portal | Collection Holding Institution (CoHI) | Enables publication of collection and/or organizational data in a standardized and shareable format. |
| 3 | Holocaust Researcher investigates person information and networks | Holocaust researcher | Enable historian/Holocaust researcher with a standardized information model and tools to analyze person and |

| | | | |
|---|---|--|---|
| | | | network information. |
| 4 | Historian wants to publish his research data and make it reusable with the DARIAH-DE repository | Researcher in History | Need for an easy-to-use metadata format for describing research data. User-friendly and intuitive to use even for researchers who are not familiar with metadata and standards. |
| 5 | Historian wants to track the dissemination of a given author's works during the Medieval and Early Modern period | Researchers working in the history disciplines, acting as data consumer | <p>Enable sharing and accessing information stored in various repositories, based on different standards; Foster standardization of reference tools for the disambiguation of names of persons and places documents (i.e.: manuscripts shelf marks); titles of texts/works</p> <p>Develop a LOD web of authors, works, documents and related information (i.e.: available information about origins and provenances of the documents)</p> <p>Develop tools to perform searches across multiple scholarly resources</p> <p>Develop tools to display the results on a map and / or timeline</p> |
| 6 | Natural Language Processing Expert wants to test her tool for semantic annotation on an available digital edition of historical texts | LT expert (Typology: a researcher that needs standards in order to achieve his / her research) | Tools need to be adaptable so as to be able to read de facto textual standards such as TEI. TEI editions need to be well formatted and documented in order to allow for tools to work correctly. |

| | | | |
|----|--|---|---|
| 7 | Create annotated digital edition | LRS researchers (LRSSRs) | Need for standards/best practices in order to support the whole process of creating and publishing a digital edition. This should cover all points in the main success scenario of the use case, especially tokenization and lemmatization of texts, performing NER (named entity recognition), annotation, publication of edition. |
| 8 | Build a corpus of linguistic data for analysis | Researcher from the domain of language-related studies (LRS) | There is a lack of recommendations for tools and the usage of standards in parts of this use case, especially for the annotation and for the visualization. Researchers need standards and best practice examples in all steps of the main success scenario. For non-experienced researchers there is the need for easy-to-use applications and how-to manuals. |
| 9 | Interoperability in literature using the TEI | Huma-Num CAHIER Consortium and its partners ⁵³ | TEI needs to be implemented; TEI skills; Some background in metadata description. |
| 10 | Linking original text in literature studies to commentary, translations and external sources | Researchers that need standards in order to express his / her research (e.g. commentary) or need formats and a standardised way to cre- | Standardized (Suggested) formats for referencing/linking between TEIP5 files, and examples of linking to external resources from TEIP5 files. |

⁵³ See <http://cahier.hypotheses.org/parte-nares>.

| | | | |
|----|--|---|--|
| | | ate and annotate links | |
| 11 | Sustainability and improved viewing of Assyrian text resources | Researcher from the field of Assyriology (who wants to share his/her data and, at the same time, wants to ensure that the data is kept sustainable) | Enabling sharing and producing otherwise inaccessible research data, by using interoperable and sustainable formats. |
| 12 | Conservation scientist wants to publish information about experimental conditions for Raman analysis of wall painting fragments and report in particular proper experimental measurement conditions for safely detecting and identifying certain types of pigments | Conservation scientist | A procedure is needed to enable the user of the digital library to inform the library authors about his / her findings (documents, graphs, images). A following procedure is needed that allows the authors to include the user findings to the digital library. |
| 13 | Researcher using lasers in conservation/restoration identifies the necessity of standardized reports of the laser application conditions and the evaluation of the obtained results. | Researchers in the field of laser cleaning in conservation/restoration of CH | Guidelines to document the laser cleaning treatments employed in the field of Cultural Heritage in a standardized and complete way are needed. The ability to publish standardized reports in IPERION CH/PARTHENOS portals is needed. |
| 14 | A dataset for the products used in conservation treatments in order to share information about their application parameters, their effective- | - Researchers who needs information about a treatment in order to know | Develop a standardized/common report, which includes the most relevant information necessary for the evaluation of conserva- |

| | | | |
|----|--|--|--|
| | ness and their durability in time, related to the type of material and its state of conservation. | <p>the state of art;</p> <ul style="list-style-type: none"> - Researchers who want to share their results in order to amplify the result of a specific research; - Researchers who want to compare their results; - SME working in the development of products for conservation that wants to test its products and looking for a good practice guidelines. | tion treatments. |
| 15 | DYAS contact CoHI and asks to integrate metadata of its digital collection into the “Humanities Resource Registries” portal | DYAS/AA | Guidelines for CoHIs and individual researchers on metadata standardization. |
| 16 | Private Foundation wants to publish the digital collections of its library and museum in the online Public Access Catalogue, Internet Culturale and CulturalItalia | <ul style="list-style-type: none"> - Private Foundation (AF) - ICCU | <p>Standards for digital collections of bibliographic and museums resources.</p> <p>Creating, extending, mapping multilingual thesauri and publishing them in SKOS format.</p> |
| 17 | Working on 3D formats for archiving and on common metadata | 3D consortium of Huma-Num (CNRS) | Enabling creation and manipulation of 3D objects. |
| 18 | Collection holding institution wants to have a standard that | Collection Holding Institution | Guidelines are needed for the collection holding institution |

| | | | |
|----|--|--|--|
| | makes cross search through cultural periods across Europe possible. | (that archives data and temporal metadata and provides international aggregators with its temporal metadata) | on how to translate their temporal metadata to the gazetteer of period definitions for linking and visualizing data called PeriodO http://perio.do/ . |
| 19 | Platform for inventorying and archiving field surveys in political science (political sociology) | ArchiPolis Humana-Num consortium ⁵⁴ | DDI standard implemented in tools and used for survey data. Knowledge in DDI format. |
| 20 | A researcher wants to share and use social science data in an effective way through a collection holding institution that has implemented the DDI standard | The collection holding institution: archiving and disseminating the data | A collection holding institutions implements DDI to enable researchers from the Social Sciences to share their data effectively. |

⁵⁴ <http://archipolis.hypotheses.org/>; <http://www.huma-num.fr/consortiums#ARCHIPOLI>

3. Interoperability, services and tools requirements

Main authors: Emiliano Degl'Innocenti (CNR-OVI, formerly SISMELE), with support by Roberta Giacomi and Veronica Boarotto (both SISMELE)

3.0. Objectives

This chapter presents the findings of the activities carried on within Task 2.3. The main objective of Task 2.3 is to gather, organize and prioritize the user requirements for the interoperability of services and tools as expressed by the research communities involved in PARTHENOS, structuring them into use cases. The research communities we focused on are listed in section 0.3 of this document (“User Communities in PARTHENOS”); the methodology we followed and the projects that were considered are described in section 0.4 (“Methods”). In order to provide concrete input for the implementation of the PARTHENOS technological framework enabling interoperability, the assessment activity of Task 2.3 has been carried out in close collaboration with the technical work-packages (WP5 and WP6), adopting the PARTHENOS vision as background and aiming at placing the requirements and use cases into the vision, in order to refine the technical architecture and make it concrete.

In the following pages, the use cases and requirements are presented in narrative form (using the Cockburn Simplified Language) in the main section of this chapter, and in tabular form in the two Technical Annexes (Annex A: Use Cases, Annex B: Requirements). The reasons for this choice are explained in the “Method” section of this chapter. In the last section we present conclusions based on the material gathered and analysed, and describe further steps towards the redaction of D2.4.

3.1. Method

In addition to the general methodology described in Section 0.3, in order to foster the process of community involvement in the activities of task 2.3, a number of different scholarly networks within the selected target groups were contacted, including but not limited to:

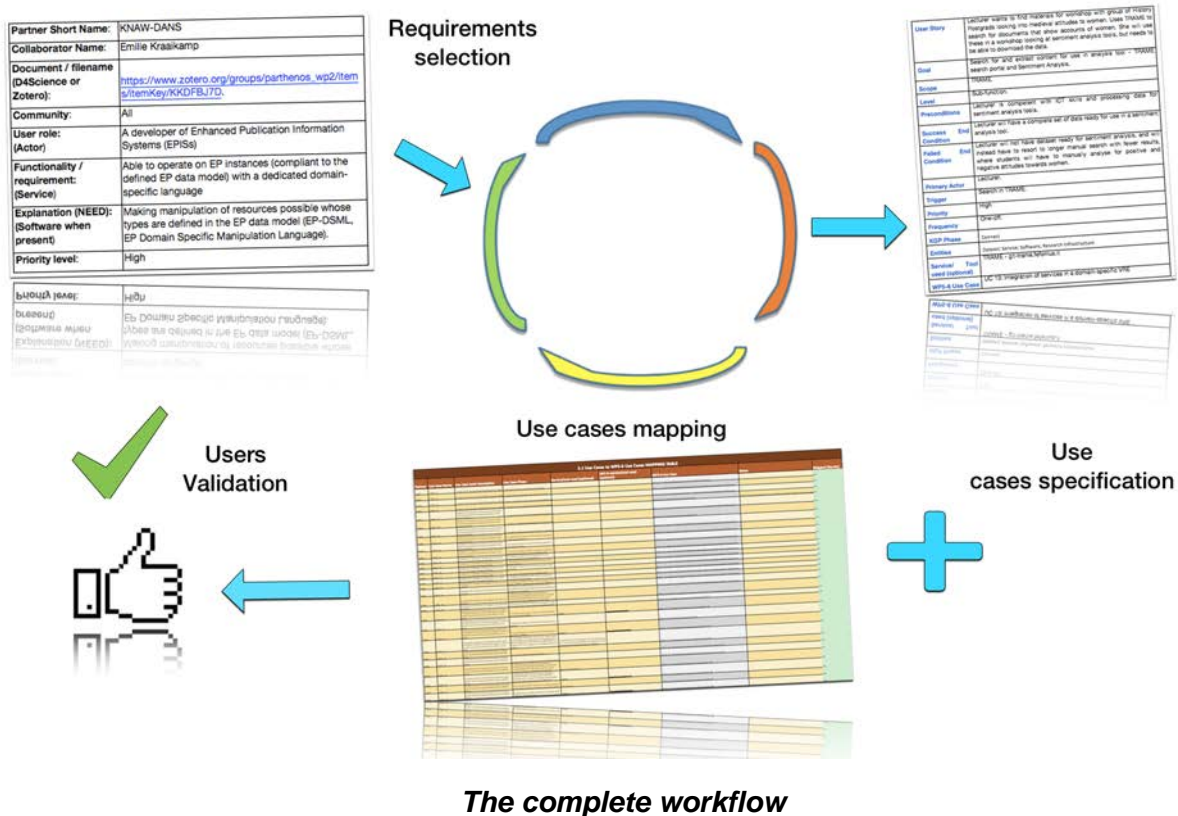
- CARMEN (Co-operative for the Advancement of Research through a Medieval European Network)⁵⁵

⁵⁵ www.carmen-medieval.net, last visited 09/22/2016

- Medievalist Sources (DARIAH-ERIC Working Group)⁵⁶
- Digital Medievalist⁵⁷

To ensure an adequate level of completeness of the requirements expressed in this chapter, relevant EU research infrastructures and projects in various research fields - such as Minerva, Athena, AthenPlus, CulturalItalia, Europeana, CENDARI, FlareNet, TRAME, DASISH, OpenAIREplus, EHRI, MUSE, COST IS10005 and ARIADNE, as well as the two ERICs in the DH field (DARIAH and CLARIN) - have been involved, both in a direct (interviews) or indirect (documentation review etc.) manner.

To facilitate the process of transition from user expectations into actual architectural design, we followed a flexible workflow, starting from the available technical documentation to extract the relevant requirements and specify the related use cases.



In a further step, the domain use cases we gathered (with the only exception of KNAW-DANS and KNAW-NIOD, for which we used requirements) were mapped against a set of more abstract use cases provided by WP5 and WP6, implementing the general-level functionalities of the PARTHENOS infrastructure, including: registering and accessing entities, setting up and using domain specific VREs, aggregating and exporting metadata from and to other RIs, curating resources. The mapping phase was aimed at verifying that the func-

⁵⁶ <http://www.medievalistsources.eu>, last visited 01/06/2016

⁵⁷ <http://www.digitalmedievalist.org>, last visited 09/22/2016

tionalities implemented by the PARTHENOS architecture actually covered the needs expressed by the users.

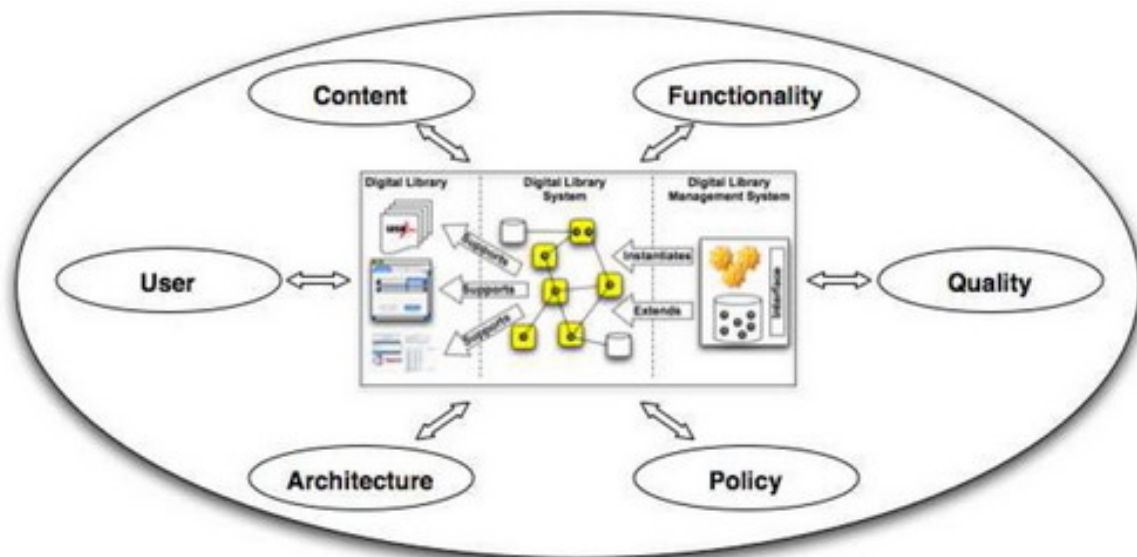
3.2. A working definition of interoperability

It is commonly understood that it is hard to find a uniform and generally accepted definition of “interoperability”. The various different attempts to define this concept in different domains and from different perspectives have resulted in a wide spectrum of definitions, each focussing on technological aspects or legal and policy contexts or content related issues, etc. For the purposes of this document we adopted the ISO/IEC 2382-2001 standard definition of interoperability as the “capability to communicate, execute program[me]s, or transfer data among various functional units in a manner that requires the user to have little or no knowledge of the unique characteristics of those units”.

3.3. PARTHENOS reference model

In collaboration with WP5 and WP6, we agreed on the necessity to use a reference model to gather information from researchers and available documentation in order to extract and present the user requirements in a form that could be effectively used by the technical teams to feed the PARTHENOS development agenda. As a basis for the development of a PARTHENOS Reference Model for Interoperability as “an abstract framework for understanding significant relationships between the entities of some universe [Digital Humanities in our case], and for the development of consistent standards and/or specifications [i.e.: interoperability requirements] supporting that universe”⁵⁸, we initially used the DELOS Reference Model.

⁵⁸ The Digital Library Reference Model website, https://workinggroups.wiki.dlorg.eu/index.php/Interoperability_Concepts, last visited on 01/06/2016



Delos reference model

The DELOS reference model was developed in 2007 by the DELOS Network of Excellence “as a necessary step towards a more systematic approach to the research on digital libraries”. Given the age of the model, we needed to update and adapt it to represent the more complex and multifaceted landscape of the Digital Humanities domain. To do so, and to organise all the pieces into a single framework to express the user requirements, we initially adopted a draft template (we’ll refer to it using the name “PARTHENOS reference model”) resulting from a combination of the Cockburn Simplified Language as a formal way of expressing the information we gathered and the PARTHENOS entities vocabulary⁵⁹ as a reference list of terms. The PARTHENOS entities (see where referenced in sections 3.6.1 “Use Cases” and 3.6.2 “Requirements”), are used to link the needs expressed by researchers to the entities involved in the PARTHENOS vision.⁶⁰ Furthermore, the elements and activities carried out under these premises are described in the next sections.

3.4. Use cases modelling and requirements extraction

Martin Fowler (2004) states “there is no standard way to write the content of a use case, and different formats work well in different cases” (Fowler (2004), together with the technical team at ISTI-CNR (WP6). Therefore we agreed to express both the use cases and the related requirements using the Cockburn Simplified Language (CSL) (Cockburn 2000),

⁵⁹ See the documents PARTHENOSEntities_CategoricalDescription_V1.11.docx

⁶⁰ For additional information, see the explanation of the elements included in the requirements, described in the below table.

which offers a good level of formalization and is considerably simpler to be handled by non-technicians than Unified Modelling Language (UML).

The requirements extraction process was based on the review of available documentation produced by ESFRI projects and other relevant initiatives (including scholarly networks, and e-infrastructures). The review looked for information about scientific needs driven by research questions, and related tools and services used in the different domains involved in PARTHENOS to support these needs in a digital environment (i.e. a VRE). In total, fifty-five documents have been gathered from fifteen projects or research infrastructures: Minerva, Athena, AthenPlus, CulturalItalia, Europeana, CENDARI, FlareNet, TRAME, DASISH, OpenAIREplus, EHRI, MUSE, COST IS10005 and ARIADNE. The two ERICs in the DH sector, DARIAH and CLARIN, were also involved in the process.

We reviewed the available documentation in order to gather relevant information on actual research practices to be supported by the PARTHENOS digital infrastructure, compile the use cases and extract the related requirements.

In this chapter, a use case is “a written description of how users will perform tasks on [a given] website [or resource]. It outlines, from a user’s point of view, a system’s behaviour as it responds to a request. Each use case is represented as a sequence of simple steps, beginning with a user’s goal and ending when that goal is fulfilled.”⁶¹ The various goals expressed in the use cases presented in this chapter are used to establish a list of features to be implemented by WP5 and WP6 within the PARTHENOS infrastructure. In a subsequent phase, together with the technical teams, we further negotiated which functions will become requirements and will be actually implemented, from the needs expressed by the users. It was agreed that the first release of the PARTHENOS infrastructure will provide cross-domain services (entities registration and access, VREs creation and use, resources curation, metadata aggregation and export) over the entities and resources in the PARTHENOS Registry, addressing more specialized needs in a second phase of development.

The level of detail provided by each use case may vary, according to the complexity of the goal(s) the user wants to achieve; a typical use case, however, describes in an easy-to-understand narrative form:

- Who is using the website / service / tool
- What the user wants to do
- The steps the user takes to accomplish a particular task

⁶¹ <http://www.usability.gov/how-to-and-tools/methods/use-cases.html>, last visited on 09/22/2016

- How the website / service / tool should respond to an action

Please note that the presented use cases do not describe any implementation specific language, nor provide details about the user interfaces or screens (Kenworthy 1997). For an overview of the used elements see the Introduction, section 0.5, “Methods of presenting user requirements”.

From the information provided by the use cases, the task 2.3 team members extracted a list of required functions and characteristics to be considered within the implementation agenda of the PARTHENOS interoperability framework (WP5 and WP6). For the scope of this document we focussed only on user and functional requirements describing:

- user expectations;
- how users will interact with the PARTHENOS infrastructure;
- how users will use the services and tools described in the use cases;
- how a given service and/or tool should behave.

A detailed overview of the elements included in the requirements is presented in the following table:

| Field Name | Explanation |
|--------------------------------------|---|
| Partner Short Name | Indicates the project partner describing the requirement |
| Collaborator Name | Indicates the individual researcher describing the requirement |
| Document / filename | Source document for the requirement (i.e.: reference to the files on D4Science and/or Zotero) |
| Community | Community expressing the requirement: History, Language related studies, Archaeology, Heritage & applied disciplines, and Social Sciences |
| Related Domain Use Case, Function ID | Reference to the use case label in the use cases table |
| User role | Indicates in which role the user has the requirement |
| Functionality / requirement | Short, unique name of the requirement |
| Explanation | Short description of the requirement |
| Priority level | High, medium, low |

| | |
|--|--|
| Macro functionality | Macro-category the requirement belongs to |
| Possible required functions | Dependencies related to other functions |
| Possible involved toolkits or components | Mapping with tools and components already existing in or known by the project consortium |
| Entities | Elements registered in the PARTHENOS entities registry, such as: RIs, Datasets, Actors, Services, Software and Knowledge Generation Processes |
| Knowledge Generation Phase | Step in the workflow established to support a given research activity, towards the production of a specific dataset, i.e.: Collect, Connect, Interpret and Present data |
| Services and tools involved | Specific (existing) tools and services needed to perform a given research task, to be integrated into the final version of the PARTHENOS research infrastructure |
| Related Functional Use Case | The functional Use Case implementing the requested functionality, i.e.: entities registration and access, VREs creation and use, resources curation, metadata aggregation and export |

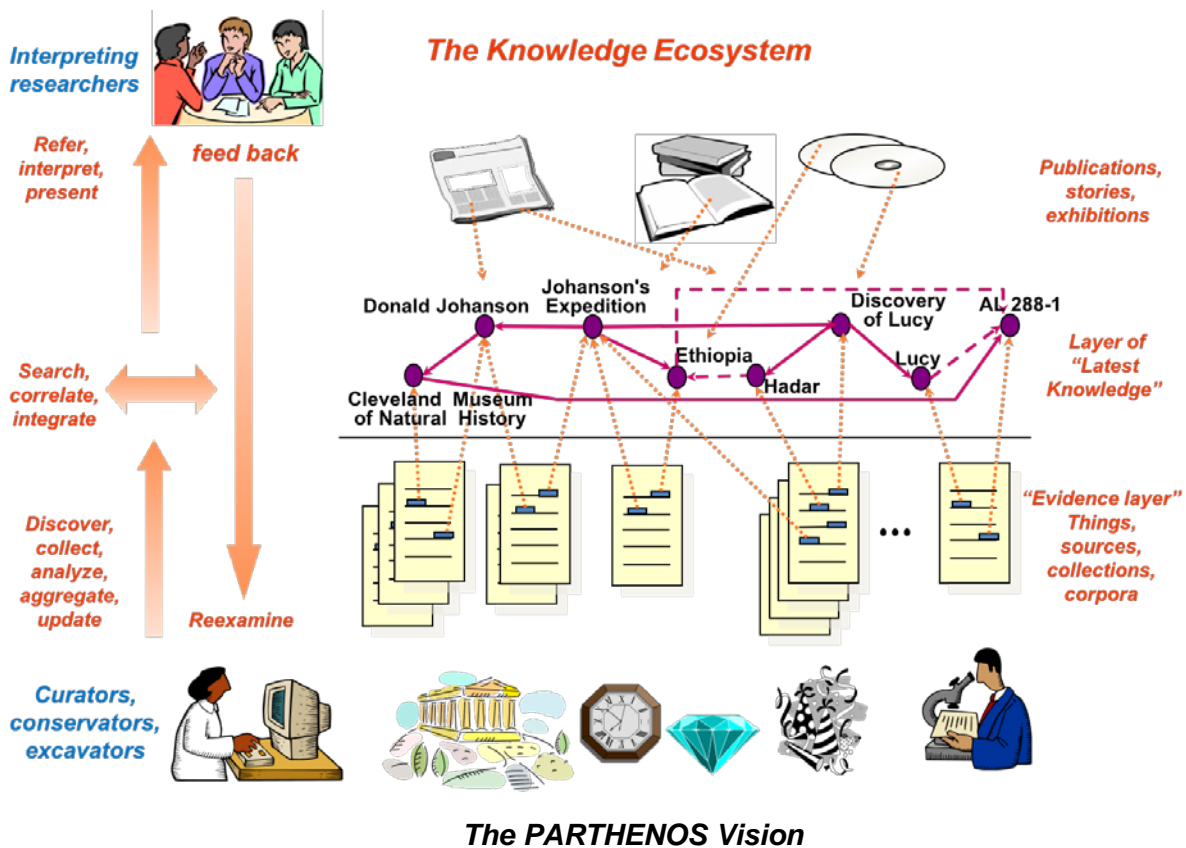
As already mentioned, the task 2.3 team members decided to present the use cases and the related requirements both in narrative and tabular form, following the methodology expressed in the PARTHENOS reference model section. The main difference between the two styles is represented by their different destination audience: the narrative form is adequate for the interaction with the researchers and better describes the process and its context (i.e.: the “Knowledge generation process” element in the PARTHENOS Reference model⁶²), while the tabular form is better suited for the consideration of the technical teams, in order to clearly shape their development agenda.

3.5. From the requirements to the architectural design

After the collection of use cases and requirements, we started a mapping activity to integrate the results of this collaborative work and making it fit to the PARTHENOS architecture. The work comprised different sources of data coming from different contexts and researchers, each with particular backgrounds, knowledge and disciplinary concerns.

⁶² See above and the PARTHENOSEntities_CategoricalDescription_V1.11.docx document.

Mapping the domain use cases provided by partners against the functional ones and distilled by members of WP5 and WP6, we made sure that the cross-domain functionalities requested by the researchers were fully covered by the PARTHENOS architecture (cf. Technical Annex C). Furthermore, according to the PARTHENOS vision ‘scientific data are components of a dynamic process, aimed at generating, evolving and consolidating knowledge’, and thus they ‘cannot be divorced and abstracted from the processes where they belong; and the researchers that execute those processes in their daily activity’. We therefore also referenced the phases of the Knowledge Generation Process (i.e.: *collect*, *connect*, *interpret* and *present*) in the existing domain use cases and in the case of KNAW-DANS and KNAW-NIOD, in the requirements.



To promote trust and to facilitate the adoption of the PARTHENOS infrastructure by the participating research communities, its architecture should support the provision of services supporting actual research practices (i.e.: actual knowledge generation processes), ensuring the scientific reliability of the contents and representing their provenance. All the relevant information related to the PARTHENOS contents will be collected, identified, described and connected in the registry that will also provide cross-domain services on registered entities.

To match the above vision and provide input for the registry establishment and the PARTHENOS infrastructure implementation, the requirements and use cases described in sections 3.6.1 and 3.6.2 have been mapped against a set of 21 more abstract functional use cases, implementing the following cross-domain functionalities:

- Entities registration
- Registered entities access
- Creation of domain specific VREs
- Use of domain specific VREs
- Metadata aggregation and export
- Resources curation

| ENTITIES REGISTRATION | |
|--|--|
| Manual registration entities in the PARTHENOS registry | <ul style="list-style-type: none"> ✓ People ✓ Services ✓ Data ✓ Metadata ✓ Software ✓ Research infrastructures |
| | ✓ Web interface |
| FUNCTIONAL USE CASES | |
| USE CASE 1: Manual registration of an entity in the PARTHENOS registry | |
| USE CASE 15: A research infrastructure joins PARTHENOS and integrates its registry | |

| REGISTERED ENTITIES ACCESS | |
|--|--|
| Retrieval / access entities registered in the PARTHENOS registry and resources in the PARTHENOS content cloud | <ul style="list-style-type: none"> ✓ People ✓ Services ✓ Data ✓ Metadata ✓ Software ✓ Research infrastructures |
| | ✓ Web interface |
| FUNCTIONAL USE CASES | |
| USE CASE 2: Search and browse the PARTHENOS registry | |
| USE CASE 3: Search and browse the PARTHENOS content cloud across several research infrastructures | |
| USE CASE 4: Retrieval/access metadata about an entity of the PARTHENOS registry or a resource in the PARTHENOS content cloud | |

USE CASE 5: Retrieval/access of resources from the PARTHENOS content cloud

| CREATION OF DOMAIN SPECIFIC VREs | |
|--|--|
| Set up, user authentication, services integration in domain specific Virtual Research Environments | <ul style="list-style-type: none"> ✓ People ✓ Services ✓ Data ✓ Metadata ✓ Software ✓ Research infrastructures |
| | ✓ Web interface |
| FUNCTIONAL USE CASES | |
| USE CASE 12: Set-up of a domain-specific VRE | |
| USE CASE 13: Integration of services in a domain-specific VRE | |
| USE CASE 7: VRE authentication and authorization | |

| USE OF DOMAIN SPECIFIC VREs | |
|--|--|
| Entities referencing, files deposition, resources sharing, dataset processing and results presentation | <ul style="list-style-type: none"> ✓ People ✓ Services ✓ Data ✓ Metadata ✓ Software ✓ Research infrastructures |
| | ✓ Web interface |
| FUNCTIONAL USE CASES | |
| USE CASE 6: Reference entities of the PARTHENOS registry in VRE posts | |
| USE CASE 10: Deposition | |
| USE CASE 11: Private and public sharing of resources deposited in the VRE workspace | |
| USE CASE 14: Process a dataset and publish results | |

| METADATA AGGREGATION AND EXPORT | |
|--|--|
| Entities referencing, files deposition, resources sharing, dataset processing and results presentation | <ul style="list-style-type: none"> ✓ Metadata ✓ Research infrastructures |
| | ✓ Web interface |
| FUNCTIONAL USE CASES | |
| USE CASE 8: Aggregate resource metadata from research infrastructures into the PARTHENOS content cloud | |

USE CASE 9: Export metadata via standard protocols

| RESOURCES CURATION | |
|--|--|
| Entities referencing, files deposition, resources sharing, dataset processing and results presentation | ✓ People ✓ Services ✓ Data ✓ Metadata ✓ Research infrastructures |
| | ✓ Web interface |
| FUNCTIONAL USE CASES | |
| USE CASE 16: Subject coverage | |
| USE CASE 17: Invite new content providers | |
| USE CASE 18: Invite curation | |
| USE CASE 20: Quality control of services (de-duplication) | |
| USE CASE 21: Quality control of services (gazetteer) | |

3.6. Requirements for interoperability: Use Cases

3.6.1. Use cases from Archaeology, Heritage and applied disciplines

3.6.1.1. AR_01

| | |
|---|--|
| Provided by: PIN Contributor(s): Paola Ronzino | |
| User Story | An archaeologist wants to search/browse available data. |
| Goal | Get a list of institutions holding archaeological information concerning excavations, objects, periods. |
| Scope | <ul style="list-style-type: none"> • Access to archaeological repository. • Access to pre-existing catalogues. • Access to online data collections. |
| Preconditions | The user has general ICT skills. |
| Success End Condition | The user can ask the holding institutions instructions on how to get access to their data. |
| Failed End Condition | No relevant datasets found. |
| Primary Actor | Archaeologist in the role of data consumer. |
| Trigger | Navigate the portal directly or indirectly by using a search engine. |
| KGP Phase | Connect |
| Entities | Dataset; Service; Actor |
| Service/ Tool used (optional) | |

| | |
|-----------------------|--|
| WP5-6 Use Case | UC ACCESS_01: Search and browse the Parthenos registry |
|-----------------------|--|

3.6.1.2. AR_02

| | |
|--|---|
| Provided by: PIN <i>Contributor(s): Paola Ronzino</i> | |
| User Story | An archaeologist wants to have a data preview. |
| Goal | See a preview of data available to allow a user determining the relevance of the data for her/his research. |
| Scope | Access to data collections and datasets and their structure |
| Preconditions | The user has general ICT skills. |
| Success End Condition | The user can identify the content of collections and datasets and their structure (DBMS, GIS, text). S/he can decide which ones are useful for her/his research question. |
| Failed End Condition | No preview is available. |
| Primary Actor | Archaeologist in the role of data consumer. |
| Trigger | The user discovered the dataset in the portal. |
| KGP Phase | Interpret |
| Entities | Dataset; Service; Actor |
| Service/ Tool used (optional) | |
| WP5-6 Use Case | UC ACCESS_02: Search and browse the Parthenos content cloud across several research infrastructures |

3.6.1.3. AR_03

| | |
|--|--|
| Provided by: PIN <i>Contributor(s): Paola Ronzino</i> | |
| User Story | An archaeologist wants to access collections. |
| Goal | Access collections to compare information about a burial site from the Iron Age with burial practices elsewhere in Europe. |
| Scope | Access to data collections. |
| Preconditions | The user has general ICT skills. |
| Success End Condition | The user can do her/his research from her/his own computer using the services available through the portal. |
| Failed End Condition | The user can't do her/his research because the datasets are not available. |
| Primary Actor | Archaeologist in the role of data consumer. |
| Trigger | The user discovered the dataset in the portal and sees the metadata and the option to download the data. |
| KGP Phase | Connect |
| Entities | Dataset; Service |
| Service/ Tool used (optional) | Catalogue of available resources |
| WP5-6 Use Case | UC ACCESS_03: Retrieval/access metadata about an entity of the Parthenos registry or a resource in the Parthenos content cloud |

3.6.1.4. AR_04

| |
|------------------|
| Provided by: PIN |
|------------------|

| | |
|--------------------------------------|--|
| <i>Contributor(s): Paola Ronzino</i> | |
| User Story | An archaeologist wants to deposit data. |
| Goal | A user wants to deposit some of the data produced in a PARTHENOS compatible archive. |
| Scope | Access to a PARTHENOS compatible archive |
| Preconditions | The user has general ICT skills. |
| Success End Condition | A user can deposit some of the data produced in a PARTHENOS compatible archive and integrate it with similar archives. |
| Failed End Condition | No archives available for depositing data. |
| Primary Actor | Archaeologist in the role of data provider. |
| Trigger | The user has data but does not have an archive for it. |
| KGP Phase | Collect |
| Entities | Dataset; Service |
| Service/ Tool used (optional) | Faceted search functionality, Catalogue of available resources |
| WP5-6 Use Case | UC VREUSE_03: Deposition |

3.6.1.5. AR_05

| | |
|--|--|
| Provided by: PIN <i>Contributor(s): Paola Ronzino</i> | |
| User Story | A VRE manager wants to search and access the services registry. |
| Goal | A VRE manager can discover tools or best practices to achieve a certain goal. |
| Scope | Access the services registry |
| Preconditions | The user has general ICT skills. |
| Success End Condition | The user identifies services useful for his research. |
| Failed End Condition | No appropriate services are found. |
| Primary Actor | VRE manager. |
| Trigger | The user navigated the portal looking for tools and best practices. Directly or via a search-engine. |
| KGP Phase | Connect |
| Entities | Service; Software |
| Service/ Tool used (optional) | Catalogue of available resources |
| WP5-6 Use Case | UC ACCESS_01: Search and browse the Parthenos registry |

3.6.1.6. AR_06

| | |
|--|--|
| Provided by: PIN <i>Contributor(s): Paola Ronzino</i> | |
| User Story | A VRE manager wants to prepare and register a new collection. |
| Goal | This case describes how an archive manager can prepare a collection to be added to the collections held in the registry. |
| Scope | Access to the registry |
| Preconditions | The VRE manager has general ICT skills. |
| Success End Condition | The user can access the documentation to prepare the collection. |
| Failed End Condition | No possibility to add a collection is offered by the registry. |
| Primary Actor | VRE manager. |
| Trigger | The user searches for the registry tool in order to understand if one of her/his |

| | |
|--------------------------------------|--|
| | new collections can be added to the infrastructure. |
| KGP Phase | Collect |
| Entities | Dataset; Service |
| Service/ Tool used (optional) | Catalogue of available resources |
| WP5-6 Use Case | UC REG_01:Manual registration of an entity in the Parthenos registry |

3.6.1.7. AR_07

| | |
|--|---|
| Provided by: PIN <i>Contributor(s): Paola Ronzino</i> | |
| User Story | An archaeologist wants to inspect and enrich visual media documents. |
| Goal | The user wants to inspect and enrich one of the visual documents stored in one of the catalogues available in the portal. |
| Scope | Access to Visual Media Documents |
| Preconditions | The user has general ICT skills. |
| Success End Condition | The user can enrich some Visual Media Document with some new information |
| Failed End Condition | No tools for enriching visual media documents are available. |
| Primary Actor | Archaeologist in the role of data consumer. |
| Trigger | The user has some information that they want to use to enrich some visual media document |
| KGP Phase | Interpret |
| Entities | Research Infrastructure; Service |
| Service/ Tool used (optional) | |
| WP5-6 Use Case | UC VREUSE_05: Process a dataset and publish results; UC CURA_03: Invite curation |

3.6.1.8. AR_08

| | |
|--|--|
| Provided by: PIN <i>Contributor(s): Paola Ronzino</i> | |
| User Story | An archaeologist wants to access information about a metadata format. |
| Goal | Access information about metadata schemas and ontologies used for archiving archaeological resources. |
| Scope | Access to archaeological resources |
| Preconditions | The user has general ICT skills. |
| Success End Condition | The user can obtain information concerning the metadata format and ontologies used for encoding and publishing archaeological information. |
| Failed End Condition | No such information is provided. |
| Primary Actor | Researcher in the role of VRE user |
| Trigger | The user discovers the datasets via the portal or a search-engine. |
| KGP Phase | Connect |
| Entities | Dataset; Service |
| Service/ Tool used (optional) | Metadata input tool, Metadata mapping tool, SKOSifier tool |
| WP5-6 Use Case | UC ACCESS_03: Retrieval/access metadata about an entity of the Parthenos registry or a resource in the Parthenos content cloud |

3.6.1.9. AR_09

| | |
|---|--|
| Provided by: PIN Contributor(s): Paola Ronzino | |
| User Story | An archaeologist wants to retrieve information from vocabularies and gazetteers. |
| Goal | Retrieve information about collections and datasets according to specific terms from a vocabulary or retrieve a location from a gazetteer. |
| Scope | Access to collections of vocabularies, gazetteers and datasets |
| Preconditions | The user has general ICT skills. |
| Success End Condition | The user can identify resources of a certain type or that are located in a specific location. |
| Failed End Condition | No tools are available. |
| Primary Actor | Archaeologist in the role of data consumer. |
| Trigger | The user discovers the datasets via the portal or a search-engine. |
| KGP Phase | Connect |
| Entities | Dataset; Service |
| Service/ Tool used (optional) | Metadata input tool, Metadata mapping tool, SKOSifier tool |
| WP5-6 Use Case | UC ACCESS_03: Retrieval/access metadata about an entity of the Parthenos registry or a resource in the Parthenos content cloud |

3.6.1.10. MINT_01

| | |
|--|--|
| Provided by: MIBACT-ICCU Contributor(s): Sara di Giorgio, Antonio Davide Madonna, Marzia Piccininno | |
| User Story | A user, acting as a content provider, is willing to aggregate a number of different metadata sets, providing them as one unified set. |
| Goal | A content provider aggregates their metadata (according to a specific data model) and disseminates them via OAI-PMH. |
| Scope | Access to OAI-PMH repository |
| Preconditions | The user has metadata but does not have the possibility to aggregate it via OAI-PMH repository. The user has good ICT skills. |
| Success End Condition | The user maps their metadata according to a specific data model via the mapping tool MINT and publishes them in the OAI-PMH repository. |
| Failed End Condition | The metadata is not available in the requested format (csv, xls, xml). |
| Primary Actor | User in the field of heritage and applied disciplines in the role of content provider. |
| Trigger | The user uploads the file in the MINT tool. |
| KGP Phase | Collect |
| Entities | Actor; Service; Software |
| Service/ Tool used (optional) | MINT http://mint.image.ece.ntua.gr/redmine/projects/mint/wiki/Mapping_Tool |
| WP5-6 Use Case | UC AGGR_01: Aggregate resource metadata from research infrastructures into the Parthenos content cloud |

3.6.1.11. MINT_02

| | |
|---|---|
| Provided by: MIBACT-ICCU <i>Contributor(s): Sara di Giorgio, Antonio Davide Madonna, Marzia Piccininno</i> | |
| User Story | A user that manages an OAI-PMH repository, acting as a content provider, is willing to aggregate a number of different metadata, providing it as one unified set. |
| Goal | A content provider aggregates their metadata (according to a specific standard) and disseminates them via OAI-PMH. |
| Scope | Access to OAI-PMH repository |
| Preconditions | The user manages an OAI-PMH repository but is not able to transform the metadata in a specific data model; The user has average ICT skills. |
| Success End Condition | The user maps their metadata according to a specific data model via the mapping tool MINT and publishes them in the OAI-PMH repository. |
| Failed End Condition | The metadata in the content provider repository are not well formed. |
| Primary Actor | User in the field of heritage and applied disciplines in the role of content provider |
| Trigger | The user provides the http address of the repository in the mapping tool. |
| KGP Phase | Collect |
| Entities | Actor; Service; Software |
| Service/ Tool used (optional) | MINT http://mint.image.ece.ntua.gr/redmine/projects/mint/wiki/Mapping_Tool |
| WP5-6 Use Case | UC AGGR_02: Export metadata via standard protocols |

3.6.1.12. MINT_03

| | |
|---|--|
| Provided by: MIBACT-ICCU <i>Contributor(s): Sara di Giorgio, Antonio Davide Madonna, Marzia Piccininno</i> | |
| User Story | A user wants to check the metadata |
| Goal | A user checks the metadata aggregated via the MINT mapping tool. |
| Scope | Access to metadata datasets. |
| Preconditions | Metadata are mapped and aggregated in the mapping tool. |
| Success End Condition | The xml files are validated according to a registered schema (XSD). |
| Failed End Condition | It isn't possible to publish metadata because the mapping tool reports error(s). |
| Primary Actor | User in the field of heritage and applied disciplines in the role of content provider. |
| Trigger | The user launches the check command. |
| KGP Phase | Interpret |
| Entities | Actor; Service |
| Service/ Tool used (optional) | MINT http://mint.image.ece.ntua.gr/redmine/projects/mint/wiki/Mapping_Tool |
| WP5-6 Use Case | UC AGGR_01: Aggregate resource metadata from research infrastructures into the Parthenos content cloud |

3.6.1.13. MINT_04

| | |
|--------------------------|--|
| Provided by: MIBACT-ICCU | |
|--------------------------|--|

| | |
|---|---|
| <i>Contributor(s): Sara di Giorgio, Antonio Davide Madonna, Marzia Piccininno</i> | |
| User Story | A user wants to preview metadata in html format |
| Goal | A user has a preview in html format of the metadata aggregated via mapping tool. |
| Scope | Access to metadata datasets. |
| Preconditions | The metadata are mapped and aggregated correctly in the mapping tool. |
| Success End Condition | The metadata are displayed in a preview window. |
| Failed End Condition | The user cannot see the preview. |
| Primary Actor | User in the field of heritage and applied disciplines in the role of content provider. |
| Trigger | The user selects the preview visualization for its data |
| KGP Phase | Present |
| Entities | Actor; Service; Software |
| Service/ Tool used (optional) | MINT http://mint.image.ece.ntua.gr/redmine/projects/mint/wiki/Mapping_Tool |
| WP5-6 Use Case | UC VREUSE_03: Deposition |

3.6.1.14. MINT_05

| | |
|---|--|
| Provided by: MIBACT-ICCU <i>Contributor(s): Sara di Giorgio, Antonio Davide Madonna, Marzia Piccininno</i> | |
| User Story | A user wants to enrich metadata aggregated in MINT using external SKOS thesauri. |
| Goal | Data aggregated in MINT are enriched by external SKOS thesauri. |
| Scope | Access to Data aggregated in MINT. |
| Preconditions | The user has general ICT skills. |
| Success End Condition | The published data are enriched with SKOS concept of external thesauri. |
| Failed End Condition | The thesauri are not available in SKOS format. |
| Primary Actor | User in the field of heritage and applied disciplines in the role of content provider. |
| Trigger | The user refers an entity to a SKOS concept. |
| KGP Phase | Interpret |
| Entities | Actor; Service; Software |
| Service/ Tool used (optional) | MINT http://mint.image.ece.ntua.gr/redmine/projects/mint/wiki/Mapping_Tool |
| WP5-6 Use Case | UC AGGR_01: Aggregate resource metadata from research infrastructures into the Parthenos content cloud; UC CURA_03: Invite curation |

3.6.1.15. CI_01

| | |
|------------------------------|---|
| Provided by: CULTURAITALIA | |
| User Story | Metadata harvesting. |
| Goal | Metadata acquisition of a data provider within Culturaitalia; metadata are shown in the Portal. |
| Scope | System. |
| Preconditions | Data provider has an OAI-PMH repository and metadata are in PICO format. |
| Success End Condition | Metadata are harvested and indexed in the harvester repository. |
| Failed End Condition | Data provider doesn't have an OAI-PMH repository or doesn't have metadata in PICO format. |

| | |
|--------------------------------------|---|
| Primary Actor | A Cultural Institution in the role of content provider; CulturalItalia, the Italian cultural portal, in the role of harvester |
| Trigger | An administrator of the harvester launches the ingestion process. |
| KGP Phase | Connect |
| Entities | Actor; Service; Software |
| Service/ Tool used (optional) | CulturalItalia Ingestion Panel |
| WP5-6 Use Case | UC ACCESS_04: Retrieval/access of resources from the Parthenos content cloud |

3.6.1.16. CI_02

| | |
|--------------------------------------|--|
| Provided by: CULTURAITALIA | |
| User Story | Metadata validation. |
| Goal | The harvester system provides an automatic check during ingestion process. |
| Scope | System. |
| Preconditions | A data provider has an OAI-PMH repository and metadata are in PICO format. |
| Success End Condition | The validation process finishes without errors. |
| Failed End Condition | The metadata structure is wrong. |
| Primary Actor | A user in the role of VRE manager. |
| Trigger | The ingestion process is launched. |
| KGP Phase | Interpret |
| Entities | Research Infrastructure; Service |
| Service/ Tool used (optional) | CulturalItalia Ingestion Panel |
| WP5-6 Use Case | UC AGGR_01: Aggregate resource metadata from research infrastructures into the Parthenos content cloud |

3.6.1.17. CI_03

| | |
|--------------------------------------|--|
| Provided by: CULTURAITALIA | |
| User Story | Repository update |
| Goal | Update metadata of a content provider. |
| Scope | System. |
| Preconditions | The repository OAI-PMH of data provider was already ingested. |
| Success End Condition | The update process finish without error: existing metadata are overwritten. Deleted metadata are discarded. |
| Failed End Condition | A metadata previously ingested is duplicated; deleted metadata are still showed |
| Primary Actor | A user in the role of VRE manager. |
| Trigger | The update of ingestion process is launched. |
| KGP Phase | Interpret |
| Entities | Service; Actor |
| Service/ Tool used (optional) | CulturalItalia Ingestion Panel |
| WP5-6 Use Case | UC 8: Aggregate resource metadata from research infrastructures into the PARTHENOS content cloud; UC 15: A research infrastructure joins PARTHENOS and integrates its registry; UC 16: Subject coverage; UC 18: Invite cura- |

| | |
|--|------|
| | tion |
|--|------|

3.6.1.18. CI_04

| Provided by: CULTURAITALIA | |
|--------------------------------------|--|
| User Story | Reporting system. |
| Goal | The system sends an email when the ingestion process is finished. |
| Scope | System. |
| Preconditions | Metadata are ingested. |
| Success End Condition | An email is sent to the repository VRE manager. |
| Failed End Condition | The email is not send or it is not complete with requested information |
| Primary Actor | A user in the role of VRE manager. |
| Trigger | The ingestion process is finished. |
| KGP Phase | Present |
| Entities | Research Infrastructure; Service |
| Service/ Tool used (optional) | Culturalitalia Ingestion Panel |
| WP5-6 Use Case | UC VRESET_02: Integration of services in a domain-specific VRE |

3.6.1.19. CI_05

| Provided by: CULTURAITALIA | |
|--------------------------------------|--|
| User Story | Discard invalid metadata. |
| Goal | All the invalid metadata records are not ingested. |
| Scope | System. |
| Preconditions | None |
| Success End Condition | A report provide a list of invalid metadata and it reports also the errors. |
| Failed End Condition | Invalid metadata are ingested in the Portal. |
| Primary Actor | A user in the role of VRE manager. |
| Trigger | The ingestion process is finished. |
| KGP Phase | Interpret |
| Entities | Dataset; Service |
| Service/ Tool used (optional) | Culturalitalia Ingestion Panel |
| WP5-6 Use Case | UC AGGR_01: Aggregate resource metadata from research infrastructures into the Parthenos content cloud |

3.6.1.20. CI_06

| Provided by: CULTURAITALIA | |
|------------------------------|---|
| User Story | Sharing ingested metadata. |
| Goal | Creation of an OAI-PMH repository to share ingested metadata in Culturalitalia. |
| Scope | System. |
| Preconditions | Metadata are ingested in the Portal and they are valid. |
| Success End Condition | The OAI-PMH repository shows metadata in grouped sets and makes them available in different formats. |
| Failed End Condition | It is not possible to create a set of metadata (in a specific format) because mandatory fields are missing. |
| Primary Actor | A user in the role of VRE manager. |
| Trigger | The process to create a new set in the repository is launched. |

| | |
|--------------------------------------|--|
| KGP Phase | Connect |
| Entities | Dataset; Service; Software |
| Service/ Tool used (optional) | OAI-PMH CulturalItalia repository http://www.culturaitalia.it/oaiProviderCI/OAIHandle |
| WP5-6 Use Case | UC VREUSE_04: Private and public sharing of resources deposited in the VRE workspace |

3.6.2. Use cases from Language-related studies

3.6.2.1. OEAW_01

| | |
|---|--|
| Provided by: OEAW <i>Contributor(s): Vanessa Hanneschläger, Klaus Illmayer</i> | |
| User Story | A researcher wants to search/browse available metadata about language resources using a combination of the faceted and text search. |
| Goal | Get a list of language resources that match search criteria. |
| Success End Condition | List of the resources matching search criteria is returned. |
| Failed End Condition | None of the resources matches search criteria. |
| Primary Actor | Researcher in the role of data consumer |
| KGP Phase | Connect |
| Entities | Dataset; Service |
| Service/ Tool used (optional) | VLO – https://vlo.clarin.eu |
| WP5-6 Use Case | UC ACCESS_01: Search and browse the Parthenos registry; UC ACCESS_02: Search and browse the Parthenos content cloud across several research infrastructures |

3.6.2.2. OEAW_02

| | |
|---|---|
| Provided by: OEAW <i>Contributor(s): Vanessa Hanneschläger, Klaus Illmayer</i> | |
| User Story | A researcher wants to preview found resources. |
| Goal | Get detailed information about resource. |
| Success End Condition | User gets all available information about resource including description, resource type, availability and link to the resource. |
| Failed End Condition | None of the resources matches search criteria. |
| Primary Actor | Researcher in the role of data consumer |
| Trigger | The user found possibly relevant resources. |
| KGP Phase | Present |
| Entities | Dataset; Service |
| Service/ Tool used (optional) | VLO – https://vlo.clarin.eu |
| WP5-6 Use Case | UC ACCESS_03: Retrieval/access metadata about an entity of the Parthenos registry or a resource in the Parthenos content cloud |

3.6.2.3. OEAW_03

| | | | | |
|--|--|---------------------|--|--|
| Provided by: OEAW Contributor(s): Vanessa Hanneschläger, Klaus Illmayer | | | | |
| User Story | A researcher wants to access a resource. | | | |
| Goal | Get the resource and use it for the research. | | | |
| Success End Condition | User uses links from preview page in order to access to the resource. | | | |
| Failed End Condition | Link is broken or doesn't point to the resource | | | |
| Primary Actor | Researcher in the role of data consumer | | | |
| Trigger | Resource is accessible. | | | |
| Extensions | The Resource has license type "request required". User uses contact information of resource provider to obtain the access. | | | |
| Main Success Scenario | | | | |
| | KGP Phase | Entities | Service/ Tool used (optional) | WP5-6 Use Case |
| Access resource from search results | Connect | Dataset; Service | VLO https://vlo.clarin.eu | UC ACCESS_04: Retrieval/access of resources from the Parthenos content cloud |
| Take the accessed resource and use it for research | Collect | Dataset; Service | | UC VREUSE_05: Process a dataset and publish results |

3.6.2.4. OEAW_04

| | | | | |
|--|--|---------------------|--|---|
| Provided by: OEAW Contributor(s): Vanessa Hanneschläger, Klaus Illmayer | | | | |
| User Story | A researcher wants to find non-digitised material in physical archives. | | | |
| Goal | Find out what relevant material exists and where it is. | | | |
| Success End Condition | User finds information about and location of material relevant to their research question. | | | |
| Failed End Condition | User cannot find out what material exists or cannot locate existing material. | | | |
| Primary Actor | Researcher in the role of data consumer | | | |
| Trigger | Research question cannot be answered by available digitised material. | | | |
| Extensions | Geo- / Temporal visualization of material distribution; filter results by time/location. | | | |
| Main Success Scenario | | | | |
| | KGP Phase | Entities | Service/ Tool used (optional) | WP5-6 Use Case |
| Search for material with help of metadata | Connect | Dataset; Service | CENDARI http://www.cendari.eu | UC ACCESS_01: Search and browse the Parthenos registry |
| Find new material on location | Connect | Dataset; Service | | UC REG_01: Manual registration of an entity in the Parthenos registry |

3.6.2.5. OEAW_05

| |
|-------------------|
| Provided by: OEAW |
|-------------------|

| | | | | |
|--|--|---|--|--|
| <i>Contributor(s): Vanessa Hanneschläger, Klaus Illmayer</i> | | | | |
| User Story | A researcher wants to reference and comment on datasets describing non-digitized material. | | | |
| Goal | Make notes about material and connect them with that material in a VRE. | | | |
| Success End Condition | User has a commented set of material to work with. | | | |
| Failed End Condition | No VRE meeting the needs of the user is available. | | | |
| Primary Actor | Researcher in the role of VRE user | | | |
| Trigger | User found information about material they want to work with. | | | |
| Extensions | VRE offers possibility to discuss notes and selection of material with other researchers. | | | |
| Main Success Scenario | | | | |
| | KGP Phase | Entities | Service/ Tool used (optional) | WP5-6 Use Case |
| Reference and comment non-digitized material | Interpret | Dataset; Service | CENDARI http://www.cendari.eu | UC VREUSE_02: Reference entities of the Parthenos registry in VRE posts |
| Upload private digitized material in a private VRE | Collect | Dataset; Service; Research Infrastructure | CENDARI http://www.cendari.eu | UC VREUSE_03: Deposition |
| Share private digitized material with selected researchers | Present | Dataset; Service; Actor | CENDARI http://www.cendari.eu | UC VREUSE_04: Private and public sharing of resources deposited in the VRE workspace |

3.6.2.6. CLARIN_01: Corpus-based Analysis of Historical Newspapers

| | |
|--|--|
| Provided by: CLARIN / BBAW | |
| <i>Contributor(s): Susanne Haaf, Axel Herold</i> | |
| User Story | <p>An interdisciplinary group of researchers (IGR) consisting of linguists and historians is interested in a certain historical newspaper (e.g. the “Neue Rheinische Zeitung”) with regard to specialities in language and ways of reporting on certain topics, events or discourses as well as political tendencies. In order to identify the characteristics of the respective newspaper and evaluate analysis results as significant or not, those results have to be compared to other documents (newspapers and texts from other text types) from the same time period (synchronic view) as well as from another time period (diachronic view). The newspaper corpus under consideration is too large to be analysed manually (~300 issues of the “Neue Rheinische Zeitung” plus corpora for comparison). Thus, the corpus has to be digitally available in order to gain corpus-based results with automatic methods. Such automatic methods are e.g. certain kinds of linguistic processing (lemmatization, morphological analysis, named-entity recognition, recognition of significant terms, recognition of grammatical structures etc.) as well as topic analysis methods in order to automatically find articles on certain events and discourses. The automatic analysis should lead to interpretable results from a linguistic and/or historical point of view, e.g. easy vs. complex sentence structures; Who is the intended audience of the news-</p> |

| | | | | |
|------------------------------|---|-----------------|---------------------------|-----------------------|
| | <p>paper?; terms of political agitation; What's the political orientation of the newspaper?; What's its opinion on certain topics/events?</p> <p>The IGR has been granted funding for the creation of the primary corpus (i.e. the digitization of the newspaper under consideration). For comparison the group has to resort to corpora which are already available otherwise. Equally, for corpus analysis with the named methods, it is planned to re-use existing tools. Thus, in order to combine the primary corpus with other corpora and to analyse all data with given tools, it is necessary that all data are interoperable and that the data formats applied are compatible with the respective analysis tools. For this process it would be beneficial if standardized formats had been used in all cases.</p> <p>The researchers of the IGR have to have knowledge about common digitization methods, standard formats and existing (corpus and software) resources.</p> <p>After finishing the project all data and workflows used should be made publicly available in order to enable the verification of project results.</p> | | | |
| Goal | Humanities researchers want to use corpus-based methods for their research. They are interested in the specifics of language, style, etc. for a certain newspaper and discourse (linguistics researchers) as well as in the specifics of reporting about certain historical events and time periods. | | | |
| Scope | Corpus-based comparative research on reporting specifics of newspapers. | | | |
| Preconditions | <ul style="list-style-type: none"> • A primary corpus (a significant number of issues of a certain historical newspaper) and corpora for comparison (other newspapers/other documents from different text types; synchronic/diachronic); • Large/satisfying amounts of data; • Access to data/corpora from different sources (in order to create corpora at reasonable cost); • Data of similar quality, in uniform or compatible formats which were created based on similar guidelines (i.e. they have to be truly interoperable); • Powerful corpus query and analysis applications which may handle standardized I/O formats; • Knowledge of project staff about common digitization methods, the application of certain tools, the usage of standardized formats etc. | | | |
| Success End Condition | Researchers were able to gather and enrich the data they needed and to use the tools and data in a reasonable manner. They gained reliable results with regard to their research question. | | | |
| Failed End Condition | The re-use of given tools and data was too complicated due to non-standard formats, lack of documentation, lack of quality or other problems. Thus, the project had to develop their individual, isolated solutions. Due to poor resources, the research results could not be based on a large enough data sample and therefore result in inadequate confidence levels for the hypotheses. | | | |
| Primary Actor | Linguists and historians in the role of data consumers. | | | |
| Trigger | Precise research question. Awareness about infrastructures which could be of use for solving the research problem. | | | |
| Main Success Scenario | | | | |
| | KGP Phase | Entities | Service/ Tool used | WP5-6 Use Case |

| | | | (optional) | |
|--|-----------|---|------------|--|
| 1. Determination of the research question and goal | Interpret | Actor | | UC VRESET_02: Integration of services in a domain-specific VRE |
| 2. Selection of the NP (and NP issues) of primary interest; creation of the primary corpus | Connect | Service | | UC VRESET_02: Integration of services in a domain-specific VRE |
| 3. Detection, selection and, if necessary, amendment of relevant data for comparison, from (a) other NPs, (b) other documents than NPs | Interpret | Dataset; Service | | UC ACCESS_02: Search and browse the Parthenos content cloud across several research infrastructures |
| 3a. Detection of relevant data sets | Connect | Dataset; Service | | UC ACCESS_02: Search and browse the Parthenos content cloud across several research infrastructures |
| 3b. Gathering of the metadata for the relevant data sets | Connect | Dataset; Service | | UC ACCESS_03: Retrieval/access metadata about an entity of the Parthenos registry or a resource in the Parthenos content cloud |
| 3c. Gathering of the relevant data sets | Connect | Dataset; Service | | UC ACCESS_04: Retrieval/access of resources from the Parthenos content cloud |
| 4. Data analysis within a shared working platform | Interpret | Dataset; Service | | UC VREUSE_04: Private and public sharing of resources deposited in the VRE workspace |
| 4a. Creation of a working platform (VRE) | Connect | Research Infrastructure | | UC VRESET_01: Set-up of a domain-specific VRE |
| 4b. Inclusion of the relevant data into the VRE | Collect | Research Infrastructure; Dataset; Service | | UC VREUSE_03: Deposition |
| 4c. Selection and inclusion of services for corpus analysis (wrt linguistic, lexical, topical, etc. features) and corpus comparison | Interpret | Dataset; Service | | UC VRESET_02: Integration of services in a domain-specific VRE |
| 4d. Corpus analysis and comparison | Interpret | Dataset; Service | | UC VREUSE_05: Process a dataset and publish results |
| 5. Publication of results and provision of the corpora | Present | Dataset; Service | | UC VREUSE_04: Private and public sharing of resources deposited in the VRE workspace |

| | |
|-----------------------|--|
| Sub-Variations | <p>2'. Researchers first need training on digitization methods and common standards.</p> <p>3'. Problems with the acquisition of corpora for comparison.</p> <p>3'a. Corpora are poorly described (in terms of their metadata) and thus difficult to retrieve – extended effort is necessary for the gathering of corpora.</p> <p>3'b. There is no sufficient number of corpora useful for comparison available – extended effort is necessary for corpus compilation.</p> <p>3'c. Relevant data are accessible but are not interoperable with the primary corpus/among themselves – extended effort is necessary for data conversion.</p> <p>4'. Problems with the acquisition of tools.</p> <p>4'a. Existing tools for data analysis can't be re-used (wrong I/O-formats) and extended effort necessary for data conversion.</p> <p>4'b. Some necessary tools for data analysis do not exist at all and extended effort necessary for software implementation.</p> |
| Extensions | <p>5a. Integration of the newly created primary corpus into an existing infrastructure (such as CLARIN); all further analyses are carried out within the respective infrastructure (by usage of tools and corpus query facilities available within the infrastructure).</p> <p>6a. Provision of the corpora used in the project (especially of the primary corpus) within a given infrastructure.</p> <p>6b. Analyses have been carried out within the infrastructure from the beginning.</p> |

3.6.3. Use cases from Studies of the Past

3.6.3.1. TCD01

| | | | | |
|--|---|-----------------|--------------------------------------|------------------------------|
| Provided by: TCD Contributor(s): Jennifer Edmond, Vicky Garnett | | | | |
| User Story | Researcher wishes to extract data on battlefield transportation from Europeana 1914-18 via the Europeana RESTful API and make notes on selected outputs of the API call in the CENDARI NTE in order to create a set of metadata for broad analysis. | | | |
| Goal | Extract data via API and import into note taking environment for further annotation (Europeana 1914-18 with CENDARI NTE). | | | |
| Scope | Europeana API and CENDARI NTE. | | | |
| Level | Summary | | | |
| Preconditions | Researcher is familiar with API use and resulting data output. | | | |
| Success End Condition | Researcher is able to extract reusable data first from Europeana, and then from CENDARI NTE once annotated. | | | |
| Failed End Condition | Europeana API data is not compatible with CENDARI NTE, and researcher has to look to another platform for analysis. | | | |
| Primary Actor | Researcher in the role of data consumer | | | |
| Trigger | API call. | | | |
| Frequency | Frequent within the scope of a single project. | | | |
| Main Success Scenario | | | | |
| | KGP Phase | Entities | Service/ Tool used (optional) | WP5-6 Use Case |
| Extract data via | Connect | Dataset; | CENDARI | UC VRESET_02: Integration of |

| | | | | |
|--|---------|----------------------------------|---|--|
| API | | Service | http://www.cendari.eu | services in a domain-specific VRE |
| Import (Previous) into note taking environment | Collect | Dataset; Service; Software | Note Management Tool | UC VRESET_02: Integration of services in a domain-specific VRE |

3.6.3.2. TCD02

| | |
|--|---|
| Provided by: TCD Contributor(s): Jennifer Edmond, Vicky Garnett | |
| User Story | Researcher wants to gather testimonies from refugees following the Hungarian Revolution of 1956. She wants to conduct a search on the Europeana Portal to look for digital content that she can download and analyse. |
| Goal | Extract data via platform portal for developing a collection (Europeana portal search). |
| Scope | Europeana Portal. |
| Level | Sub-function. |
| Preconditions | Researcher has some ICT skills, but is not familiar with API-use or coding. |
| Success End Condition | Researcher will have comprehensive list of testimonies available via Europeana. |
| Failed End Condition | Researcher will not be able to identify or find testimonies due to poor metadata. |
| Primary Actor | Researcher |
| Trigger | Search in portal. |
| KGP Phase | Connect |
| Entities | Dataset; Service |
| Service/ Tool used (optional) | |
| WP5-6 Use Case | UC ACCESS_03: Retrieval/access metadata about an entity of the Parthenos registry or a resource in the Parthenos content cloud |

3.6.3.3. TCD03

| | |
|--|--|
| Provided by: TCD Contributor(s): Jennifer Edmond, Vicky Garnett | |
| User Story | A lecturer wants to find materials for a workshop with a group of History Postgrads looking into medieval attitudes to women. Uses TRAME to search for documents that show accounts of women. She will use these in a workshop looking at sentiment analysis tools, but needs to be able to download the data. |
| Goal | Search for and extract content for use in analysis tool – TRAME search portal and Sentiment Analysis. |
| Scope | TRAME. |
| Level | Sub-function. |
| Preconditions | Lecturer is competent with ICT skills and processing data for sentiment analysis tools. |
| Success End Condition | Lecturer will have a complete set of data ready for use in a sentiment analysis tool. |
| Failed End Condition | Lecturer will not have dataset ready for sentiment analysis, and will instead have to resort to longer manual search with fewer results, where students will have to |

| | | | | |
|---------------------------------|---|--|--------------------------------------|--|
| | manually analyse for positive and negative attitudes towards women. | | | |
| Primary Actor | Lecturer as VRE user | | | |
| Trigger | Search in TRAME. | | | |
| Priority | High | | | |
| Frequency | One-off. | | | |
| Main Success Scenario | | | | |
| | KGP Phase | Entities | Service/ Tool used (optional) | WP5-6 Use Case |
| Integrate TRAME search portal | Collect | Research Infrastructure; Service; Software | | UC REG_02: A research infrastructure joins Parthenos and integrates its registry |
| Use Trame search portal | Connect | Service; Software | TRAME – git-trame.fefonlus.it | UC VRESET_02: Integration of services in a domain-specific VRE |
| Download the data | Connect | Dataset; Service | | UC AGGR_02: Export metadata via standard protocols |
| Use the Sentiment analysis tool | Interpret | Dataset; Service; Software | | UC VRESET_02: Integration of services in a domain-specific VRE |

3.6.3.4. TCD04

| | |
|---|--|
| Provided by: TCD <i>Contributor(s): Jennifer Edmond, Vicky Garnett</i> | |
| User Story | Researcher wants to structure their multiple datasets in order to make it interoperable, and uses the FlareNet guidelines to determine appropriate standards. |
| Goal | Preparation of multiple datasets |
| Scope | Undetermined (Flarenet) |
| Level | Primary Task. |
| Preconditions | Researcher is aware of the importance of standards and data structure for future reuse in an interoperable context. Researcher is aware of the FlareNet listing. |
| Success End Condition | Researcher's data can be used and reused in combination. |
| Failed End Condition | Researcher chooses an inappropriate standard and their data cannot be made interoperable by the researcher or others. |
| Primary Actor | Researcher |
| Trigger | Data Creation (recognition of the need for choosing of a standard). |
| Priority | High |
| Frequency | Once |
| KGP Phase | Connect |
| Entities | Dataset; Service; Software; Research Infrastructure |
| Service/ Tool used (optional) | TRAME – git-trame.fefonlus.it |
| WP5-6 Use Case | UC AGGR_01: Aggregate resource metadata from research infrastructures into the Parthenos content cloud |

3.6.3.5. SISME01

| |
|--|
| Provided by: SISME01 <i>Contributor(s): Emiliano Degl'Innocenti, Roberta Giacomi, Veronica Boarotto</i> |
|--|

| | | | | |
|--|--|---------------------|--------------------------------------|---|
| User Story | <p>A research team has been established to produce a digital edition of Cassiodorus' <i>Institutiones</i>. The <i>Institutiones</i>, considered the most important work by Cassiodorus, were written around 560 in <i>Vivarium</i>, a monastery with a relevant <i>scriptorium</i> founded by Cassiodorus himself. The manuscripts held by the <i>scriptorium</i> were scattered in many countries and libraries.</p> <p>The digital edition will record not only the critical text but also the manuscript tradition.</p> | | | |
| Goal | <p>The digital edition will provide:</p> <ul style="list-style-type: none"> • a map of where the manuscripts are held today and a map of where they come from originally with linked catalographic information; • a stemmatological analysis of the evolution of the text; • images and transcriptions of the most important manuscripts that helped to reconstruct the text; • a timeline of the diffusion of the text. | | | |
| Scope | <ul style="list-style-type: none"> • Access to virtual libraries (with digitised manuscripts); • Access to manuscript catalogues; • Access to manuscript repositories. | | | |
| Preconditions | <ul style="list-style-type: none"> • Reference tools for the disambiguation of: <ul style="list-style-type: none"> • names of persons and places, • documents (i.e.: manuscripts shelfmarks), • titles of texts/works; • Access to a LOD web of authors, works, documents and related information (i.e.: available information about origins and provenances of the documents); • Tool to perform searches across multiple scholarly resources; • Tool to display the results on a map and / or timeline; • Knowledge on the structure of the involved databases and resources. | | | |
| Success End Condition | Establish a digital edition provided with a stemmatological analysis, maps and a timeline. | | | |
| Failed End Condition | Unavailability of useful tools to perform the research. | | | |
| Primary Actor | A team of researchers (typology: data consumer). | | | |
| Trigger | Perform a research within the available tools. | | | |
| Main Success Scenario | | | | |
| | KGP Phase | Entities | Service/ Tool used (optional) | WP5-6 Use Case |
| 1: Get a list of all the manuscripts containing « <i>Institutiones</i> » | Connect | Dataset; Service | | UC REG_01:Manual registration of an entity in the Parthenos registry |
| 2: Search timeline Information about the manuscripts | Connect | Dataset; Service | | UC ACCESS_02: Search and browse the Parthenos content cloud across several research infrastructures |
| 3: Search information about places, dates and bibliography | Connect | Dataset; Service | | UC ACCESS_02: Search and browse the Parthenos content cloud across several research infrastructures |
| 4: Compare Text of Cassiodorus with others author's work | Interpret | Dataset; Service | | UC VREUSE_05: Process a dataset and publish results |

3.6.3.6. SISMEI_02

| | | | | |
|---|---|-------------------------|--------------------------------------|---|
| Provided by: SISMEI Contributor(s): <i>Emiliano Degl'Innocenti, Roberta Giacomi, Veronica Boarotto</i> | | | | |
| User Story | The goal of the research is to find out how many Italian 15th century preachers employed philosophy in their preaching activity; to select relevant collections of sermons (edited texts and unedited, i.e. preserved in manuscripts and incunabula): the relevance is not represented only by the number of the manuscripts within the collections, but by the cultural significance of the sources. These data will allow users to conduct an analysis of the meaning of philosophical knowledge and its change in late medieval preaching. | | | |
| Goal | As a scholar interested in late-medieval preaching, a researcher wants to be able to find the most relevant collections of manuscripts and rare books held by European libraries, with special focus on 15th century sermons collections. | | | |
| Scope | <ul style="list-style-type: none"> • Access to European libraries; • Access to digitization of sources; • Access to bibliographical databases; • Access to manuscript catalogues; • Access to other research projects related to the study of late medieval preaching; • Access to authority lists about authors and titles. | | | |
| Preconditions | <ul style="list-style-type: none"> • Reference tools for the disambiguation of: <ul style="list-style-type: none"> ○ names of persons and places, ○ documents (i.e.: manuscripts shelfmarks), ○ titles of texts/works; • Tool to perform searches across multiple scholarly resources; • Tool to display the results on a map and / or timeline. | | | |
| Success End Condition | A dossier exists providing information about the most relevant collections of 15th century Italian sermons and about the author's life, the places where the preachers were active; a geo-chronological map to trace their activity. | | | |
| Failed End Condition | The researcher can't have access to the repositories or the tools s/he needs. | | | |
| Primary Actor | A researcher (typology: data consumer). | | | |
| Trigger | Perform an analysis on the available material. | | | |
| Main Success Scenario | | | | |
| | KGP Phase | Entities | Service/ Tool used (optional) | WP5-6 Use Case |
| 1: A researcher wants to find out how many collections of unedited sermons are held in European libraries | Connect | Dataset; Actor; Service | | UC ACCESS_02: Search and browse the Parthenos content cloud across several research infrastructures |
| 2: He wants to publish a critical edition of 15th century Italian collections of sermons held in Italian libraries. | Present | Actor; Service | | UC VREUSE_05: Process a dataset and publish results |
| 3: He wants to have a list of Italian preachers | Connect | Actor, Dataset; Ser- | | UC ACCESS_02: Search and browse the Parthenos content |

| | | | | |
|---|---------|-------------------------|--|---|
| that employ philosophical concepts and quotations in collections of sermons. | | vice | | cloud across several research infrastructures |
| 4: A researcher wants to search for Aristotle's quotations in 15th century sermons | Connect | Actor, Dataset; Service | | UC ACCESS_02: Search and browse the Parthenos content cloud across several research infrastructures |
| 5: A researcher wants to search for Plato's quotations in 15th century sermons | Connect | Actor, Dataset; Service | | UC ACCESS_02: Search and browse the Parthenos content cloud across several research infrastructures |
| 6: A researcher wants to search for classical, literary and philosophical quotations in late medieval sermons | Connect | Actor, Dataset; Service | | UC ACCESS_02: Search and browse the Parthenos content cloud across several research infrastructures |

3.6.3.7. SISMELO_03

| | | | | |
|--|--|-----------------|--------------------------------------|-----------------------|
| Provided by: SISMELO | | | | |
| Contributor(s): <i>Emiliano Degl'Innocenti, Roberta Giacomi, Veronica Boarotto</i> | | | | |
| User Story | A researcher is interested in finding information about Ramon's Llull works and textual tradition: to have a list of his works, with special focus on the Catalan and Arabic philosophical and theological production; the indication of the language and the place in which they were written. The researcher also wants to know if, where and when the original works were translated, and – in case of translations – he wants to know by whom. | | | |
| Goal | The researcher wants to have a geo-chronological map where he can place Llull's works; for every work he wants to be able to have chronological and bibliographical information: year of composition, language, translation, editions and related manuscripts (digitized, if available). | | | |
| Scope | <ul style="list-style-type: none"> • Access to virtual libraries, possibly with digitized manuscripts; • Access to bibliographical databases; • Access to manuscript catalogues; • Access to multilingual resources and repositories. | | | |
| Preconditions | <ul style="list-style-type: none"> • Tool to perform searches across multiple scholarly resources; • Tool to display the results on a map and / or timeline; • Access to a LOD web of authors, works, documents and related information (i.e.: available information about origins and provenances of the documents). | | | |
| Success End Condition | The researcher can draw a line of development of Llull's philosophical production and see how it changes over time and space. | | | |
| Failed End Condition | The researcher can't have access to the repositories or the tools he needs. | | | |
| Primary Actor | A researcher (typology: data consumer). | | | |
| Trigger | Perform an analysis on the available material. | | | |
| Main Success Scenario | | | | |
| | KGP Phase | Entities | Service/ Tool used (optional) | WP5-6 Use Case |

| | | | | |
|---|---------|-------------------------|--|---|
| 1: A researcher wants to find all the works of Lull in which he employs the Ars combinatoria. | Connect | Actor, Dataset; Service | | UC ACCESS_02: Search and browse the Parthenos content cloud across several research infrastructures |
| 2: He wants to find how many Lullian works are held in German Libraries. | Connect | Actor, Dataset; Service | | UC ACCESS_02: Search and browse the Parthenos content cloud across several research infrastructures |
| 3: He wants to know how many Latin and Catalan works Lull wrote. | Connect | Actor, Dataset; Service | | UC ACCESS_02: Search and browse the Parthenos content cloud across several research infrastructures |
| 4: He wants to know when Lull travelled to Italy. | Connect | Actor, Dataset; Service | | UC ACCESS_02: Search and browse the Parthenos content cloud across several research infrastructures |
| 5: He wants to know what other works are copied with the <i>Liber de amico et amato</i> . | Connect | Actor, Dataset; Service | | UC ACCESS_02: Search and browse the Parthenos content cloud across several research infrastructures |

3.6.3.8. SISMELE_04: Tracking of the circulation of the legend of Barlaam and Josaphat.

| | |
|---|---|
| Provided by: SISMELE Contributor(s): Emiliano Degl'Innocenti, Roberta Giacomi, Veronica Boarotto | |
| User Story | A researcher wants to track the legend of Buddha in Multilanguage manuscripts and prints using pre-existent tools. |
| Goal | The researcher wants to track in space and times texts about the legend of Buddha using information stored in different multi-lingual databases. |
| Scope | Access information stored in catalogues held by contemporary libraries; Access to virtual libraries with many digitized reproductions of manuscripts and incunabula. |
| Preconditions | <ul style="list-style-type: none"> • Reference tools for the disambiguation of: <ul style="list-style-type: none"> ○ names of persons and places, ○ documents (i.e.: manuscripts shelfmarks), ○ titles of texts/works; • Tool to perform searches across multiple scholarly resources; • Tool to display the results on a map and / or timeline. |
| Success End Condition | The researcher obtains a map with geo and time coordinates of the circulation of the legend. |
| Failed End Condition | Information about sources and / or secondary literature is not accessible. User has to perform many different searches over a number of dispersed resources. |
| Primary Actor | Researchers working in the history disciplines (typology: data con- |

| | | | | |
|--|---|----------------------------------|--------------------------------------|---|
| | sumers). | | | |
| Trigger | A researcher is interested in tracking the circulation of the legend. | | | |
| Main Success Scenario | | | | |
| | KGP Phase | Entities | Service/ Tool used (optional) | WP5-6 Use Case |
| 1: Access to all existent editions of the manuscript in the Gallica, BVMM, E-codices, Manuscriptorium databases. | Connect | Dataset; Service; Software | | UC ACCESS_02: Search and browse the Parthenos content cloud across several research infrastructures |
| 2: Access to all existent editions of the prints from Incunabula Short-Title Catalogue (ISTC) and MEI (Material Evidence in Incunabula). | Connect | Dataset; Service; Software | | UC ACCESS_02: Search and browse the Parthenos content cloud across several research infrastructures |
| 3: Assess how many medieval and renaissance manuscripts of this legend survive today in our libraries using the META-OPAC CERL Portal to access a wide number of electronic catalogues of manuscripts. | Interpret | Service | | UC ACCESS_02: Search and browse the Parthenos content cloud across several research infrastructures |
| 4: Ensure that the CERL Thesaurus is running at the back of the above listed tools to assure inclusiveness of data. | Connect | Research Infrastructure; Service | | UC VRESET_02: Integration of services in a domain-specific VRE |

3.6.3.9. SISMELO5

| | |
|---|--|
| Provided by: SISMELO5 Contributor(s): <i>Emiliano Degl'Innocenti, Roberta Giacomi, Veronica Boarotto</i> | |
| User Story | A researcher wants to know who was the writer of the first western treatise on how retard old age, to collect his possible sources and reconstruct the diffusion of his work. |
| Goal | The researcher wants to know what kind of works are transmitted together with the <i>De retardatione accidentium senectutis</i> . He wants to then outline a hierarchy in the textual tradition of the works offering a possible prolongation of life, to analyse the cultural and social context of the transmission of those Western fundamental works on prolonging life and to map the geographical and social itineraries of a more complete literary history of Western prolonging life and immortality. |
| Scope | Access information stored in catalogues of manuscripts and prints held by contemporary libraries; assess what was available during the Medieval and Renaissance period by accessing information in catalogues of medieval libraries; access related primary sources reproductions and descriptions; access related secondary literature; access authority lists and repertories of medieval authors. |

| | | | | |
|--|---|-------------------------------|--|--|
| Preconditions | <ul style="list-style-type: none"> Reference tools for the disambiguation of: <ul style="list-style-type: none"> names of persons and places, documents (i.e.: manuscripts shelfmarks), titles of texts/works; Tool to perform searches across multiple scholarly resources; Tool to display the results on a map and / or timeline; Knowledge on the structure of the involved databases and re-sources. | | | |
| Success End Condition | A dossier with all the relevant resources related to the <i>De retardatione accidentium senectutis</i> , including primary sources and secondary literature, is produced. A hierarchy in the textual tradition of the works offering a possible prolongation of life is outlined. It could be possibly exported and re-used. A map and / or timeline displaying the information is available. | | | |
| Failed End Condition | Information about sources and / or secondary literature is not accessible. User has to perform many different searches over a number of dispersed re-sources. | | | |
| Primary Actor | Researchers working in the disciplines of the past (typology: data consumers). | | | |
| Other Actor | <p>Researchers and institutions producing data on authors, texts, sources etc. (typology: data providers).</p> <p>D/H community involved on the same field (typology: standards developers).</p> | | | |
| Trigger | A researcher is interested in tracking, analysing and mapping the textual tradition of a given text. | | | |
| Main Success Scenario | | | | |
| | KGP Phase | Entities | Service/ Tool used (optional) | WP5-6 Use Case |
| 1: Trace the manuscripts in Latin or in French thanks to the search engine TRAME | Interpret | Service; Software | TRAME – http://git-trame.fefonlus.it/ | UC VRESET_02: Integration of services in a domain-specific VRE |
| 2: Extend the queries in the virtual manuscripts libraries available | Interpret | Dataset; Service; | | UC VRESET_01: Set up domain-specific VRE; UC CURA_01: Subject coverage |
| 3: Use authority lists of authors and of titles of works as available in MIRABLE | Interpret | Dataset; Service; Software | MIRABILE – http://mirabile-fe.netseven.it/ | UC VRESET_02: Integration of services in a domain-specific VRE |
| 4: Undertake a systematic re- | Interpret | Dataset; Service; | Incunabula Short-Title Catalogue (ISTC), the ERC program on “The | UC VRESET_02: |

| | | | | |
|---|--|----------|--|--|
| search from manuscripts to print, using the Incunabula Short-Title Catalogue (ISTC), the ERC program on “The 15th-century Book Trade”, the English Short Title Catalogue and the digital collections of libraries | | Software | 15th-century Book Trade”, the English Short Title Catalogue and the digital collections of libraries | Integration of services in a domain-specific VRE; UC CURA_01: Subject coverage |
|---|--|----------|--|--|

3.6.3.10. SISMEI_06

Generic Use Case: A historian wants to track the dissemination of a given author’s works during the Medieval and Early Modern period.

| | |
|---|--|
| Provided by: SISMEI Contributor(s): <i>Emiliano Degl’Innocenti, Roberta Giacomi, Veronica Boarotto</i> | |
| User Story | <p>Within the disciplines of the past community, scholars are interested in the accessibility of research data on authors, sources (i.e.: manuscripts and printed books) and transmitted works. Other related information, coming from repertoires and hand lists, authority lists and bibliographies are important as well to provide additional context and are to be integrated. Dealing with multilingual contents, access to both Latin and vernacular resources is required.</p> <p>Our researcher is interested in tracking the dissemination of Donatus’ <i>Ars minor</i> – a Medieval condensation of the late Roman schoolbook, in which a series of dialogues conveyed the rudiments of the language in the Medieval and early modern era.</p> <p>The goal is to investigate the spread of literacy in early modern Western European society, since <i>Ars minor</i> was quite possibly the first book printed with moveable type both in Germany and in Italy. Unfortunately, the editions have been lost, but the researcher can compensate for the loss of evidence today with the use of documentary material made available by focused initiatives and other scholarly projects and databases.</p> |
| Goal | Address the question of the spread of literacy in early modern European society using a combination of digital resources. |
| Scope | Access information stored in catalogues of manuscripts held by contemporary libraries; assess what was available during the Medieval and Renaissance period by accessing information in catalogues of medieval libraries; access related primary sources reproductions and descriptions; access related secondary literature. |
| Preconditions | <ul style="list-style-type: none"> • Budget; • Time; • NER service to extract relevant entities (i.e. names of persons and places, |

| | | | | |
|---|---|----------------------------|--------------------------------------|--|
| | titles of works etc.); <ul style="list-style-type: none"> Reference tools for the disambiguation of: <ul style="list-style-type: none"> names of persons and places, documents (i.e.: manuscripts shelfmarks), titles of texts/works; Access to a LOD web of authors, works, documents and related information (i.e.: available information about origins and provenances of the documents); Tool to perform searches across multiple scholarly resources; Tool to display the results on a map and / or timeline; Knowledge on the structure of the involved databases and resources. | | | |
| Success End Condition | A dossier with all the relevant resources related to the textual tradition of <i>Ars Minor</i> , including primary sources and secondary literature, is produced. It could be possibly exported and re-used. A map and / or timeline displaying the information is available. | | | |
| Failed End Condition | Information about sources and / or secondary literature is not accessible. User has to perform many different searches over a number of dispersed resources. | | | |
| Primary Actor | Researchers working in the Disciplines of the past (typology: data consumers). | | | |
| Other Actor | Researchers and institutions producing data on authors, texts, sources etc. (typology: data providers). Holding institutions preserving sources (typology: GLAMs, holding institutions). DH community involved on the same field (typology: standards developers). | | | |
| Trigger | A researcher is interested in tracking the textual tradition of a given text or the transmission of a given manuscript. | | | |
| Main Success Scenario | | | | |
| | KGP Phase | Entities | Service/ Tool used (optional) | WP5-6 Use Case |
| 1: Survey all existent editions of Donatus | Connect | Dataset; Service; | | UC ACCESS_04: Retrieval/access of resources from the Parthenos content cloud |
| 2: Assess the 15th and 16th-century use of these editions | Interpret | Dataset; Service; | | UC ACCESS_04: Retrieval/access of resources from the Parthenos content cloud |
| 3: Assess how many medieval and renaissance manuscripts of this work survive today in our libraries using the META-OPAC CERL Portal to access a wide number of electronic cata- | Interpret | Dataset; Service; Software | META-OPAC CERL Portal | UC VRESET_02: Integration of services in a domain-specific VRE; UC CURA_01: Subject coverage |

| | | | | |
|---|-----------|----------------------------|---|--|
| logues of manuscripts | | | | |
| 4: Assess the presence of this work in catalogues of medieval libraries in Europe, to understand the popularity and circulation of this work in the medieval and early modern period by using TRAME tool. | Interpret | Dataset; Service; Software | TRAME – http://git-trame.fefonlus.it/ | UC VRESET_02: Integration of services in a domain-specific VRE; UC CURA_01: Subject coverage |
| 5: Ensure that the CERL Thesaurus is running at the back of the above listed tools to assure inclusiveness of data. | Connect | Dataset; Service; Software | CRL thesaurus | UC VRESET_02: Integration of services in a domain-specific VRE; UC CURA_01: Subject coverage |
| 6: Linking out to secondary literature on this work using TRAME and Bibliissima tools. | Connect | Dataset; Service; Software | TRAME – http://git-trame.fefonlus.it/ | UC VRESET_02: Integration of services in a domain-specific VRE |

3.7. Requirements for interoperability: Mapped requirements

3.7.1. Mapped requirements from Studies of the Past

3.7.1.1. KNAW-NIOD

| | |
|--|--|
| Provided by: KNAW-NIOD <i>Contributor(s): Annelies van Nispen</i> | |
| Document / filename (D4Science or Zotero): | CENDARI_ParticipatoryDesignWWI_report.pdf; CENDARI_D8.2_Functional Description.docx; CENDARI_D8.2_Functional_Description_Visualisation.doc |
| Community: | Studies of the Past; Heritage and Applied disciplines |
| User role: (Actor) | Researcher/archivist/librarian as VRE manager |
| Functionality / requirement: (Service) | Visualize search results |
| Explanation (NEED): (Software when present) | Being able to visualize objects and location |
| Priority level: | High |

| | |
|--------------------------------------|---|
| KGP Phase | Interpret |
| Entities | Service; Software |
| Service/ Tool used (optional) | 3D-visualisation; virtual reality and immersive environments |
| WP5-6 Use Case | UC VRESET_02: Integration of services in a domain specific VRE; UC VREUSE_05: Process a dataset and publish results |

3.7.1.2. KNAW-NIOD

| | |
|--|--|
| Provided by: KNAW-NIOD <i>Contributor(s): Annelies van Nispen</i> | |
| Document / filename (D4Science or Zotero): | CENDARI_ParticipatoryDesignWWI_report.pdf; CENDARI_D8.2_Functional Description.docx; CENDARI_D8.2_Functional_Description_Visualisation.doc |
| Community: | Studies of the Past; Heritage and Applied disciplines |
| User role: (Actor) | Researcher/archivist/librarian as VRE manager |
| Functionality / requirement: (Service) | Visualize search results |
| Explanation (NEED): (Software when present) | Being able to visualize search paths |
| Priority level: | High |
| KGP Phase | Interpret |
| Entities | Service/Software |
| Service/ Tool used (optional) | Geo-visualisation tool |
| WP5-6 Use Case | UC VRESET_02: Integration of services in a domain specific VRE; UC VREUSE_05: Process a dataset and publish results |

3.7.1.3. KNAW-NIOD

| | |
|--|--|
| Provided by: KNAW-NIOD <i>Contributor(s): Annelies van Nispen</i> | |
| Document / filename (D4Science or Zotero): | CENDARI_ParticipatoryDesignWWI_report.pdf; CENDARI_D8.2_Functional Description.docx; CENDARI_D8.2_Functional_Description_Visualisation.doc |
| Community: | Studies of the Past; Heritage and Applied disciplines |
| User role: (Actor) | Researcher/archivist/librarian as VRE manager |
| Functionality / requirement: (Service) | Being able to use tools to crowdsource translation of (archival) documents |
| KGP Phase | Interpret |
| Entities | Service; Software |
| Service/ Tool used (optional) | |
| WP5-6 Use Case | UC VRESET_02: Integration of services in a domain specific VRE; UC VREUSE_05: Process a dataset and publish results |

3.7.1.4. KNAW-NIOD

| | |
|--|--|
| Provided by: KNAW-NIOD <i>Contributor(s): Annelies van Nispen</i> | |
| Document / filename (D4Science or Zotero): | CENDARI_ParticipatoryDesignWWI_report.pdf; CENDARI_D8.2_Functional Description.docx; CENDARI_D8.2_Functional_Description_Visualisation.doc |
| Community: | Studies of the Past; Heritage and Applied disciplines |
| User role: (Actor) | Researcher/archivist/librarian as VRE manager |
| Functionality / requirement: (Service) | Visualize search results |
| Explanation (NEED): (Software when present) | Being able to map archive location and type in a tool |
| KGP Phase | Interpret |
| Entities | Service; Software |
| Service/ Tool used (optional) | Geo-search tool |
| WP5-6 Use Case | UC VRESET_02: Integration of services in a domain specific VRE; UC VREUSE_05: Process a dataset and publish results |

3.7.1.5. KNAW-NIOD

| | |
|--|--|
| Provided by: KNAW-NIOD <i>Contributor(s): Annelies van Nispen</i> | |
| Document / filename (D4Science or Zotero): | CENDARI_ParticipatoryDesignWWI_report.pdf; CENDARI_D8.2_Functional Description.docx; CENDARI_D8.2_Functional_Description_Visualisation.doc |
| Community: | Studies of the Past; Heritage and Applied disciplines |
| User role: (Actor) | Researcher/archivist/librarian as VRE manager |
| Functionality / requirement: (Service) | Visualize search results |
| Explanation (NEED): (Software when present) | The researcher can understand and display the spatial or chronological relationships between documents |
| Priority level: | High |
| KGP Phase | Connect |
| Entities | Service; Software; Dataset |
| Service/ Tool used (optional) | Geotime visualisation tool |
| WP5-6 Use Case | UC VRESET_02: Integration of services in a domain specific VRE; UC VREUSE_05: Process a dataset and publish results |

3.7.1.6. KNAW-NIOD

| | |
|--|--|
| Provided by: KNAW-NIOD <i>Contributor(s): Annelies van Nispen</i> | |
| Document / filename | CENDARI_ParticipatoryDesignWWI_report.pdf; CENDARI_D8.2_Functional |

| | |
|--|---|
| (D4Science or Zotero): | Description.docx; CENDARI_D8.2_Functional_Description_Visualisation.doc |
| Community: | Studies of the Past; Heritage and Applied disciplines |
| User role: (Actor) | Researcher/archivist/librarian as VRE manager |
| Functionality / requirement: (Service) | Visualize search results |
| Explanation (NEED): (Software when present) | The researcher can understand and present the spatial and chronological relationships between documents |
| Priority level: | High |
| KGP Phase | Present |
| Entities | Service; Software; Dataset |
| Service/ Tool used (optional) | 3D-visualization |
| WP5-6 Use Case | UC VRESET_02: Integration of services in a domain specific VRE; UC VREUSE_05: Process a dataset and publish results |

3.7.1.7. KNAW-NIOD

| | |
|--|--|
| Provided by: KNAW-NIOD <i>Contributor(s): Annelies van Nispen</i> | |
| Document / filename (D4Science or Zotero): | CENDARI_ParticipatoryDesignWWI_report.pdf; CENDARI_D8.2_Functional_Description.docx; CENDARI_D8.2_Functional_Description_Visualisation.doc |
| Community: | Studies of the Past; Heritage and Applied disciplines |
| User role: (Actor) | Researcher/archivist/librarian as VRE manager |
| Functionality / requirement: (Service) | Citation of dataset |
| Explanation (NEED): (Software when present) | Researcher can use CENDARI material in a presentation or publication without having to figure out citation format |
| Priority level: | High |
| KGP Phase | Present |
| Entities | Service; Dataset |
| Service/ Tool used (optional) | Citation tool |
| WP5-6 Use Case | UC VRESET_02: Integration of services in a domain specific VRE; UC VREUSE_05: Process a dataset and publish results |

3.7.1.8. KNAW-NIOD

| | |
|--|--|
| Provided by: KNAW-NIOD <i>Contributor(s): Annelies van Nispen</i> | |
| Document / filename (D4Science or Zotero): | CENDARI_ParticipatoryDesignWWI_report.pdf; CENDARI_D8.2_Functional_Description.docx; CENDARI_D8.2_Functional_Description_Visualisation.doc |
| Community: | Studies of the Past; Heritage and Applied disciplines |
| User role: | Researcher/archivist/librarian as VRE manager |

| | |
|--|---|
| (Actor) | |
| Functionality / requirement: (Service) | Planning of research |
| Explanation (NEED): (Software when present) | Create a schedule/calendar for how much time the researcher will need for each archive, when the archives are open, and their contact details, national/religious holidays when they will be closed, etc. Link the results of my searches (archives I want to visit and when) to real-world information for planning (calendars, travel websites for airline, fares and train reservations, hotels, etc.) |
| KGP Phase | Collect |
| Entities | Service; Dataset; Actor |
| Service/ Tool used (optional) | "Voyager travel agent application" |
| WP5-6 Use Case | UC VRESET_02: Integration of services in a domain specific VRE; UC VREUSE_05: Process a dataset and publish results |

3.7.1.9. KNAW-NIOD

| | |
|--|---|
| Provided by: KNAW-NIOD <i>Contributor(s): Annelies van Nispen</i> | |
| Document / filename (D4Science or Zotero): | EHRI D17.1 Report on standards including survey.pdf; EHRI D17.2 Metadata schema for the portal site.pdf |
| Community: | Studies of the Past; Heritage and Applied disciplines |
| User role: (Actor) | Research Infrastructure admin |
| Functionality / requirement: (Service) | Enable (federated) search services |
| Explanation (NEED): (Software when present) | Being able to search (keywords, persons, location/geographical information, events, time/dates...) |
| Priority level: | High |
| KGP Phase | Connect |
| Entities | Service; Research Infrastructure |
| Service/ Tool used (optional) | |
| WP5-6 Use Case | UC ACCESS_02: Search and browse the Parthenos content cloud across several research infrastructures |

3.7.1.10. KNAW-NIOD

| | |
|--|---|
| Provided by: KNAW-NIOD <i>Contributor(s): Annelies van Nispen</i> | |
| Document / filename (D4Science or Zotero): | EHRI D17.1 Report on standards including survey.pdf; EHRI D17.2 Metadata schema for the portal site.pdf |
| Community: | Studies of the Past; Heritage and Applied disciplines |
| User role: (Actor) | Research Infrastructure admin |

| | |
|--|---|
| Explanation (NEED): (Software when present) | Being able to upload/harvest/integrate data from CHI into EHRI |
| Priority level: | High |
| Macro functionality: | Data interoperability and data integration |
| KGP Phase | Connect |
| Entities | Service; Research Infrastructure |
| Service/ Tool used (optional) | |
| WP5-6 Use Case | UC REG_02: A research infrastructure joins Parthenos and integrates its registry; UC CURA_02: Invite new content providers; UC CURA_03: Invite curation |

3.7.1.11. KNAW-NIOD

| | |
|--|---|
| Provided by: KNAW-NIOD <i>Contributor(s): Annelies van Nispen</i> | |
| Document / filename (D4Science or Zotero): | EHRI D17.1 Report on standards including survey.pdf; EHRI D17.2 Metadata schema for the portal site.pdf |
| Community: | Studies of the Past; Heritage and Applied disciplines |
| User role: (Actor) | Research Infrastructure admin |
| Explanation (NEED): (Software when present) | Being able to share and collaborate with other researchers on EHRI documents |
| Priority level: | High |
| Macro functionality: | Data integration |
| KGP Phase | Connect |
| Entities | Service; Dataset; Research Infrastructure |
| Service/ Tool used (optional) | |
| WP5-6 Use Case | UC VREUSE_04: Private and public sharing of resources deposited in the VRE workspace |

3.7.1.12. KNAW-NIOD

| | |
|--|---|
| Provided by: KNAW-NIOD <i>Contributor(s): Annelies van Nispen</i> | |
| Document / filename (D4Science or Zotero): | EHRI D17.1 Report on standards including survey.pdf; EHRI D17.2 Metadata schema for the portal site.pdf |
| Community: | Studies of the Past; Heritage and Applied disciplines |
| User role: (Actor) | Research Infrastructure admin |
| Explanation (NEED): (Software when present) | Being able to present data in an online portal |
| Priority level: | High |
| KGP Phase | Present |
| Entities | Service; Dataset; Research Infrastructure |

| | |
|--------------------------------------|---|
| Service/ Tool used (optional) | |
| WP5-6 Use Case | UC VREUSE_05: Process a dataset and publish results |

3.7.1.13. KNAW-NIOD

| | |
|--|---|
| Provided by: KNAW-NIOD <i>Contributor(s): Annelies van Nispen</i> | |
| Document / filename (D4Science or Zotero): | EHRI D17.1 Report on standards including survey.pdf; EHRI D17.2 Metadata schema for the portal site.pdf |
| Community: | Studies of the Past; Heritage and Applied disciplines |
| User role: (Actor) | Research Infrastructure admin; access management |
| Explanation (NEED): (Software when present) | Services to be able to authenticate and identify users and set authorization levels |
| Priority level: | High |
| KGP Phase | Connect |
| Entities | Service; Research Infrastructure |
| Service/ Tool used (optional) | |
| WP5-6 Use Case | UC VRESET_03: VRE authentication and authorization |

3.7.2. Mapped requirements from Social Sciences

3.7.2.1. KNAW-DANS

| | |
|---|--|
| Provided by: KNAW-DANS <i>Contributor(s): Emilie Kraaikamp</i> | |
| Document / filename (D4Science or Zotero): | https://www.zotero.org/groups/parthenos_wp2/items/itemKey/KKDFBJ7D ; https://www.zotero.org/groups/parthenos_wp2/items/itemKey/Q9XI2PU3 . |
| Community: | All |
| User role: (Actor) | VRE Manager |
| Functionality / requirement: (Service) | Supporting different back-ends for data storage |
| Explanation (NEED): (Software when present) | The framework should provide a storage management module for the configuration of the storage back-ends to be used. Depending on the functional requirements of the target Enhanced Publication Information System (EPIS), a type of back-end, may be preferable to another. |
| Priority level: | High |
| KGP Phase | Interpret |
| Entities | Service; Software |
| Service/ Tool used (optional) | |
| WP5-6 Use Case | UC VRESET_02: Integration of services in a domain-specific VRE |

3.7.2.2. KNAW-DANS

| | |
|---|--|
| Provided by: KNAW-DANS <i>Contributor(s): Emilie Kraaikamp</i> | |
| Document / filename (D4Science or Zotero): | https://www.zotero.org/groups/parthenos_wp2/items/itemKey/KKDFBJ7D ; https://www.zotero.org/groups/parthenos_wp2/items/itemKey/Q9XI2PU3 . |
| Community: | All |
| User role: (Actor) | VRE Manager |
| Functionality / requirement: (Service) | Offering data definition, manipulation, and access languages |
| Explanation (NEED): (Software when present) | The framework should provide a language for the definition of EP data models (EP-DMDL, EP Data Model Definition Language) |
| Priority level: | High |
| KGP Phase | Interpret |
| Entities | Dataset; Service |
| Service/ Tool used (optional) | |
| WP5-6 Use Case | UC VRESET_02: Integration of services in a domain-specific VRE |

3.7.2.3. KNAW-DANS

| | |
|---|---|
| Provided by: KNAW-DANS <i>Contributor(s): Emilie Kraaikamp</i> | |
| Document / filename (D4Science or Zotero): | https://www.zotero.org/groups/parthenos_wp2/items/itemKey/KKDFBJ7D |
| Community: | All |
| User role: (Actor) | A developer of Enhanced Publication Information Systems (EPIs) as VRE user |
| Functionality / requirement: (Service) | Able to operate on EP instances (compliant to the defined EP data model) with a dedicated domain-specific language |
| Explanation (NEED): (Software when present) | Making manipulation of resources possible whose types are defined in the EP data model (EP-DSML, EP Domain Specific Manipulation Language). |
| Priority level: | High |
| KGP Phase | Interpret |
| Entities | Service; Software |
| Service/ Tool used (optional) | |
| WP5-6 Use Case | UC VRESET_02: Integration of services in a domain-specific VRE; UC CURA_01: Subject coverage |

3.7.2.4. KNAW-DANS

| | |
|---|---|
| Provided by: KNAW-DANS <i>Contributor(s): Emilie Kraaikamp</i> | |
| Document / filename | https://www.zotero.org/groups/parthenos_wp2/items/itemKey/KKDFBJ7D ; |

| | |
|--|---|
| (D4Science or Zotero): | https://www.zotero.org/groups/parthenos_wp2/items/itemKey/Q9XI2PU3 . |
| Community: | All |
| User role: (Actor) | VRE Manager |
| Functionality / requirement: (Service) | Enabling data sharing |
| Explanation (NEED): (Software when present) | Need for supporting the export of content via different standard APIs and protocols to serve third-party applications. |
| Priority level: | High |
| KGP Phase | Interpret |
| Entities | Service; Dataset |
| Service/ Tool used (optional) | |
| WP5-6 Use Case | UC VREUSE_04: Private and public sharing of resources deposited in the VRE network |

3.7.2.5. KNAW-DANS

| | |
|---|--|
| Provided by: KNAW-DANS Contributor(s): <i>Emilie Kraaikamp</i> | |
| Document / filename (D4Science or Zotero): | https://www.zotero.org/groups/parthenos_wp2/items/itemKey/KKDFBJ7D ; https://www.zotero.org/groups/parthenos_wp2/items/itemKey/Q9XI2PU3 . |
| Community: | All |
| User role: (Actor) | VRE Manager |
| Functionality / requirement: (Service) | Supporting data portability |
| Explanation (NEED): (Software when present) | Support is needed for open standards for the representation of data |
| Priority level: | High |
| KGP Phase | Interpret |
| Entities | Service; Dataset |
| Service/ Tool used (optional) | |
| WP5-6 Use Case | UC VREUSE_04: Private and public sharing of resources deposited in the VRE network |

3.7.2.6. KNAW-DANS

| | |
|---|--|
| Provided by: KNAW-DANS Contributor(s): <i>Emilie Kraaikamp</i> | |
| Document / filename (D4Science or Zotero): | https://www.zotero.org/groups/parthenos_wp2/items/itemKey/KKDFBJ7D ; https://www.zotero.org/groups/parthenos_wp2/items/itemKey/Q9XI2PU3 . |
| Community: | All |
| User role: | VRE Manager |

| | |
|--|--|
| (Actor) | |
| Functionality / requirement: (Service) | Supporting the integration of heterogeneous data sources |
| Explanation (NEED): (Software when present) | Data sources export different typologies of content according to different formats and via different protocols. EPMSs should support developers in the integration of such diverse content. |
| Priority level: | High |
| KGP Phase | Connect |
| Entities | Dataset; Service; Actor |
| Service/ Tool used (optional) | |
| WP5-6 Use Case | UC AGGR_01: Aggregate resource metadata from research infrastructures into the Parthenos content cloud; UC REG_02: A research infrastructure joins Parthenos and integrates its registry; UC CURA_01: Subject coverage |

3.7.2.7. KNAW-DANS

| | |
|---|--|
| Provided by: KNAW-DANS <i>Contributor(s): Emilie Kraaikamp</i> | |
| Document / filename (D4Science or Zotero): | https://www.zotero.org/groups/parthenos_wp2/items/itemKey/KKDFBJ7D . |
| Community: | All |
| User role: (Actor) | VRE Manager |
| Functionality / requirement: (Service) | Support the management of dynamic data sources. |
| Explanation (NEED): (Software when present) | Data source management functionality is needed to ease the administrative operations needed to take care of the dynamic nature of the data sources. |
| Priority level: | High |
| KGP Phase | Interpret |
| Entities | Dataset; Service |
| Service/ Tool used (optional) | |
| WP5-6 Use Case | UC AGGR_01: Aggregate resource metadata from research infrastructures into the Parthenos content cloud; UC REG_02: A research infrastructure joins Parthenos and integrates its registry; UC CURA_01: Subject coverage |

3.7.2.8. KNAW-DANS

| | |
|---|---|
| Provided by: KNAW-DANS <i>Contributor(s): Emilie Kraaikamp</i> | |
| Document / filename (D4Science or Zotero): | https://www.zotero.org/groups/parthenos_wp2/items/itemKey/KKDFBJ7D . |
| Community: | All |
| User role: (Actor) | VRE Manager |

| | |
|--|--|
| Functionality / requirement: (Service) | Support the integration of content |
| Explanation (NEED): (Software when present) | Due to the heterogeneity of the content, a transformation and harmonization module is necessary in order to massage the incoming material and transform it in a homogeneous format, so that further operations can be performed on content without tackling again the peculiarities of each data source. |
| Priority level: | High |
| KGP Phase | Connect |
| Entities | Dataset; Service |
| Service/ Tool used (optional) | |
| WP5-6 Use Case | UC AGGR_01: Aggregate resource metadata from research infrastructures into the Parthenos content cloud |

3.7.2.9. KNAW-DANS

| | |
|---|--|
| Provided by: KNAW-DANS <i>Contributor(s): Emilie Kraaikamp</i> | |
| Document / filename (D4Science or Zotero): | https://www.zotero.org/groups/parthenos_wp2/items/itemKey/KKDFBJ7D ; https://www.zotero.org/groups/parthenos_wp2/items/itemKey/Q9XI2PU3 . |
| Community: | All |
| User role: (Actor) | VRE Manager |
| Functionality / requirement: (Service) | Enable the customization of the EP data model |
| Explanation (NEED): (Software when present) | Tools are needed for the definition of EP data models |
| Priority level: | High |
| KGP Phase | Interpret |
| Entities | Service; Software |
| Service/ Tool used (optional) | |
| WP5-6 Use Case | UC VRESET_02: Integration of services in a domain-specific VRE |

3.7.2.10. KNAW-DANS

| | |
|---|--|
| Provided by: KNAW-DANS <i>Contributor(s): Emilie Kraaikamp</i> | |
| Document / filename (D4Science or Zotero): | https://www.zotero.org/groups/parthenos_wp2/items/itemKey/KKDFBJ7D ; https://www.zotero.org/groups/parthenos_wp2/items/itemKey/Q9XI2PU3 . |
| Community: | All |
| User role: (Actor) | VRE Manager |
| Functionality / requirement: | Support the enrichment and curation of content |

| | |
|--|--|
| (Service) | |
| Explanation (NEED): (Software when present) | To create a high quality content it is needed to better the quality of the EPs and enrich the original content |
| Priority level: | High |
| KGP Phase | Interpret |
| Entities | Dataset; Service |
| Service/ Tool used (optional) | |
| WP5-6 Use Case | UC CURA_02: Invite new content providers; UC CURA_03: Invite curation |

3.7.2.11. KNAW-DANS

| | |
|---|---|
| Provided by: KNAW-DANS <i>Contributor(s): Emilie Kraaikamp</i> | |
| Document / filename (D4Science or Zotero): | https://www.zotero.org/groups/parthenos_wp2/items/itemKey/Q9XI2PU3 . |
| Community: | All |
| User role: (Actor) | VRE Manager |
| Functionality / requirement: (Service) | Supporting the addition of new domain-specific functionalities |
| Explanation (NEED): (Software when present) | Based on the requirements of existing EPISs |
| Priority level: | High |
| KGP Phase | Interpret |
| Entities | Service; Software |
| Service/ Tool used (optional) | |
| WP5-6 Use Case | UC VRESET_02: Integration of services in a domain-specific VRE |

3.7.2.12. KNAW-DANS

| | |
|---|---|
| Provided by: KNAW-DANS <i>Contributor(s): Emilie Kraaikamp</i> | |
| Document / filename (D4Science or Zotero): | EHRI D16.4 Researcher Practices and User Requirements.pdf |
| Community: | Studies of the Past |
| User role: (Actor) | Content Provider |
| Functionality / requirement: (Service) | Confidentiality |
| Explanation (NEED): (Software when present) | Requirements are needed which state that some sensitive information may not be disclosed to unauthorized parties. |
| Function ID: | UR1 |

| | |
|--------------------------------------|--|
| KGP Phase | Connect |
| Entities | Dataset; Service |
| Service/ Tool used (optional) | |
| WP5-6 Use Case | UC VRESET_03: VRE authentication and authorization |

3.7.2.13. KNAW-DANS

| | |
|---|---|
| Provided by: KNAW-DANS <i>Contributor(s): Emilie Kraaikamp</i> | |
| Document / filename (D4Science or Zotero): | EHRI D16.4 Researcher Practices and User Requirements.pdf |
| Community: | Studies of the Past |
| User role: (Actor) | Content Provider |
| Functionality / requirement: (Service) | Availability |
| Explanation (NEED): (Software when present) | Requirements are needed which state that some information or resource can be used at any point in time when it is needed and its usage is authorized. |
| Function ID: | UR2 |
| KGP Phase | Connect |
| Entities | Dataset; Service |
| Service/ Tool used (optional) | |
| WP5-6 Use Case | UC VRESET_03: VRE authentication and authorization |

3.7.2.14. KNAW-DANS

| | |
|---|--|
| Provided by: KNAW-DANS <i>Contributor(s): Emilie Kraaikamp</i> | |
| Document / filename (D4Science or Zotero): | EHRI D16.4 Researcher Practices and User Requirements.pdf |
| Community: | Studies of the Past |
| User role: (Actor) | Content Provider |
| Functionality / requirement: (Service) | Reliability |
| Explanation (NEED): (Software when present) | Requirements are needed which constrain the software to operate as expected over long periods of time. |
| Function ID: | UR3 |
| KGP Phase | Interpret |
| Entities | Service; Software |
| Service/ Tool used (optional) | |
| WP5-6 Use Case | UC CURA_03: Invite curation |

3.7.2.15. KNAW-DANS

| | |
|---|--|
| Provided by: KNAW-DANS Contributor(s): <i>Emilie Kraaikamp</i> | |
| Document / filename (D4Science or Zotero): | EHRI D16.4 Researcher Practices and User Requirements.pdf |
| Community: | Studies of the Past |
| User role: (Actor) | Content Provider |
| Functionality / requirement: (Service) | Accuracy |
| Explanation (NEED): (Software when present) | Requirements are needed which constrain the state of the information processed by the software to reflect the state of the corresponding physical information in the environment accurately. |
| Function ID: | UR4 |
| KGP Phase | Connect |
| Entities | Service; Software |
| Service/ Tool used (optional) | |
| WP5-6 Use Case | UC VRESET_02: Integration of services in a domain-specific VRE; UC CURA_02: Invite new content providers |

3.7.2.16. KNAW-DANS

| | |
|---|---|
| Provided by: KNAW-DANS Contributor(s): <i>Emilie Kraaikamp</i> | |
| Document / filename (D4Science or Zotero): | EHRI D16.4 Researcher Practices and User Requirements.pdf |
| Community: | Studies of the Past |
| User role: (Actor) | Content Provider |
| Functionality / requirement: (Service) | Usability |
| Explanation (NEED): (Software when present) | For human interaction, usability requirements are needed which prescribe input/output formats and user dialogues to fit the abstractions, abilities and expectations of the target users. |
| Function ID: | UR5 |
| KGP Phase | Interpret |
| Entities | Service; Software; Actor |
| Service/ Tool used (optional) | |
| WP5-6 Use Case | UC VRESET_02: Integration of services in a domain-specific VRE |

3.7.2.17. KNAW-DANS

| | |
|---|---|
| Provided by: KNAW-DANS Contributor(s): <i>Emilie Kraaikamp</i> | |
| Document / filename | EHRI D16.4 Researcher Practices and User Requirements.pdf |

| | |
|--|--|
| (D4Science or Zotero): | |
| Community: | Studies of the Past |
| User role: (Actor) | Content Provider |
| Functionality / requirement: (Service) | Architectural |
| Explanation (NEED): (Software when present) | Requirements are needed which impose structural constraints on the software-to-be to fit its environment |
| Function ID: | UR6 |
| KGP Phase | Interpret |
| Entities | Service; Software |
| Service/ Tool used (optional) | |
| WP5-6 Use Case | UC VRESET_02: Integration of services in a domain-specific VRE |

3.7.2.18. KNAW-DANS

| | |
|---|--|
| Provided by: KNAW-DANS <i>Contributor(s): Emilie Kraaikamp</i> | |
| Document / filename (D4Science or Zotero): | EHRI D16.5 Data Requirements.pdf |
| Community: | Studies of the Past |
| User role: (Actor) | Content Provider |
| Functionality / requirement: (Service) | Archival information on collections |
| Explanation (NEED): (Software when present) | Providing as much information as possible about archives that hold collections of interest helps researchers to be prepared |
| KGP Phase | Connect |
| Entities | Service; Dataset; Actor |
| Service/ Tool used (optional) | |
| WP5-6 Use Case | UC ACCESS_03: Retrieval/access metadata about an entity of the Parthenos registry or a resource in the Parthenos content cloud; UC CURA_02: Invite new content providers |

3.7.2.19. KNAW-DANS

| | |
|---|----------------------------------|
| Provided by: KNAW-DANS <i>Contributor(s): Emilie Kraaikamp</i> | |
| Document / filename (D4Science or Zotero): | EHRI D16.5 Data Requirements.pdf |
| Community: | Studies of the Past |
| User role: (Actor) | Content Provider |

| | |
|--|--|
| Functionality / requirement: (Service) | Archival information on how archives manage and describe the holdings |
| Explanation (NEED): (Software when present) | Providing this information helps researchers to be prepared for working in an archive |
| KGP Phase | Connect |
| Entities | Service; Dataset; Actor |
| Service/ Tool used (optional) | |
| WP5-6 Use Case | UC ACCESS_03: Retrieval/access metadata about an entity of the Parthenos registry or a resource in the Parthenos content cloud |

3.7.2.20. KNAW-DANS

| | |
|---|---|
| Provided by: KNAW-DANS <i>Contributor(s): Emilie Kraaikamp</i> | |
| Document / filename (D4Science or Zotero): | EHRI D16.5 Data Requirements.pdf |
| Community: | Studies of the Past |
| User role: (Actor) | Content Provider |
| Functionality / requirement: (Service) | As much archival information as possible on archival holdings |
| Explanation (NEED): (Software when present) | Providing this information helps to enable the researchers to undertake an initial assessment of the value of the archival holdings for their research |
| KGP Phase | Connect |
| Entities | Service; Dataset; Actor |
| Service/ Tool used (optional) | |
| WP5-6 Use Case | UC ACCESS_03: Retrieval/access metadata about an entity of the Parthenos registry or a resource in the Parthenos content cloud; UC CURA_01: Subject coverage; UC CURA_03: Invite curation |

3.7.2.21. KNAW-DANS

| | |
|---|---|
| Provided by: KNAW-DANS <i>Contributor(s): Emilie Kraaikamp</i> | |
| Document / filename (D4Science or Zotero): | EHRI D16.5 Data Requirements.pdf |
| Community: | Studies of the Past |
| User role: (Actor) | Content Provider |
| Functionality / requirement: (Service) | Facilitate sharing, categorising, and indexing research questions and/or topics |
| Explanation (NEED): | Users of EHRI would benefit from having access to these along with the |

| | |
|--------------------------------------|---|
| (Software when present) | sources selected to assist in answering a question or addressing a particular topic. |
| KGP Phase | Connect |
| Entities | Service |
| Service/ Tool used (optional) | |
| WP5-6 Use Case | UC VRESET_01: Set-up of a domain-specific VRE; UC CURA_02: Invite new content providers |

3.7.2.22. KNAW-DANS

| | |
|---|---|
| Provided by: KNAW-DANS <i>Contributor(s): Emilie Kraaikamp</i> | |
| Document / filename (D4Science or Zotero): | EHRI D16.5 Data Requirements.pdf |
| Community: | Studies of the Past |
| User role: (Actor) | Content Provider |
| Functionality / requirement: (Service) | Facilitate sharing, categorising, and indexing additional information about a research project |
| Explanation (NEED): (Software when present) | Users of EHRI would benefit from having access to these along with the sources selected to assist in answering a question or addressing a particular topic. |
| KGP Phase | Connect |
| Entities | Service |
| Service/ Tool used (optional) | |
| WP5-6 Use Case | UC VRESET_01: Set-up of a domain-specific VRE; UC CURA_02: Invite new content providers |

3.7.2.23. KNAW-DANS

| | |
|---|---|
| Provided by: KNAW-DANS <i>Contributor(s): Emilie Kraaikamp</i> | |
| Document / filename (D4Science or Zotero): | EHRI D16.5 Data Requirements.pdf |
| Community: | Studies of the Past |
| User role: (Actor) | Content Provider |
| Functionality / requirement: (Service) | Facilitate sharing, categorising, and indexing notes and annotations on sources at various levels |
| Explanation (NEED): (Software when present) | They are perceived as valuable for research |
| Priority level: | Medium |
| KGP Phase | Connect |
| Entities | Service |

| | |
|--------------------------------------|---|
| Service/ Tool used (optional) | |
| WP5-6 Use Case | UC VRESET_01: Set-up of a domain-specific VRE; UC CURA_02: Invite new content providers |

3.7.2.24. KNAW-DANS

| | |
|---|--|
| Provided by: KNAW-DANS <i>Contributor(s): Emilie Kraaikamp</i> | |
| Document / filename (D4Science or Zotero): | EHRI D16.5 Data Requirements.pdf |
| Community: | Studies of the Past |
| User role: (Actor) | Content Provider |
| Functionality / requirement: (Service) | Facilitate sharing, categorising, and indexing details (citations) of researchers' publications |
| Explanation (NEED): (Software when present) | To assist in the 'chaining' process of moving from published works to other works, and to archival sources |
| KGP Phase | Connect |
| Entities | Service; Dataset |
| Service/ Tool used (optional) | |
| WP5-6 Use Case | UC VRESET_01: Set-up of a domain-specific VRE; UC CURA_02: Invite new content providers |

3.7.2.25. KNAW-DANS

| | |
|---|---|
| Provided by: KNAW-DANS <i>Contributor(s): Emilie Kraaikamp</i> | |
| Document / filename (D4Science or Zotero): | EHRI D16.5 Data Requirements.pdf |
| Community: | Studies of the Past |
| User role: (Actor) | Content Provider |
| Functionality / requirement: (Service) | Facilitate sharing, categorising, and indexing of researcher bibliographies |
| KGP Phase | Connect |
| Entities | Service; Dataset; Actor |
| Service/ Tool used (optional) | |
| WP5-6 Use Case | UC VRESET_01: Set-up of a domain-specific VRE; UC CURA_02: Invite new content providers |

3.7.2.26. KNAW-DANS

| | |
|---|--|
| Provided by: KNAW-DANS <i>Contributor(s): Emilie Kraaikamp</i> | |
|---|--|

| | |
|--|---|
| Document / filename (D4Science or Zotero): | CENDARI_D8.1 Functional description, portal and VRE final.pdf |
| Community: | Studies of the Past |
| User role: (Actor) | Content Provider |
| Functionality / requirement: (Service) | Tool for finding sources |
| Explanation (NEED): (Software when present) | Studies reveal that functions facilitating early research for example, finding, organizing, and displaying sources are the most used and sought after in the scholarly community. |
| KGP Phase | Interpret |
| Entities | Service; Dataset |
| Service/ Tool used (optional) | |
| WP5-6 Use Case | UC ACCESS_02: Search and browse the Parthenos content cloud across several research infrastructures |

3.7.2.27. KNAW-DANS

| | |
|---|--|
| Provided by: KNAW-DANS Contributor(s): <i>Emilie Kraaikamp</i> | |
| Document / filename (D4Science or Zotero): | CENDARI_D8.1 Functional description, portal and VRE final.pdf |
| Community: | Studies of the Past |
| User role: (Actor) | Content Provider |
| Functionality / requirement: (Service) | Tool for organizing resource |
| Explanation (NEED): (Software when present) | Studies reveal that functions facilitating early research (for example, finding, organizing, and displaying sources) are the most used and sought after in the scholarly community. This allows the user to take dynamic notes, organize them in useful ways, and link her/his research to data in the CENDARI data space. |
| KGP Phase | Collect |
| Entities | Dataset; Service |
| Service/ Tool used (optional) | |
| WP5-6 Use Case | UC VRESET_01: Set-up of a domain-specific VRE; UC CURA_02: Invite new content providers |

3.7.2.28. KNAW-DANS

| | |
|---|---|
| Provided by: KNAW-DANS Contributor(s): <i>Emilie Kraaikamp</i> | |
| Document / filename (D4Science or Zotero): | CENDARI_D8.1 Functional description, portal and VRE final.pdf |
| Community: | Studies of the Past |

| | |
|--|---|
| User role: (Actor) | Content Provider |
| Functionality / requirement: (Service) | Tool for displaying own research |
| Explanation (NEED): (Software when present) | Studies reveal that functions facilitating early research (for example, finding, organizing, and displaying sources) are the most used and sought after in the scholarly community. This not only displays the user's research in provoking ways, but can also reveal connections and patterns that may inform the conclusions of his/her research or guide further research. |
| KGP Phase | Present |
| Entities | Dataset; Service |
| Service/ Tool used (optional) | |
| WP5-6 Use Case | UC VRESET_01: Set-up of a domain-specific VRE; UC VRESET_02: Integration of services in a domain-specific VRE; UC VREUSE_05: Process a dataset and publish results |

3.7.2.29. KNAW-DANS

| | |
|---|--|
| Provided by: KNAW-DANS <i>Contributor(s): Emilie Kraaikamp</i> | |
| Document / filename (D4Science or Zotero): | CENDARI_D8.1 Functional description, portal and VRE final.pdf |
| Community: | Studies of the Past |
| User role: (Actor) | Content Provider |
| Functionality / requirement: (Service) | Tool for accessing existing data resources (published or not) |
| Explanation (NEED): (Software when present) | Studies reveal that functions facilitating early research (for example, finding, organizing, and displaying sources) are the most used and sought after in the scholarly community. |
| KGP Phase | Connect |
| Entities | Dataset; Service |
| Service/ Tool used (optional) | |
| WP5-6 Use Case | (Published) UC ACCESS_04: Retrieval/access of resources from the Parthenos content cloud; (Not published) UC VREUSE_04: Private and public sharing of resources deposited in the VRE workspace |

3.7.2.30. KNAW-DANS

| | |
|---|---|
| Provided by: KNAW-DANS <i>Contributor(s): Emilie Kraaikamp</i> | |
| Document / filename (D4Science or Zotero): | CENDARI_D8.1 Functional description, portal and VRE final.pdf |
| Community: | Studies of the Past |
| User role: | Content Provider |

| | |
|--|---|
| (Actor) | |
| Functionality / requirement: (Service) | Tool for productively connecting with other researchers |
| Explanation (NEED): (Software when present) | Studies reveal that functions facilitating early research (for example, finding, organizing, and displaying sources) are the most used and sought after in the scholarly community. |
| KGP Phase | Connect |
| Entities | Service; Actor |
| Service/ Tool used (optional) | |
| WP5-6 Use Case | UC VRESET_01: Set-up of a domain-specific VRE; UC VRESET_02: Integration of services in a domain-specific VRE |

3.7.2.31. KNAW-DANS

| | |
|---|---|
| Provided by: KNAW-DANS <i>Contributor(s): Emilie Kraaikamp</i> | |
| Document / filename (D4Science or Zotero): | CENDARI_D8.1 Functional description, portal and VRE final.pdf |
| Community: | Studies of the Past |
| User role: (Actor) | Content Provider |
| Functionality / requirement: (Service) | Tools to search and browse at a general level |
| Explanation (NEED): (Software when present) | To find institutions and collections and other information |
| KGP Phase | Connect |
| Entities | Service; Software |
| Service/ Tool used (optional) | |
| WP5-6 Use Case | UC ACCESS_01: Search and browse the Parthenos registry |

3.7.2.32. KNAW-DANS

| | |
|---|---|
| Provided by: KNAW-DANS <i>Contributor(s): Emilie Kraaikamp</i> | |
| Document / filename (D4Science or Zotero): | CENDARI_D8.1 Functional description, portal and VRE final.pdf |
| Community: | Studies of the Past |
| User role: (Actor) | Content Provider |
| Functionality / requirement: (Service) | Tools to search and browse at a more detailed level |
| Explanation (NEED): (Software when present) | To find detailed information on a given subject/work |

| | |
|--------------------------------------|---|
| sent) | |
| KGP Phase | Connect |
| Entities | Service; Software |
| Service/ Tool used (optional) | |
| WP5-6 Use Case | UC ACCESS_02: Search and browse the Parthenos content cloud across several research infrastructures |

3.7.2.33. KNAW-DANS

| | |
|---|---|
| Provided by: KNAW-DANS Contributor(s): <i>Emilie Kraaikamp</i> | |
| Document / filename (D4Science or Zotero): | CENDARI_D8.1 Functional description, portal and VRE final.pdf |
| Community: | Studies of the Past |
| User role: (Actor) | Content Provider |
| Functionality / requirement: (Service) | Provide indication of the language and the place in which works were written (origin/provenance) |
| Explanation (NEED): (Software when present) | To help the user in their research work |
| KGP Phase | Connect |
| Entities | Service; Actor |
| Service/ Tool used (optional) | |
| WP5-6 Use Case | UC ACCESS_02: Search and browse the Parthenos content cloud across several research infrastructures; UC CURA_01: Subject coverage |

3.7.2.34. KNAW-DANS

| | |
|---|---|
| Provided by: KNAW-DANS Contributor(s): <i>Emilie Kraaikamp</i> | |
| Document / filename (D4Science or Zotero): | CENDARI_D8.1 Functional description, portal and VRE final.pdf |
| Community: | Studies of the Past |
| User role: (Actor) | Content Provider |
| Functionality / requirement: (Service) | Provide the year of composition of a work |
| Explanation (NEED): (Software when present) | To help the user in their research work |
| KGP Phase | Connect |
| Entities | Service |
| Service/ Tool used (optional) | |

| | |
|-----------------------|---|
| WP5-6 Use Case | UC ACCESS_02: Search and browse the Parthenos content cloud across several research infrastructures; UC CURA_01: Subject coverage |
|-----------------------|---|

3.7.2.35. KNAW-DANS

| | |
|---|---|
| Provided by: KNAW-DANS <i>Contributor(s): Emilie Kraaikamp</i> | |
| Document / filename (D4Science or Zotero): | CENDARI_D8.1 Functional description, portal and VRE final.pdf |
| Community: | Studies of the Past |
| User role: (Actor) | Content Provider |
| Functionality / requirement: (Service) | Show the availability of printed editions of a work |
| Explanation (NEED): (Software when present) | To help the user in their research work |
| KGP Phase | Connect |
| Entities | Service; Dataset |
| Service/ Tool used (optional) | |
| WP5-6 Use Case | UC ACCESS_02: Search and browse the Parthenos content cloud across several research infrastructures; UC CURA_01: Subject coverage |

3.7.2.36. KNAW-DANS

| | |
|---|---|
| Provided by: KNAW-DANS <i>Contributor(s): Emilie Kraaikamp</i> | |
| Document / filename (D4Science or Zotero): | CENDARI_D8.1 Functional description, portal and VRE final.pdf |
| Community: | Studies of the Past |
| User role: (Actor) | Content Provider |
| Functionality / requirement: (Service) | Show manuscripts related to a work |
| Explanation (NEED): (Software when present) | To help the user in their research work |
| KGP Phase | Connect |
| Entities | |
| Service/ Tool used (optional) | Service; Dataset |
| WP5-6 Use Case | UC ACCESS_02: Search and browse the Parthenos content cloud across several research infrastructures; UC CURA_01: Subject coverage |

3.7.2.37. KNAW-DANS

| |
|------------------------|
| Provided by: KNAW-DANS |
|------------------------|

| | |
|--|---|
| <i>Contributor(s): Emilie Kraaikamp</i> | |
| Document / filename (D4Science or Zotero): | CENDARI_D8.1 Functional description, portal and VRE final.pdf |
| Community: | Studies of the Past |
| User role: (Actor) | Content Provider |
| Functionality / requirement: (Service) | Show author, place and time of translations of a work |
| Explanation (NEED): (Software when present) | To help the user in his research work |
| KGP Phase | Connect |
| Entities | Dataset; Service |
| Service/ Tool used (optional) | |
| WP5-6 Use Case | UC ACCESS_02: Search and browse the Parthenos content cloud across several research infrastructures; UC CURA_01: Subject coverage |

3.7.2.38. KNAW-DANS

| | |
|---|---|
| Provided by: KNAW-DANS <i>Contributor(s): Emilie Kraaikamp</i> | |
| Document / filename (D4Science or Zotero): | CENDARI_D8.1 Functional description, portal and VRE final.pdf |
| Community: | Studies of the Past |
| User role: (Actor) | Content Provider |
| Functionality / requirement: (Service) | Show availability of digital objects related to a work |
| Explanation (NEED): (Software when present) | To help the user in their research work |
| KGP Phase | Connect |
| Entities | Dataset; Service |
| Service/ Tool used (optional) | |
| WP5-6 Use Case | UC ACCESS_02: Search and browse the Parthenos content cloud across several research infrastructures; UC CURA_01: Subject coverage |

3.7.2.39. KNAW-DANS

| | |
|---|---|
| Provided by: KNAW-DANS <i>Contributor(s): Emilie Kraaikamp</i> | |
| Document / filename (D4Science or Zotero): | CENDARI_D8.1 Functional description, portal and VRE final.pdf |
| Community: | Studies of the Past |
| User role: | Content Provider |

| | |
|--|---|
| (Actor) | |
| Functionality / requirement: (Service) | Show a bibliography |
| Explanation (NEED): (Software when present) | To help the user in their research work |
| KGP Phase | Connect |
| Entities | Dataset; Service |
| Service/ Tool used (optional) | |
| WP5-6 Use Case | UC ACCESS_02: Search and browse the Parthenos content cloud across several research infrastructures; UC CURA_01: Subject coverage |

3.7.2.40. KNAW-DANS

| | |
|---|---|
| Provided by: KNAW-DANS <i>Contributor(s): Emilie Kraaikamp</i> | |
| Document / filename (D4Science or Zotero): | CENDARI_D8.1 Functional description, portal and VRE final.pdf |
| Community: | Studies of the Past |
| User role: (Actor) | Content Provider |
| Functionality / requirement: (Service) | Advanced tools for research and discovery |
| Explanation (NEED): (Software when present) | To enable new forms of research and discovery |
| KGP Phase | Interpret |
| Entities | Service; Software |
| Service/ Tool used (optional) | |
| WP5-6 Use Case | UC ACCESS_02: Search and browse the Parthenos content cloud across several research infrastructures |

3.8. Requirements for interoperability: General requirements

3.8.1. Requirements from Archaeology, Heritage and applied disciplines

3.8.1.1. PIN

| | |
|---|--|
| Partner Short Name: | PIN |
| Collaborator Name: | Paola Ronzino |
| Document / filename (D4Science or Zotero): | D13.1 Service Design_ARIADNE.pdf; D12.1 Use requirement_ARIADNE.pdf; D2.1 First report on users needs_ARIADNE .pdf; D2.2 Second report on user |

| | |
|---|---|
| | needs_ARIADNE.pdf. |
| Community: | Studies of the Past |
| User role: (Actor) | Data consumer |
| Functionality / requirement: (Service) | Data accessibility |
| Explanation (NEED): (Software when present) | The required data(sets) are available in an uncomplicated way |
| Priority level: | High |
| Macro functionality: | Data interoperability and data integration |
| Possible required functions: | Data transparency International dimension |
| Possible involved toolkits or components: | Use of web-based resources Possibility to download the enriched data |
| Comments / remarks: | Content should be provided using the Creative Commons licence suite |
| Function ID: | AR_03, AR_07 |
| Related Use Case: (Knowledge generation process) | Access collections Enriching Visual Media Documents |

3.8.1.2. PIN

| | |
|--|--|
| Partner Short Name: | PIN |
| Collaborator Name: | Paola Ronzino |
| Document / filename (D4Science or Zotero): | D13.1 Service Design_ARIADNE.pdf; D12.1 Use requirement_ARIADNE.pdf; D2.1 First report on users needs_ARIADNE.pdf; D2.2 Second report on user needs_ARIADNE.pdf. |
| Community: | Studies of the Past |
| User role: (Actor) | Data provider |
| Functionality / requirement: (Service) | Data accessibility |
| Explanation (NEED): (Software when present) | The required data(sets) are available in an uncomplicated way |
| Priority level: | High |
| Macro functionality: | Data interoperability and data integration |

| | |
|---|--|
| Possible required functions: | Data transparency International dimension |
| Possible involved toolkits or components: | Faceted search functionality Catalogue of available resources |
| Comments / remarks: | While guidelines for depositing data may differ between archives, a generic set of rules is needed. When redirected, the user should follow the specific guidelines of the archive |
| Function ID: | AR_04 |
| Related Use Case: (Knowledge generation process) | Deposit data |

3.8.1.3. PIN

| | |
|---|--|
| Partner Short Name: | PIN |
| Collaborator Name: | Paola Ronzino |
| Document / filename (D4Science or Zotero): | D13.1 Service Design_ARIADNE.pdf; D12.1 Use requirement_ARIADNE.pdf; D2.1 First report on users needs_ARIADNE.pdf; D2.2 Second report on user needs_ARIADNE.pdf. |
| Community: | Studies of the Past |
| User role: (Actor) | Archive manager |
| Functionality / requirement: (Service) | Data accessibility |
| Explanation (NEED): (Software when present) | The required data(sets) are available in an uncomplicated way |
| Priority level: | High |
| Macro functionality: | Data interoperability and data integration |
| Possible required functions: | Data transparency International dimension |
| Possible involved toolkits or components: | Catalogue of available resources |
| Comments / remarks: | The portal should ensure storing space and long-term availability of data |
| Function ID: | AR_05, AR_06 |
| Related Use Case: (Knowledge generation process) | Search and access the services registry Prepare and register a new collection |

3.8.1.4. PIN

| | |
|---|---|
| Partner Short Name: | PIN |
| Collaborator Name: | Paola Ronzino |
| Document / filename (D4Science or Zotero): | D13.1 Service Design_ARIADNE.pdf; D12.1 Use requirement_ARIADNE.pdf; D2.1 First report on users needs_ARIADNE.pdf; D2.2 Second report on user needs_ARIADNE.pdf. |
| Community: | Studies of the Past |
| User role: (Actor) | Data consumer |
| Functionality / requirement: (Service) | Metadata quality |
| Explanation (NEED): (Software when present) | The available data(sets) are well described |
| Priority level: | High |
| Macro functionality: | Data interoperability and data integration |
| Possible required functions: | Data quality |
| Possible involved toolkits or components: | Metadata input tool Metadata mapping tool SKOSifier tool |
| Comments / remarks: | Metadata records should be published under a CC0 licence to enable integration of multiple datasets within the metadata repository, support resource discovery and enable LOD |
| Function ID: | AR_08, AR_09 |
| Related Use Case: (Knowledge generation process) | Metadata format Vocabularies and gazetteer |

3.8.1.5. PIN

| | |
|---|--|
| Partner Short Name: | PIN |
| Collaborator Name: | Paola Ronzino |
| Document / filename (D4Science or Zotero): | D13.1 Service Design_ARIADNE.pdf; D12.1 Use requirement_ARIADNE.pdf; D2.1 First report on users needs_ARIADNE.pdf; D2.2 Second report on user needs_ARIADNE.pdf. |
| Community: | Studies of the Past |
| User role: (Actor) | Archive manager |

| | |
|--|--|
| Functionality / requirement: (Service) | Data quality |
| Explanation (NEED): (Software when present) | The available data(sets) are complete and well organised |
| Priority level: | High |
| Macro functionality: | Data interoperability and data integration |
| Possible required functions: | Metadata quality |
| Possible involved toolkits or components: | Catalogue of available resources |
| Comments / remarks: | According to D2.1: even when data is available online, it still failed to be useful because the data is structured in different ways, not up to date, is incomplete or lacking important details |
| Function ID: | AR_06 |
| Related Use Case: (Knowledge generation process) | Prepare and register a new collection |

3.8.1.6. PIN

| | |
|---|--|
| Partner Short Name: | PIN |
| Collaborator Name: | Paola Ronzino |
| Document / filename (D4Science or Zotero): | D13.1 Service Design_ARIADNE.pdf; D12.1 Use requirement_ARIADNE.pdf; D2.1 First report on users needs_ARIADNE.pdf; D2.2 Second report on user needs_ARIADNE.pdf. |
| Community: | Studies of the Past |
| User role: (Actor) | Data consumer |
| Functionality / requirement: (Service) | International dimension |
| Explanation (NEED): (Software when present) | Having easy access to international data(sets) |
| Priority level: | High |
| Macro functionality: | Data interoperability and data integration |
| Possible required functions: | Metadata quality Data quality Data accessibility |

| | |
|---|--|
| Possible involved toolkits or components: | Catalogue of available resources |
| Comments / remarks: | European/international dimension of resources may be an advantage with regard to attracting portal users |
| Function ID: | AR_03, AR_08 |
| Related Use Case: (Knowledge generation process) | Access collections Metadata format |

3.8.1.7. ICCU-MIBACT

| | |
|---|--|
| Partner Short Name: | ICCU-MIBACT |
| Functionality / requirement: (Service) | Metadata interoperability via mapping tool |
| Explanation (NEED):(Software when present) | Enable an automatic mapping between different standards to ensure the best dissemination of research data, providing services of data checking, data preview and data enrichment |
| Collaborator Name: | Sara Di Giorgio, Antonio Davide Madonna |
| Document / filename (D4Science or Zotero): | Athena_Digitisation_Standards_landscape.pdf; AthenaPlus_D7_2_Analysis, scenarios use cases, opportunities of innovative services for DCH, and future development_rev_2014_06_15.pdf The MINT ingestion platform; http://www.athenaplus.eu/index.php?en/156/deliverables-and-documents |
| Function ID: | #MINT 01, #MINT 02, #MINT 03, #MINT 04, #MINT 05, \\ |
| Related Use Case: (Knowledge generation process) | MINT (mapping tool): Interoperability with local file (csv, xls, xml); Interoperability with OAI-PMH repositories; Checking data; Preview data; Data enrichment; |
| Community: | Heritage & Applied Disciplines |
| User role: (Actor) | User in the role of Content Provider |
| Priority level: | High |
| Macro functionality: | Data interoperability Metadata quality |
| Possible required functions: | Transformed data preview; Checking Data; Data enrichment |
| Possible involved toolkits or components: | MINT mapping tool |
| Comments / remarks: | The possibility to ingest data via local file is crucial to support research data dissemination. MINT provides metadata in EDM, LIDO and DC format; other metadata formats could be requested by the community. Type: Architectural |

3.8.1.8. ICCU-MIBACT

| | |
|---|--|
| Partner Short Name: | ICCU-MIBACT |
| Functionality / requirement: (Service) | Metadata acquisition and interoperability via OAI-PMH repository |
| Explanation (NEED): (Software when present) | Enable the acquisition of metadata from an OAI-PMH repository in a specified format providing services of metadata validation, reporting, update and managing of invalid metadata. |
| Collaborator Name: | Sara Di Giorgio, Antonio Davide Madonna |
| Document / filename (D4Science or Zotero): | LineeguidaintegrazioneCulturaitalia.pdf |
| Function ID: | #CI 01, #CI 02, #CI 03, #CI 04, #CI 05 |
| Related Use Case: (Knowledge generation process) | Metadata Harvesting; Metadata validation; Repository update; Reporting system; Discard invalid Metadata |
| Community: | Heritage & Applied Disciplines |
| User role: (Actor) | User in the role of Content Provider User in the role of data collection manager |
| Priority level: | High OAI-PMH Data provider (see: http://www.culturaitalia.it/opencms/documentazione_tecnica_it.jsp?language=it HYPERLINK |
| Macro functionality: | Data interoperability; Metadata quality |
| Possible required functions: | Metadata validation; Repository update; Reporting system |
| Possible involved toolkits or components: | OAI-PMH Data provider (see: http://www.culturaitalia.it/opencms/documentazione_tecnica_it.jsp?language=it) OAI-PMH Harvester dashboard |
| Comments / remarks: | Type: Architectural |

3.8.1.9. ICCU-MIBACT

| | |
|--|---|
| Partner Short Name: | ICCU-MIBACT |
| Functionality / requirement: (Service) | Metadata sharing by collection manager via OAI-PMH repository |
| Explanation (NEED): (Software when present) | Enable the harvester to share data acquired by other content providers via OAI-PMH repository with different metadata formats |
| Collaborator Name: | Sara Di Giorgio, Antonio Davide Madonna |
| Document / filename | LineeguidaintegrazioneCulturaitalia.pdf |

| | |
|---|---|
| (D4Science or Zotero): | |
| Function ID: | #CI_06 |
| Related Use Case: (Knowledge generation process) | Sharing ingested metadata |
| Community: | Heritage & Applied Disciplines |
| User role: (Actor) | User in the role of data collection manager |
| Priority level: | Medium |
| Macro functionality: | Data interoperability |
| Possible required functions: | Metadata validation; Repository update |
| Possible involved toolkits or components: | OAI-Cat repository software |
| Comments / remarks: | Type: Architectural |

3.8.2. Requirements from Language-related Studies

3.8.2.1. CLARIN

| | |
|--|---|
| Partner Short Name: | CLARIN |
| Collaborator Name: | Susanne Haaf |
| Document / filename (D4Science or Zotero): | What_researchers_want.pdf |
| Community: | All |
| User role: (Actor) | Researcher |
| Functionality / requirement: (Service) | Improvement of research data storage |
| Explanation (NEED): (Software when present) | Fields of required improvements include protection of (dynamic) data during the research project phase; storage of (static) data after the research project phase; easy access to data stored; creation of sustainable data |
| Priority level: | High |
| Macro functionality: | Data preservation; re-analysis of research data |
| Possible required functions: | Possibility to keep control over data stored; possibility to store data in a protected area; easy to use services; services that suit the researcher's workflow; support; backup solutions; ease of access |
| Possible involved toolkits or components: | There should be a set of available services (rather than a top-down solution) for the researcher to choose from |

| | |
|----------------------------|--|
| Comments / remarks: | Reasons for data preservation are ensuring data re-use, the value of the data collected, authorities (e.g. Academic journals or funding bodies) requiring the preservation of data |
| Function ID: | WRW1 |

3.8.2.2. CLARIN

| | |
|--|--|
| Partner Short Name: | CLARIN |
| Collaborator Name: | Susanne Haaf |
| Document / filename (D4Science or Zotero): | What_researchers_want.pdf |
| Community: | All |
| User role: (Actor) | Researcher |
| Functionality / requirement: (Service) | Enabling data re-use |
| Explanation (NEED): (Software when present) | Ensuring free and easy access to research data so that they can be re-used by others |
| Priority level: | High/medium (depending on the community) |
| Macro functionality: | Data sustainability; data sharing |
| Possible required functions: | Research data storage facilities; open access |
| Possible involved toolkits or components: | Tools to easily share data with certain colleagues, communities or everyone |
| Comments / remarks: | Lowering costs, reduction of research duplication, better cooperation, value/uniqueness of the data, educational purposes count as reasons for this requirement; the report also summarizes obstacles against data sharing |
| Function ID: | WRW2 |

3.8.2.3. CLARIN

| | |
|---|---------------------------|
| Partner Short Name: | CLARIN |
| Collaborator Name: | Susanne Haaf |
| Document / filename (D4Science or Zotero): | What_researchers_want.pdf |
| Community: | All |
| User role: (Actor) | Researcher |

| | |
|--|---|
| Functionality / requirement: (Service) | Quality assurance of research data/datasets by the community |
| Explanation (NEED): (Software when present) | Peer-reviewed descriptions of datasets, enabling comments by users (to be published with the dataset), open access availability and citation possibilities of datasets, code of conduct for researchers on data management and availability of datasets |
| Priority level: | High |
| Function ID: | WRW3 |

3.8.2.4. CLARIN

| | |
|--|--|
| Partner Short Name: | CLARIN |
| Collaborator Name: | Susanne Haaf |
| Document / filename (D4Science or Zotero): | CENDARI-_D5.1-Archive-Directory_final.pdf |
| Community: | Studies of the Past; Heritage and Applied disciplines |
| User role: (Actor) | Researcher |
| Functionality / requirement: (Service) | Digitally available information on: access to historical sources, the existence of historical sources, their contents |
| Explanation (NEED): (Software when present) | The provision of the named information by libraries and archives and the sharing of these data should be augmented, so that researchers can be sure that they don't miss sources relevant for their research |
| Priority level: | High |
| Macro functionality: | Data availability and accessibility |
| Possible required functions: | Standards of describing digital archival data |
| Possible involved toolkits or components: | Catalogue of available resources |
| Comments / remarks: | The paper explains how a respective catalogue (i.e. the CENDARI archive directory) was created with regard to this requirement |
| Function ID: | Cen1 |

3.8.2.5. CLARIN

| | |
|----------------------------|--------------|
| Partner Short Name: | CLARIN |
| Collaborator Name: | Susanne Haaf |

| | |
|--|---|
| Document / filename (D4Science or Zotero): | CENDARI-_D5.1-Archive-Directory_final.pdf |
| Community: | Studies of the Past; Heritage and Applied disciplines |
| User role: (Actor) | Librarian, archivist |
| Functionality / requirement: (Service) | Ontologies, selection criteria |
| Explanation (NEED): (Software when present) | Ontologies which reflect existing classifications and vocabularies used by researchers working on a certain topic (here: World War 1, Medieval Manuscripts); selection criteria for organising and displaying information |
| Priority level: | High |
| Macro functionality: | Data organization |
| Function ID: | Cen2 |

3.8.2.6. CLARIN

| | |
|--|--|
| Partner Short Name: | CLARIN |
| Collaborator Name: | Susanne Haaf |
| Document / filename (D4Science or Zotero): | CENDARI-_D5.1-Archive-Directory_final.pdf |
| Community: | Studies of the Past; Heritage and Applied disciplines |
| User role: (Actor) | Librarian, archivist |
| Functionality / requirement: (Service) | Visibility of an institution |
| Explanation (NEED): (Software when present) | A website about an institution (library or archive) is not enough; it is also necessary to provide information on the collections or individual sources (digitally) available within the respective information; information should be available in different languages (esp. English) |
| Priority level: | Medium (for the aspect of interoperability) |
| Macro functionality: | Data provision and presentation |
| Possible required functions: | Digitized (meta)data (historical sources or catalogues of historical sources) |
| Function ID: | Cen3 |

3.8.2.7. CLARIN

| | |
|--|---|
| Partner Short Name: | CLARIN |
| Collaborator Name: | Susanne Haaf |
| Document / filename (D4Science or Zotero): | D8S-3.1_Transnational Coordination and Collaboration with Third Parties.pdf |
| Community: | Studies of the Past; Heritage and Applied disciplines |
| User role: (Actor) | Researcher; data manager/archivist/librarian |
| Functionality / requirement: (Service) | Mapping archival networks by location |
| Explanation (NEED): (Software when present) | Being able to map archival networks by location of the archives |
| Priority level: | High |
| Comments / remarks: | No requirements suitable for the current issue found in the document |
| Function ID: | Cen3 |

3.8.2.8. CLARIN

| | |
|---|--|
| Partner Short Name: | CLARIN |
| Functionality / requirement: (Service) | Powerful tools and tool maintenance |
| Explanation (NEED):(Software when present) | Researchers wish to have access to better tools and to have their tools maintained |
| Collaborator Name: | Piotr Banski, Susanne Haaf |
| Document/filename (D4Science or Zotero): | 814_Paper_Encompassing a spectrum of LT users.pdf |
| Function ID: | 814-2 |
| Related Use Case: (Knowledge generation process) | Historical newspapers |
| Community: | All (Humanities' researchers with more or less technical expertise) |
| User role: (Actor) | Researcher |
| Priority level: | High |
| Macro functionality: | Sustainability of tools |
| Possible required func- | Tools should provide word/text statistics, NER, geospatial visu- |

| | |
|---------|-----------|
| itions: | alization |
|---------|-----------|

3.8.2.9. CLARIN

| | |
|---|---|
| Partner Short Name: | CLARIN |
| Functionality / requirement: (Service) | Guidance on IPR (intellectual property rights) clearance |
| Explanation (NEED):(Software when present) | Necessity to avoid IPR problems in collecting and sharing data; guidance on what constitutes an "IPR-free" resource, and what the (various levels of) freedom imply |
| Collaborator Name: | Piotr Banski, Susanne Haaf |
| Document / filename (D4Science or Zotero): | 814_Paper_Encompassing a spectrum of LT users.pdf |
| Function ID: | 814-3 |
| Related Use Case: (Knowledge generation process) | |
| Community: | All (Humanities' researchers with more or less technical expertise) |
| User role: (Actor) | Researcher |
| Priority level: | High |
| Macro functionality: | Data accessibility |
| Possible required functions: | |
| Possible involved toolkits or components: | "Legal Helpdesk" (like e.g. provided in CLARIN) |
| Comments / remarks: | This requirement is rather implicit in the text (not explicitly highlighted), cf. p. 2177: "The most interesting resources found at the departments are those that are free of property right problems, [...]. For other resources a priority list for negotiation of access rights will be made. " |

3.8.2.10. CLARIN

| | |
|---|--|
| Partner Short Name: | CLARIN |
| Functionality / requirement: (Service) | Workflow guidelines and workflow sharing |
| Explanation (NEED):(Software when present) | Guidelines for researchers to create workflows, tools to support the creation of workflows and the ability to share workflows with other researchers |
| Collaborator Name: | Piotr Banski, Susanne Haaf |

| | |
|---|--|
| Document / filename (D4Science or Zotero): | 814_Paper_Encompassing a spectrum of LT users.pdf |
| Function ID: | 814-4 |
| Related Use Case: (Knowledge generation process) | |
| Community: | ALL (Humanities' researchers with more or less technical expertise) |
| User role: (Actor) | Researcher |
| Priority level: | High |
| Macro functionality: | Workflow management |
| Possible required functions: | |
| Possible involved toolkits or components: | Workflow planner (provided in CLARIN-DK) which should be useful for all but essential for novice users |
| Comments / remarks: | This is part of what is perceived as enabling turning an archive into an infrastructure |

3.8.2.11. CLARIN

| | |
|---|---|
| Partner Short Name: | CLARIN |
| Functionality / requirement: (Service) | Different search methods |
| Explanation (NEED):(Software when present) | Depending on goals and circumstances, simple text search ("Google-style"), advanced search (e.g. with access to metadata fields and/or to the results of text analysis) and exploratory search (by browsing) are all needed |
| Collaborator Name: | Piotr Banski, Susanne Haaf |
| Document / filename (D4Science or Zotero): | 814_Paper_Encompassing a spectrum of LT users.pdf |
| Function ID: | 814-5 |
| Related Use Case: (Knowledge generation process) | Hist. newspapers |
| Community: | ALL (Humanities' researchers with more or less technical expertise) |
| User role: (Actor) | Researcher |
| Priority level: | High |
| Macro functionality: | Data analysis |

| | |
|--|--|
| Possible required functions: | Text analysis, query management |
| Possible involved toolkits or components: | Corpus analysis systems; Clarin Federated Content Search |
| Comments / remarks: | The main issue concerning interoperability here is the necessity of interoperable data for reliable search results |

3.8.2.12. CLARIN

| | |
|---|---|
| Partner Short Name: | CLARIN |
| Functionality / requirement: (Service) | Functionalities that help with the provision of metadata |
| Explanation (NEED):(Software when present) | The needs of researchers and purposes for data creation vary and require different metadata fields to be filled in upon artefact record creation (or entering them into storage); different pre-formatted metadata profiles are the way to capture the similarities and provide for dissimilarities; tools may help to extract metadata for a text semi-automatically |
| Collaborator Name: | Piotr Banski, Susanne Haaf |
| Document / filename (D4Science or Zotero): | 814_Paper_Encompassing a spectrum of LT users.pdf; CLARIN-Prep-D5R-2.pdf |
| Function ID: | 814-6/CPrep2 |
| Related Use Case: (Knowledge generation process) | |
| Community: | ALL (Humanities' researchers with more or less technical expertise) |
| User role: (Actor) | Novice and expert researcher |
| Priority level: | High |
| Macro functionality: | Description and visibility of resources |
| Possible required functions: | "Wizards" for selecting metadata templates for the user to fill in |
| Possible involved toolkits or components: | Metadata templates, metadata editors, Clarin CMDI, software for metadata extraction from texts |
| Comments / remarks: | For many users, creating structured metadata is a demanding task; assistance in this process is essential, together with a degree of flexibility with respect to the level of expertise and the purpose of the resource (e.g. storing data only for download vs. detailed description of data) |

3.8.2.13. CLARIN

| | |
|---|--|
| Partner Short Name: | CLARIN |
| Functionality / requirement: (Service) | Data analysis/visualisation toolkits, with clearly defined input structures (in terms of standardized formats) |
| Explanation (NEED):(Software when present) | Search and analysis results need to be displayed in a manageable way |
| Collaborator Name: | Piotr Banski, Susanne Haaf |
| Document / filename (D4Science or Zotero): | 814_Paper_Encompassing a spectrum of LT users.pdf; CLARIN-Prep-D5R-2.pdf |
| Function ID: | 814-7/CPrep5 |
| Related Use Case: (Knowledge generation process) | |
| Community: | ALL |
| User role: (Actor) | Researcher, student, instructor |
| Priority level: | High |
| Macro functionality: | Data visualization, data interpretation |
| Possible required functions: | Data analysis |
| Possible involved toolkits or components: | Concordancer, statistical package(s), graph visualization |
| Comments / remarks: | Featured by practically all scenarios, to various degrees |

3.8.2.14. CLARIN

| | |
|---|---|
| Partner Short Name: | CLARIN |
| Functionality / requirement: (Service) | Flexible tools to analyse various data sets linguistically |
| Explanation (NEED):(Software when present) | Pre-processing (spelling normalization, stemming, lemmatization), lexical and syntactic analysis (NER, shallow parsing etc.), information search/content extraction tools (e.g. corpus query functionalities), tools for translation and comparative corpus studies, tools for analysing corpora of speech and visual resources (e.g. for the annotation of gestures, prosody etc.) |
| Collaborator Name: | Susanne Haaf |
| Document / filename (D4Science or Zotero): | CLARIN-Prep-D5R-2.pdf; CLARIN-D3C-6.1.pdf |
| Function ID: | CPrep3/D3C-2 |

| | |
|---|--|
| Related Use Case: (Knowledge generation process) | Hist. newspapers |
| Community: | Language-related studies; All humanities, working with language resources |
| User role: (Actor) | Researcher |
| Priority level: | High |
| Macro functionality: | Text analysis, data processing |
| Possible required functions: | Interoperability among the components of a tool pipeline, data-format standardization, tool input/output format standardization |
| Possible involved toolkits or components: | CLARIN infrastructure; Corpus managing tool, Dictionary editing system: XML-editor; XML-database, (inferred) triple/quadruple store, standard formats for flexible and reliable import/export; corpora (raw, tagged), concordancers (on-line, with a simple query language/interface), tagging and lemmatization tools. Ontologies and lexical database (FrameNet or WordNet) |
| Comments / remarks: | The evaluation of possible CLARIN usage scenarios showed that tools for language analysis would be helpful not only to linguists but also to other HSS researchers. HSS researchers operate on a variety of datasets, sometimes cutting across language stages and coming from various sources. A uniform way to apply statistical methods, for example, to this data is needed. |

3.8.2.15. CLARIN

| | |
|---|--|
| Partner Short Name: | CLARIN |
| Functionality / requirement: (Service) | Standardized/compatible formats |
| Explanation (NEED):(Software when present) | The usage of standardized (input/output) formats; the possibility to convert between formats (for texts, audio or visual data) in general as well as specifically the possibility to convert from non-standard to standard formats |
| Collaborator Name: | Susanne Haaf |
| Document / filename (D4Science or Zotero): | CLARIN-Prep-D5R-2.pdf |
| Function ID: | CPrep4 |
| Related Use Case: (Knowledge generation process) | Hist. newspapers |
| Community: | Language-related studies; All humanities, working with language resources |

| | |
|--|--------------------------------------|
| User role: (Actor) | Researcher, developer |
| Priority level: | High |
| Macro functionality: | Data standardization and interchange |
| Possible required functions: | |
| Possible involved toolkits or components: | Conversion tools |

3.8.2.16. CLARIN

| | |
|---|--|
| Partner Short Name: | CLARIN |
| Functionality / requirement: (Service) | Guidance on LRT methods and practices |
| Explanation (NEED):(Software when present) | A whole new set of concepts and principles need to be communicated and understood; methods for data structuring need to be explained and taught |
| Collaborator Name: | Piotr Banski, Susanne Haaf |
| Document / filename (D4Science or Zotero): | CLARIN-D3C-6.1.pdf |
| Function ID: | D3C-1 |
| Related Use Case: (Knowledge generation process) | Hist. newspapers |
| Community: | ALL |
| User role: (Actor) | Researcher, assistant, developer |
| Priority level: | High |
| Macro functionality: | Cross-disciplinarity, knowledge transfer |
| Possible required functions: | |
| Possible involved toolkits or components: | “Novice mode” in UI, interactive tutorials, schema-aware XML editors with auto-completion; instruction on XML and standard document/data formats; instruction in statistics |
| Comments / remarks: | LRT is applied NLP, a separately evolved discipline that is not intuitive to regular HSS specialists; complex/abstract HSS issues may be difficult to operationalize in LRT terms; advice is needed on selecting the proper tools and data formats for the task at hand, to address project goals and (potentially) maximise reusability |

3.8.2.17. CLARIN

| | |
|---|--|
| Partner Short Name: | CLARIN |
| Functionality / requirement: (Service) | Optimization of data collection from corpora |
| Explanation (NEED):(Software when present) | Time and learning curve are obstacles in learning the existent variety of query languages for average non-technical researchers; a way to cater for this is needed, either by providing a common query language or by ensuring interoperability at the level of query interpretation |
| Collaborator Name: | Piotr Banski, Susanne Haaf |
| Document / filename (D4Science or Zotero): | CLARIN-D3C-6.1.pdf |
| Function ID: | D3C-3 |
| Related Use Case: (Knowledge generation process) | (Historic newspapers) |
| Community: | ALL |
| User role: (Actor) | Researcher |
| Priority level: | Medium |
| Macro functionality: | Standardized corpus query |
| Possible required functions: | Data format validation, data visualisation |
| Possible involved toolkits or components: | (Distributed) corpus analysis system, standardized interpreter for different query strings (in different query languages) |
| Comments / remarks: | Instead of learning a corpus query language, it would be nice to formulate a unique query and this query would be sent to all the existing corpora available at the CLARIN repositories |

3.8.3. Requirements from Studies of the Past

3.8.3.1. SISMELE

| | |
|--|---|
| Partner Short Name: | SISMELE |
| Functionality / requirement: (Service) | Possibility to compare the text of an author with works by different authors |
| Explanation (NEED): (Software when present) | Preparation: search/browse for content and create a virtual collection with found works; create relationships; compare the works like in, document the comparison, add this document to the collection. Usage: search/browse this collection presentation as graph presentation with map/timeline; make annotations; add comments; possibly modify the vir- |

| | |
|---|--|
| | tual collection or add new documents (further preparation) |
| Collaborator Name: | Emiliano degl'Innocenti |
| Document / filename (D4Science or Zotero): | Annex01_USECASES.doc |
| Function ID: | SISMEL_01, SISMEL_03, SISMEL_05 |
| Community: | Studies of the Past |
| User role: (Actor) | Researcher |
| Macro functionality: | Compare |

SISMEL

| | |
|--|---|
| Partner Short Name: | SISMEL |
| Functionality / requirement: (Service) | Get the most information possible about the manuscript: date and place of copy, history of the library where it comes from, bibliography, past editions of the text, images of the code and of the writing. |
| Explanation (NEED): (Software when present) | Summary of content (metadata and annotations); create relationships; presentation of relationships; navigate through graph |
| Collaborator Name: | Emiliano degl'Innocenti |
| Document / filename (D4Science or Zotero): | Annex01_USECASES.doc |
| Function ID: | SISMEL_01 to SISMEL_06 |
| Community: | Studies of the Past |
| User role: (Actor) | Researcher |
| Macro functionality: | Investigation, relationships, navigation |

3.8.3.2. SISMEL

| | |
|--|---|
| Partner Short Name: | SISMEL |
| Functionality / requirement: (Service) | The researcher wants to know when the manuscripts have been produced |
| Explanation (NEED): (Software when present) | Focus on range of time based; search/browsing; presentation with map and timeline |
| Collaborator Name: | Emiliano degl'Innocenti |
| Document / filename (D4Science or Zotero): | Annex01_USECASES.doc |
| Function ID: | SISMEL_01, SISMEL_04, SISMEL_05 |
| Community: | Studies of the Past |

| | |
|-----------------------------|------------------|
| User role: (Actor) | Researcher |
| Macro functionality: | Search, browsing |

3.8.3.3. SISMEL

| | |
|--|--|
| Partner Short Name: | SISMEL |
| Functionality / requirement: (Service) | Get a list of all the manuscripts containing a certain text |
| Explanation (NEED): (Software when present) | Do faceted browsing on manuscripts and institutions; find content, not yet considered in editions: assumed that all editions declare a relation "is_edition_of" to the original; search all content, from this person, which don't declare this relationship |
| Collaborator Name: | Emiliano degl'Innocenti |
| Document / filename (D4Science or Zotero): | Annex01_USECASES.doc |
| Function ID: | SISMEL_01, SISMEL_04 |
| Community: | Studies of the Past |
| User role: (Actor) | Researcher |
| Macro functionality: | Search, browsing |

3.8.3.4. SISMEL

| | |
|--|--|
| Partner Short Name: | SISMEL |
| Functionality / requirement: (Service) | Find out how many collections of unedited texts are held in European libraries |
| Explanation (NEED): (Software when present) | Faceted browsing on ...; find works which are not part of a relationship "is_edition_of" |
| Collaborator Name: | Emiliano degl'Innocenti |
| Document / filename (D4Science or Zotero): | Annex01_USECASES.doc |
| Function ID: | SISMEL_01, SISMEL_04 |
| Community: | Studies of the Past |
| User role: (Actor) | Researcher |
| Macro functionality: | Search, browse, relationship, unedited work |

3.8.3.5. SISMEL

| | |
|-------------------------------------|-----------------------------------|
| Partner Short Name: | SISMEL |
| Functionality / requirement: | Search for philosophical concepts |

| | |
|--|---|
| (Service) | |
| Explanation (NEED): (Software when present) | Faceted browsing on ...; and search for works annotated with keyword "philosophical", about diffusion of works create relationships between philosophical concepts and named authorities; so that users can navigate through the graph of related content |
| Collaborator Name: | Emiliano degl'Innocenti |
| Document / filename (D4Science or Zotero): | Annex01_USECASES.doc |
| Function ID: | SISMEL_02 |
| Community: | Studies of the Past |
| User role: (Actor) | Researcher |
| Macro functionality: | Search, faceted browsing, relationships |

3.8.3.6. SISMEL

| | |
|--|--|
| Partner Short Name: | SISMEL |
| Functionality / requirement: (Service) | Search for quotations in any century texts |
| Explanation (NEED): (Software when present) | Combination of faceted browsing and search on works of the 15th century which are annotated with, for example, Aristotle or have a relationship to Aristotle; create a virtual collection with the results of the search; compare findings; collaborate on this task; create relationships to authorities, e.g. people, places, events, etc., so that users can navigate through the graph of related content to identify which preachers have used Aristotle's texts and florilegia |
| Collaborator Name: | Emiliano degl'Innocenti |
| Document / filename (D4Science or Zotero): | Annex01_USECASES.doc |
| Function ID: | SISMEL_02, SISMEL_04, SISMEL_05 |
| Community: | Studies of the Past |
| User role: (Actor) | Researcher |
| Macro functionality: | Search, faceted browsing, relationships |

3.8.3.7. SISMEL

| | |
|---|---|
| Partner Short Name: | SISMEL |
| Functionality / requirement: (Service) | To find something out about any author |
| Explanation (NEED): (Software) | Search the repository, possibly prepare a virtual collection, |

| | |
|---|--|
| when present) | make notes, create saved search, make annotations, build relationships, creation of bibliography, list of related collections, content, archives/libraries, list of groups or users, list of topics or research areas, list of research questions – provide general information about an author or artist. Something like a summary page, with personal information (biography, ...) |
| Collaborator Name: | Emiliano degl'Innocenti |
| Document / filename (D4Science or Zotero): | Annex01_USECASES.doc |
| Function ID: | SISMEL_01, SISMEL_03, SISMEL_05 |
| Community: | Studies of the Past |
| User role: (Actor) | Researcher |
| Macro functionality: | Search |

3.8.3.8. SISMEL

| | |
|--|---|
| Partner Short Name: | SISMEL |
| Functionality / requirement: (Service) | Know in which manuscripts and repository the texts are held |
| Explanation (NEED): (Software when present) | More focussed on special questions: which manuscripts held by which institution faceted browsing based on: archives, collection, in the boundaries of countries, language/translation, date, range of time, ... present the diffusion of work, possibly with a map and timeline |
| Collaborator Name: | Emiliano degl'Innocenti |
| Document / filename (D4Science or Zotero): | Annex01_USECASES.doc |
| Function ID: | SISMEL_01 to SISMEL_06 |
| Community: | Studies of the Past |
| User role: (Actor) | Researcher |
| Macro functionality: | Search, faceted browsing |

3.8.3.9. SISMEL

| | |
|--|--|
| Partner Short Name: | SISMEL |
| Functionality / requirement: (Service) | Find in which language the works are written |
| Explanation (NEED): (Software when present) | With focus on languages faceted browsing on ... and languages, the search and browsing can be supported by language related metadata and annotations |

| | |
|---|---|
| Collaborator Name: | Emiliano degl'Innocenti |
| Document / filename (D4Science or Zotero): | Annex01_USECASES.doc |
| Function ID: | SISMEL_01, SISMEL_02, SISMEL_03, SISMEL_04, SISMEL_05 |
| Community: | Studies of the Past |
| User role: (Actor) | Researcher |
| Macro functionality: | Search |

3.8.3.10. SISMEL

| | |
|--|--|
| Partner Short Name: | SISMEL |
| Functionality / requirement: (Service) | Harvest biographical information on ... (any Author) |
| Explanation (NEED): (Software when present) | Support a possibility for semantic annotation of content: automatic or manual named entity recognition (person, organization, place, date, event, do faceted browsing on this information present this information with a map and timeline |
| Collaborator Name: | Emiliano degl'Innocenti |
| Document / filename (D4Science or Zotero): | Annex01_USECASES.doc |
| Function ID: | SISMEL_01, SISMEL_03, SISMEL_05 |
| Community: | Studies of the Past |
| User role: (Actor) | Researcher |
| Macro functionality: | Content analysis, faceted browsing |

3.8.3.11. SISMEL

| | |
|--|--|
| Partner Short Name: | SISMEL |
| Functionality / requirement: (Service) | To know what kind of texts are transmitted with an author's work |
| Explanation (NEED): (Software when present) | Follow relationships; see what works are related to an individual's work; do faceted browsing on country, region, range of time, etc. to understand the cultural context |
| Collaborator Name: | Emiliano degl'Innocenti |
| Document / filename (D4Science or Zotero): | Annex01_USECASES.doc |
| Function ID: | SISMEL_01 to SISMEL_06 |
| Community: | Studies of the Past |
| User role: (Actor) | Researcher |

| | |
|-----------------------------|--------------------|
| Macro functionality: | Presentation, view |
|-----------------------------|--------------------|

3.9. Conclusions

The use cases in this document are documenting requirements expressed by a vast number of disciplines in the Digital Humanities community, leveraging on the documentation made available by different partners and networks i.e.: ARIADNE (PIN, MIBACT-ICCU) for Archaeology, Heritage and Applied Disciplines, CENDARI (TCD, SISMELE) and EHRI (KNAW-DANS) for History, CLARIN for Language related studies, Huma-Num (CNRS) for Social Sciences, etc.

Despite the different approach and methodological focus, we found a number of general-level requirements, shared across several use cases and disciplines, expressing the same needs e.g.: data quality, availability, accessibility and enrichment, as well as other specific needs (i.e.: visual media documents enrichment, integration of authority lists, gazetteers and reference tools and/or resources) driven by specific disciplinary concerns. Other requirements both from the backend (i.e.: like storage and preservation) and the frontend (i.e.: tools for collaborative work and data analysis) perspective were gathered. The same can be said for tools, where we've found a similar situation with a shared set of priorities at the general level (i.e. search and information display tools) as well as some detailed, domain driven requirements (i.e. tools to prepare digital editions etc.).

Finally, a set of not (only) technical requirements, such as the sustainability of tools and datasets were expressed by the researchers: we plan to consider them as action points for other WPs (namely WP3), and insert them in the agenda for the development of mid and long-term actions and policies.

According to the PARTHENOS vision, we mapped the domain specific use cases provided by the partners (in the case of KNAW-DANS and KNAW-NIOD, also requirements) against a set of more abstract use cases from WP5 and WP6, addressing general, cross-domain requirements (entities registration and access, VREs creation and use, resources curation, metadata aggregation and export), and verified that the cross-domain functionalities requested by the researchers were fully covered by the PARTHENOS architecture (cfr. Technical Annex C).

4. Definition of education & training requirements

Main author: Jenny Oltersdorf (FHP)

4.0. Introduction

This chapter presents the results of work conducted by Parthenos Task 2.4 (hereafter T2.4). The task aims to compile user needs and requirements regarding training and education in the field of Digital Humanities and affine subjects. Thus members of T2.4, namely partners from AA, CLARIN, FHP and TCD, analysed current provision of infrastructural skills and training on the basis of project reports provided by the PARTHENOS community. It is not the task's aim to collect a tool list and respective requirements but rather to take a more general, abstract view on user needs. Thus a preliminary overview of training needs (in terms of topics) and suggestions for their implementation was compiled. In doing so, the task serves in particular the needs of WP7 "Skills, Professional Development and Advancement". One objective of WP7 is to provide appropriate training and professional development opportunities for researchers at early, mid and advanced career stages. Deliverable 7.1 reports an initial training and education plan developed early in Task 7.1 to organize the project's training activities.

The method for gathering information is presented in section 4.1, divided across sections 4.1.1 – "Text analysis", 4.1.2 – "Survey", 4.1.3 - "Desktop Research", 4.1.4 "User study" and 4.1.5 "European Summer University in Digital Humanities, Leipzig". Findings of the document analysis are expressed in section 4.2 followed by the analysis itself. Section 4.3 presents the results of the conducted survey, whereas section 4.4 provides information about the user study conducted at University of Copenhagen in 2016. Then Section 4.5 covers information on existing platforms for training and education materials as well as a bibliography that contains relevant research literature. Chapter 4.6 provides insight into the European Summer University in Digital Humanities, Leipzig. (ESU). A summary of main results and important aspects can be found in the last section, 4.7, of this chapter.

4.1. Method

The general methodological approach is described in chapter 0. Procedures and information specific to Task 2.4 on the data basis are given in the next two sections.

4.1.1. Text analysis

The team members of T2.4 agreed on a qualitative approach based on the prior decisions of project management to focus the information collection process on already existing outputs of the PARTHENOS community. The projects and initiatives reviewed with regard to their training and education requirements are: AthenaPlus, CENDARI, CLARIN, DARIAH, DASISH, DigCurV, ECLAP and Europeana Cloud. The document analysis procedure was based on a template developed under PARTHENOS WP2 T2.3, and modified using information from the CENDARI project's Deliverable 4.2 - "Domain Use Cases" to reflect the functions of each of the agents in the scenarios. Both the CENDARI work and the T2.3 template applied a simplified Cockburn approach to data gathering around user requirements, placing a strong emphasis on who the user is, what they are trying to achieve and how they will pursue this aim.

The PARTHENOS T2.4 requirements table was therefore developed to capture the following basic information from the documents, namely:

- A unique identifier
- A reference to the document from which the data was taken
- A definition of an actor role (researcher, manager etc.)
- A function (what the actor sought to do)
- An explanation (clarifying the underlying requirement of the actor driving the choice of function)
- A comments field to clarify the interpretation of the PARTHENOS researcher entering the data
- A 'macrofunctionality' statement to clarify how data related to another task within PARTHENOS WP2.

In the course of developing this template, the T2.4 team was required to define an appropriate list of actors for any eventual training programme. This work developed out of a discussion of the PARTHENOS research domain areas (see introduction of deliverable, chapter 0), from which we decided that these fields didn't give us a robust enough basis for determining whose requirements we were gathering.

It was therefore decided that work in T2.4 would focus on gathering training and education requirements for the following target actor roles across the domain communities shared by the full project:

- Researchers, who will use research infrastructures to conduct their work
- Content holders / professional employees in cultural heritage institutions, who will expose their content through research infrastructures

- Technical developers, who will develop specific tools and services for use within research infrastructures
- Administrative managers and decision makers (e.g. faculty deans, university presidents, research funding agencies etc.) who need to understand the role and functions of research infrastructures in order to be effective advocates for them.

The initial inspiration for this set of roles was the DigCurV Curriculum Framework for Digital Curation, which outlines three different ‘lenses’ through which to view curation education, the “Executive,” “Manager” and “Practitioner” lenses. This resonated with, but did not directly map to, the manner in which the CENDARI project and others view their key user communities as comprising the domain researcher, the technical developer and the collections expert. The PARTHENOS T2.4 list presented target actor roles and brings these two (DigCurV and CENDARI approach) together to represent what we feel is a rich view of the ecosystem of actors in which research infrastructures operate, and in which PARTHENOS will be able to make an effective intervention with training, education and awareness raising activities.

4.1.2. Survey

Based on the results of the document analysis, the team of T2.4 decided that further information on training and education requirements is necessary to obtain a more complete overview and get insight into the process of setting up training courses / modules. Thus representatives from ten PARTHENOS-related projects plus two highly relevant projects in the field of training and education were approached to gather further information. The PARTHENOS-affiliated projects ARIADNE, CENDARI, CLARIN, DARIAH, DASISH, EHRI, IPERION, DCH-RP⁶³, PERICLES, NEDIMAH as well as individuals from the projects AthenaPlus and DigCurV were asked for their experiences. The project representatives were asked by email to answer the following questions:

- 1) What kind of training or education services have been put in place within your project?
- 2) What are the main target groups for the training?
- 3) In which way did you identify training needs?
- 4) How did you transform the mentioned needs into training material?
- 5) What didn't work in terms of methods of training, and why (if applicable)?

⁶³ The representative asked for information on the DCH-RP project provided us with data on the Central Institute for the Union Catalogue of Italian Libraries and Bibliographic Information (ICCU) as the DCH-RP project ended in 2014 and no representative of that project was available anymore to pass the request.

- 6) Is there anything else you would like to say to influence how we might provide training through PARTHENOS?

4.1.3. Desktop Research

A third, additional approach to collect relevant information about training and education is the gathering and analysis of information, already available in print or published on the Internet. To complete the information obtained from the project reports and the survey, and to get an outside view of the topic beyond the discussions within PARTHENOS, it was decided at a team meeting together with WP7 at Trinity College Dublin that an overview of research publications as well as an inventory of existing platforms would be helpful. The results of the literature review are presented in the form of a Zotero library. Since the training material developed in PARTHENOS will cover face to face consultations as well as asynchronous teaching material in the form of video tutorials, literature lists etc. an overview of existing platforms seemed necessary. The overview of already existing platforms will help to avoid reproducing services that already exist and enable the identification of potential cooperation scenarios. For further information see section 4.5.

4.1.4. User study

In 2016 the University of Copenhagen conducted a user study in order to understand the digital approaches of researchers within the Humanities. These studies were arranged as a series of open meetings for all interested staff, but primarily for the researchers. Participants came from the universities in Copenhagen, Aarhus, Aalborg and Kolding. Additionally smaller, more focussed meetings with researchers from each department at the Faculty of Humanities in Copenhagen were organized to get insight into the epistemological practices and understand which requirements and needs exist when it comes to the interaction with digital research infrastructures.

A wide range of disciplines were represented from different institutions in Denmark. The group of researcher came from the following departments: Department of English, Germanic and Romance Studies, Department of Scandinavian Studies and Linguistics, Department of Media Cognition and Communication; Department of Design and Communication, Department of Philosophy, Department of Aesthetics and Communication, Department of Culture and Society and the Centre for User-driven Innovation Learning and Design. This study underpins the results obtained by the first 3 mentioned approaches within this task. Since it is a very recent analysis, not only focussed on the Danish universities

but on the services of CLARIN (one of the major PARTHENOS partners), results are presented in section 4.5.

4.1.5. European Summer University in Digital Humanities, Leipzig

Since 2009 the University of Leipzig has offered a two-week summer school in Digital Humanities (http://www.culingtec.uni-leipzig.de/ESU_C_T/node/481). The ESU is directed at an international audience. Students in their final year, graduates, postgraduates and doctoral students as well as postdocs are welcome. The ESU addresses teachers, librarians and technical assistants, engineers and computer scientists. The Summer School is structured in two (mostly independent) blocks of workshops. Some run for one week only, some are planned for two weeks and are structured into two blocks. In close cooperation with the colleagues from Leipzig (namely Stefanie Löpke and Elisabeth Burr) data from past ESUs were analysed with regard to the attendees' background, motivation and expectations on the Summer University. This data is valuable since we get first-hand information on the development, popularity or decline of topics and biographical information about people who are actively interested in training and education programmes. Cooperation with ESU to test and implement the training and education activities of WP7 is planned for the next year. The results of the data analysis can be found in section 4.7.

4.2. Document analysis

Relevant reports have been gathered to collect user needs and requirements. In total 16 documents from 8 projects or research infrastructures were mined for information (Athena Plus, CENDARI, CLARIN, DARIAH, DARIAH-Teach DASISH, DigCurV, ECLAP and European Cloud). Three documents were deemed to not be as relevant as expected at first glance, namely:

Höckendorff, Mareike, Stefan Pernes, and Marcus Held, *Konzept Dissemination und Lehrmittelsammlung Cluster 5 Big Data in den Geisteswissenschaften* (IEG Mainz, DHd/UniHH, 4 2015). – This report deals with the description, presentation and implementation of teaching materials in the field of "Big Data in the Humanities". In addition, a concept for the dissemination of such a collection is presented. The paper was developed within the DARIAH-DE project. The dissemination concept is not of high relevance for T2.4 but may instead be useful for T2.5. The description of the development of teaching material collection is very much tailored to

the use cases developed in DARIAH-DE work package 5 and thus not of general interest for the aims of T2.4.

Dierickx, Barbara, and Maria Teresa Natale, *Athena Plus - Report on User Needs and Requirements in Relation to the Creative Applications for the (re)use of Digital Cultural Heritage Content* (Athena Plus Project, 2013). – It seems difficult for the authors to find the link between user needs described and the various digital exhibitions presented in the document. It was therefore decided to not take it into account for further analysis.

CLARIN: The Knowledge Sharing Infrastructure [KSI], 2014. This report is not relevant in this context, being a brief outline of that the Knowledge Sharing Infrastructure is to be and a short manual of how to apply to become a Knowledge Sharing Centre in the CLARIN organization.

The following section presents the most obvious user needs in terms of topics for training and education as well as suggestions for their implementation followed by the document analysis itself. The results derive from the documents produced in the projects mentioned above. The authors have summarized the documents and formulated the main idea in a short “take away message”. Please note that this outcome should not be considered as a complete overview of the user needs of the research communities in regard to training and education.

The document analysis showed that there is a considerable lack of available training material. The awareness of training needs over all target groups is surprisingly low. Researchers who are experienced with digital techniques “don’t know what they don’t know” thus they do not see the necessity of guided training programmes and do not articulate training needs. They do not actively desire or seek out training because they don’t know what the methods can achieve until shown to them or given a chance to apply them in context. Considering that the target groups do not see clear needs for training, it is perhaps unsurprising that information on training and education requirements could be found in multiple reports but most of them are intrinsic, and not explicitly stated. If training demands are mentioned in the texts, they mostly centre on specific tools developed in the projects.

If one wants to generalize the mentioned tools and needs, training seems to be wanted in the fields of Data Enrichment, Data Quality, Data Archiving, Security & Policies and Management.

Regarding scholarly education programmes one can see that meta-analysis on university DH programmes / curricula have already been done and resulted in a list of relevant teach-

ing points. They range from typical knowledge of humanities subjects to expertise in data modelling and preparation for further digital use including strength in the appropriate presentation of research results. The typical contents of DH programmes can be split into the four groups:

- 1) Basics and general skills,
- 2) Core issues of Digital Humanities,
- 3) Application area of Digital Humanities,
- 4) Technical skills.

With regards to concrete implementation of training, the documents show a clear tendency towards face-to-face interactions, which promise to be most effective. Small group workshops and hands-on training courses are favoured. "Learning by doing" is the preferred method of training for those who are not aware of training needs.

The following section is about the analysis itself. It includes the summary of the project reports as well as a short key take away message.

Thaller, Manfred, *Digitale Geisteswissenschaften* (Cologne: Cologne Center for eHumanities, November 2011) (<http://www.cceh.uni-koeln.de/Dokumente/BroschuereWeb.pdf>)

Key takeaway messages:

- There is a list of skills that need to be trained in DH university courses.
- Skills range from typical knowledge in the humanities subjects to expertise in data modelling and preparation for further digital use including strength in the appropriate presentation of research results.
- A specific target group regarding research disciplines is not mentioned.

This brochure was compiled within the inter-institutional initiative "Digital Humanities Curriculum" supported by the Cologne Center for eHumanities (CCEH) at the University of Cologne under the direction of Prof. Dr. Manfred Thaller and the project DARIAH-DE. The document deals with four main questions:

- 1) What are Digital Humanities?
- 2) How can Digital Humanities be studied?
- 3) What are teaching points in the field?
- 4) Which career opportunities do exist for graduates?

After a general discussion of these issues a list of Digital Humanities degree programmes in Germany including description of the administrative issues and content is presented. Although no explicit target group is mentioned, it is clear from the structure that it addresses an audience interested in studying Digital Humanities. In the second section of the doc-

ument it is said that people will usually study the field of Digital Humanities in conjunction with a "traditional" humanities subject. Thus the most common combination is a Digital Humanities degree programme as the major subject together with a traditional subject as minor field of study. Based on that information, general teaching points are presented. This part is the most relevant regarding training and education requirements. The following areas of teaching are specified:

- Knowledge in common methodological approaches in humanities subjects
- Skills to model relevant data in an open and machine readable manner
- Data preparation for long-term preservation
- Skills to technically execute data which includes abilities in software programming and the use and development of technical architecture solutions
- Knowledge in the analysis of research output
- Qualifications to appropriately present research results.

This list can be understood as an overview of the main relevant topics in the training of Digital Humanities at universities.

Sahle, Patrick, ***DH studieren! Auf dem Weg zu einem Kern- und Referenzcurriculum*** (Köln: HKI / CCEH Köln, 7 2013)

Key takeaway messages:

- The typical contents of DH programmes can be split into the four groups:
 - o Basics and general skills,
 - o Core issues of Digital Humanities,
 - o Application area of Digital Humanities,
 - o Technical skills.
- The paper addresses administrative managers who want to launch or establish DH courses at their university as well as all individuals who are teaching in the field of Digital Humanities

This document by Patrick Sahle was developed within the DARIAH-DE project. The paper addresses administrative managers who want to launch or establish DH courses at their university as well as all individuals who are teaching in the field of Digital Humanities, not only in German-speaking countries, but also in Europe. The document states that one important step towards establishing Digital Humanities as a discipline is the integration into teaching at university level. This is already happening in a wide range of pedagogical formats: from single courses and modules, to focussed offerings, certificates, and summer schools, all the way to BA, MA, and Ph.D. programmes. In the report, a need for reference curricula for the different types of DH programmes is mentioned. Such curricula will help to

improve their coherence and visibility. In order to achieve this goal “an empirical review of existing offerings, an analytical framework for their examination, models for the different basic types, an initial collection of the typical course content and targeted skills, and consideration of how to build a new program of study”⁶⁴ is needed and the reports is a first attempt to achieve that aim. Among others it discusses the geographical distribution of Digital Humanities university programmes and comes to the conclusion that there is a group of countries with clearly recognizable DH training structures, which includes Great Britain, Germany, Canada, the USA, France, Ireland and Italy. Several programmes exist at the BA / MA level in all these countries. Furthermore, isolated programmes can also be found (or were previously available) in Finland, Japan, the Netherlands, Norway, Austria, Portugal, Spain and Sweden. In addition to the overall goal to collect, compare and analyse common curricula, the report aims at a detailed description of the course content. User requirements are not given in an explicit way but can be derived from the analysis of existing courses mentioned in chapter IV. "Inhalte und Curricula". The typical contents of DH programmes can be split into four groups:

- Basics and general skills.
 - General competencies like scientific work; information retrieval; information management; communication; how to write; foreign languages etc.
 - General Humanities methods
 - Subject specific methods
- Core issues of Digital Humanities.
 - What do we mean by digital society, culture and science?
 - Overview of DH as a research field
 - Theory, methods, questions of DH
 - Subject areas and themes in their digital transformation
 - Tools and resources (for specific research areas)
- Application area of Digital Humanities.
 - Core set of applications - e.g. digitization, digital libraries; information systems; digital edition; visualization)
 - Digital objects & data (texts, images, audio objects; geographical / semantic information etc.) including the processes of modelling, coding and (re)use
 - Project-based learning and project practices (modelling, project design, implementation of technical solutions, data analysis, evaluation)
- Technical skills. (This area includes those technologies that are DH relevant in a more practical way of application and usage, but are actually also part of other training programmes in Computer Science.

⁶⁴ http://cceh.uni-koeln.de/files/DARIAH-M2-3-3_DH-programs_1_2_0.pdf, p.1

- o Web Technologies (networks / client - server / HTML / CSS / Javascript)
- o Publication technologies (design, web publishing, CMS repository systems, etc.)
- o Databases and data structure
- o Coding and software engineering

Jakub Beneš, Kathleen Smith, Andrea Buchner, Klaus Richter, and Pavlina Bobič, *CENDARI Report on Archival Research Practices*, 25 June 2013 and CENDARI Project, '*CENDARI Project: Domain Use Cases*'.

Key takeaway messages:

- Peer-to-peer advocacy is important in raising awareness of the usefulness of digital techniques (either in digital curation or the wider digital humanities context).
- Researchers may be unaware of the benefit they might gain from formal training in advanced methods.
- Researchers do not actively desire or seek out training because they don't know what the methods can achieve until shown them or given a chance to apply them in context.
- Learning-by-doing due to necessity is the key driver in developing new techniques.

These documents summarise the outcomes of the user-centred design process and underlying methodological foundations of the key user groups for the CENDARI project. Although they contain a wealth of insight as such, they are not the most useful documents for the consideration of training and education needs, however. There are a number of reasons for this. The former document is very much focussed on archival practice, and the skills and training needed to undertake research in a traditional, analogue archive. Here, the researchers spoken to felt very little need for training. Even when the document does discuss digital methodologies, the subjects informing the work seemed to have very little awareness of or desire for any formal training: "Typical was the response of a medievalist who said that introduction to these methods proceeded "in an ad hoc way, because I haven't had any training in digital humanities.... I began the job and just started reading and looking at other projects." (7). Three interviewees cited the importance of colleagues in introducing them to such methods. One remarked that the university's choices in providing software to staff determined the sort of methods used (12)." (63) The assumption implicit in this statement, essentially that researchers do not actively desire or seek out training because they don't know what the methods can achieve until shown them or given a chance to apply them in context (learning how to do so along the way) is borne out by the for targeted work underpinning the Domain Use Cases. Throughout the account of the three participatory design sessions and 13 user scenarios/user stories, training needs are

not mentioned once. This is in part an artefact of the method underlying the data gathering exercise, which seeks to draw our current practice from the users, rather than what they think they might do in a digital environment, but it also reinforces the interpretation put forward above, that is that researchers may be unaware of the benefit they might gain from formal training in advanced methods.

Europeana Cloud Project, '*Europeana Cloud Report on User Needs*' (Angelis et al. 2015)

Key takeaway messages:

- Researchers who are inexperienced with digital techniques “don't know what they don't know”
- Unguided training takes longer and is less accurate
- Not all technicians recommend training in APIs, as ‘a little knowledge can be a dangerous thing’.

This extensive document covers user requirements, which range from content, tools and services to training needs. Topics that cover training needs within this report could be found in multiple places, most of which are intrinsic and not explicitly stated within the document. Exceptions to this occur in Section 5 (APIs in Humanities and Social Science Research), which discusses the needs of researchers to ‘skill up’ in digital data mining methodologies. Questions such as:

- “What support is available for researchers who don't have a high level of technical expertise?”
- Should training be made available to researchers to give them the means to access and re-use this data themselves in a manner best fitted to their research methodology?
 - o If so, training in what?
 - o To what level?

This report suggests that a lack of technical expertise can be a barrier to advanced methods of data reuse, and that training in API-use should be provided in addition to providing the API itself. When reviewing free online training resources, the rationale for Software Carpentry showed that while people may take courses at undergraduate level, there are no formal training modules in data techniques or programming at postgraduate level or beyond as “they are expected to pick up programming on their own”.⁶⁵ With such unguided

⁶⁵ <http://software-carpentry.org/blog/2013/06/lessons-learned.html> (accessed by eCloud: 24th September 2014)

training, there is a tendency for researchers to spend far longer trying to work out what they actually need to learn, or simply that they don't know where to start. Furthermore, it can be difficult for someone who is unfamiliar with a particular field to even know what they don't know.

Summer schools put on by projects such as CLARIN-DE can be useful, and the British Library has put together training for its staff so that they can better help customers visiting the library.

Recommendations were made for Europeana as a result of this document. However these recommendations can be used in a wider context. Of those that relate directly to Education and Training in the section on APIs and data reuse, the following were pertinent:

- Offer tutorials with clear technical prerequisites, and pointers toward other sources of technical training
- Host periodic training 'workshops' (either online or in person) to allow those keen to learn new digital techniques alongside [Europeana] content (p.124).

Engelhardt, Claudia, Katie McCadden, and Stefan Strathmann, ***DigCurV - Report and Analysis of the Survey of Training Needs*** (Goettingen)

Key takeaway messages:

- "Storing and Managing Data" and "Project Management" training is considered important
- "Learning by doing" is the preferred method of training
- Small group workshops considered more effective
- Target group: practitioners working in the CHI sector

This is a report on training needs survey conducted within the CHI sector into Digital Curation training. While not directly relevant, it does show the methods used, and the responses that could be taken at a broader level into general Digital Humanities (DH) training. For example, among the methods of training suggested, small group workshops were considered the most effective. However, in addition to the suggested methods, 'learning by doing' was also mentioned by a few participants.

Among the more technical topics for discussion, 'storing and managing data' was considered important. This could have cross-disciplinary relevance, particularly for digital humanists. Likewise, 'project management' is also important.

This is a very comprehensive account of findings regarding user-needs training and education in the field of Digital Curation and Preservation. This therefore makes it very specific to these needs. Many of the generic skills listed would be relevant to any researcher

(grants and funding applications, communications and networking skills, project management), but as these are so generic it doesn't tell us exactly what needs a DH researcher would have of an infrastructure. That said, the methodologies used are useful here, along with the findings regarding the duration of the training involved.

Victoria Arranz, Daan Broeder, Bertrand Gaiffe, Maria Gavrilidou, Monica Monachini, and Thorsten Trippel, "**Describing LRs with Metadata: Towards Flexibility and Interoperability in the Documentation of LR**", 2012

Key Takeaway messages:

- Annotation of standardized metadata to language resources is important
- Hands-on annotation training courses are needed
- Targeted user groups are researchers, content holders and technical developers

This compilation of documents strives at addressing issues and challenges in the concrete work with metadata for language resources in a broad perspective. Several of the contributions contain elements that would be relevant to use and incorporate in a training curriculum.

Annotation of standardized metadata to a given language resource leads to more advantages. Not only will it ease and enable a precise search for the resources that the user wants, it will also enable exchange of data as an efficient way to avoid a waste of effort. The need for training of more user groups with respect to assigning standardized metadata information to language resources is a well-established fact. Not only researchers but also content holders and technical developers would benefit from learning about standardized metadata annotation.

Lina Henriksen, Dorte H. Hansen, Bente Maegaard, Bolette S. Pedersen, and Claus Povlsen, "**Encompassing a Spectrum of LT Users**", *Proceedings of the 9th Conference on Language Resources and Evaluation : LREC 2014*

Key takeaway messages:

- To reach HSS researchers implementation of user-friendly interfaces is crucial
- Design of research platforms should be made in close contact with the users
- The target user group is technical developers

This paper is directed towards technology developers and it concerns the design of a user-friendly interface and functionalities for the CLARIN platform with the aim of reaching the HSS (Humanities and Social Sciences) researchers.

The paper points out that although language-processing tools are available from the CLARIN platform and elsewhere, and even if researchers know about the platform and the types of tools accessible from the platform, the Humanities and Social Sciences (HSS) researchers still do not use the platform or the tools. The paper emphasizes that other measures must be established in order to reach the HSS researchers and bridge the gap between available language resources and tools on the one hand and traditional HSS research methods on the other hand. More specifically, the document suggests a number of concrete functionalities which should be included in the design of platforms such as CLARIN in order to reach the HSS research community.

Kemp-Snijders, Marc, and Lothar Lemnitzer, ***CLARIN - Usage and Workflow Scenarios***, 30 June 2009

Key takeaway messages:

- Interaction and dialogues between Humanities and Social Sciences (HSS) researchers and Natural Language Processing (NLP) researchers are important
- Implementation of visualization tools is useful both for the HSS researchers (methodology) and for making NLP understandable for the HSS researchers
- Target group: Researchers in language related research areas

This report concerns the identification of user needs in preparation for the creation of user scenario examples for the CLARIN platform. The reported user needs findings indicate that HSS researchers within domains such as history, linguistics and other language related areas and researchers within Natural Language Processing (NLP) related areas have severe difficulties in understanding each other's research. In connection to this, the report advocates the need for visualization tools and methods. Visualization tools can contribute to gaining new insights into data and to the methods behind automatic data analysis. In other words, visualization tools contribute to understanding methods of natural language processing (giving researchers the ability to carry out research within their own domain faster or with new insights, new questions etc.).

Váradi, Tamás, and Piroska Lendvai, ***CLARIN - Integrated Strategic Plan for Supporting HSS Research***, 16 May 2011

Key takeaway messages:

- Close collaboration between NLP experts and HSS researchers promotes use of LPT in the HSS community

- Researchers in the field of Humanities and Social Sciences need to know more about statistics, annotation, and XML

The broad target group treated in this report is scholars within HSS that, in their research and methodology, focus on text analysis. This group of researchers count as members of not only the Language-related community but also the communities from Studies of the Past and Social Sciences.

In general, the report provides an overview of the results achieved by a one-year collaboration between CLARIN experts as advisors and Humanities and Social Sciences researchers as users. Important information revealed during this co-operation shows that users need knowledge and competencies within the following topics and issues: statistics and methodologies of language processing, including knowledge of XML and annotation.

Existing Language Processing Tools (LPT) are, even if relevant for some research tasks, rarely used by HSS researchers. Therefore, a qualitative study of HSS researchers' use of LPT was carried out. The study shows that researchers need a better understanding of the ideas and techniques underlying LPT, which in effect entails that they cannot (or at least can often not) meaningfully use the tools/methods in their research.

The report does not give specific recommendations to the contents of the courses in statistics, XML and annotation. But the observations and recommendations of the report are very well in line with our own experiences here at the University of Copenhagen. In our experience, a number of researchers are aware of the existence of methods and tools and they know about the CLARIN platform, but their knowledge about preprocessing of resources and about methodologies behind tools is insufficient.

Gnadt, Timo, and Claudia Engelhardt, ***DASISH - Data Service Infrastructure for the Social Sciences and Humanities*** (Goettingen)

Key takeaway messages:

- There is not much training material available yet among the ESFRI projects or their participating institutions, which DASISH could directly use or integrate into an existing platform. The responses revealed a considerable lack of available training material.
- The main topic areas of interest for training and education are Data enrichment, Data quality, Data archiving, Security & Policies, and Management.
- The results of the study refer to researchers in the ESFRI community, which mainly means researchers in the field of Social Sciences and Humanities.

The report deals with the outcome of DASISH task 7.1 "Training Modules". It was dedicated to develop online training modules for topics and target groups relevant to the Social

Sciences and Humanities communities. In section 2 of the paper, the processes of assessing the training needs of and available material from ESFRI communities are described.

The assessment of training requirements and needs was carried out for the target groups of “developers and managers of data archives and repositories, decision makers from research and education institutions as well as researchers”.⁶⁶ The respective ESFRI projects / institutions are:

- CESSDA - GESIS, NSD
- CLARIN - UiB, OEAW
- DARIAH - KCL, UGOE
- ESS – NSD
- SHARE - MPG-MEA

In order to assess the concrete training needs of the target groups, a two-round survey among the SSH ESFRI projects was conducted. The analysis of the first questionnaire resulted in the derivation of five main topic areas, each with between three to six sub-topics:

- Data enrichment,
- Data quality,
- Data archiving,
- Security & Policies, and
- Management

In a second step, people were asked to indicate the relevance for certain target groups, the kinds of desired activity, available training material and to give further comments. People felt a need for training in the fields of “Access Policies”, “Licensing”, “Persistent Identifiers”, “Data analysis/harmonization”, “Workflows”, “Linked Data”, “Authentication and Authorization Infrastructure”, “Metadata standards and usage”, “Publication/Open Access” and “Deposit services and SLA negotiations”. In addition, one result of the questionnaire was the awareness of a considerable lack of available training material. Thus, DASISH WP7 designed online training modules on “Access Policies and Licensing”, “Authentication and Authorization Infrastructures” and “Persistent Identifiers”.

McCrae, John, Jorge Gracia, Roberto Navigli, and Paul Buitelaar, “**Reconciling Heterogeneous Descriptions of Language Resources**”, 2015

Key Takeaway messages:

⁶⁶ DASISH Report 7.1_training_needs, p.2.

- Harmonization of metadata is necessary for efficient discovery and finding of language resources.
- Target group: Technical developers

This document forms a paper of the Proceedings of a workshop on linked data in linguistics

that took place in the 53rd Annual Meeting of the Association of Computational Linguistics and the 7th International Joint Conference on Natural Language Processing in China (July 2015). As a result, it is directed at technology developers working on the harmonization of language resources, and suggests methods and techniques for the improvement of data quality of metadata records. More specifically, the document wants to break the dichotomy between curatorial and crowd-sourced resources by suggesting a set of properties (resource type, language, intended use, licensing conditions) for description and discovery of relevant language resources. It also seeks to detect and remove the duplicates of metadata within and across different repositories of language resources (using META-SHARE, CLARIN, LRE-Map, Datahub.io, OLAC, ELRA and LCD Catalogues) and harmonize them, by applying NLP techniques. Its final aim is to render metadata queryable and browseable on the Web in an efficient way. The subject of the document is quite technical and not orientated to education and training requirements. Its target group is the technology developers and NLP engineers and only indirectly does it concern language researchers.

Belice Baltussen, Lotte, Maia Borelli, Irene Scaturro, Ferruccio Marotti, Emanuele Bellini, Katia Maratea, and others, "***ECLAP-DE2-1-2-User-Requirements-and-Use-Cases-v1-1-Final***", 14 February 2012

Key Takeaway messages:

- Few education and training needs are extracted from use cases, and user requirements has no separate section
- Training for students and researchers to use content for preparing essays and papers
- Training for cultural content managers to create and curate virtual exhibition
- Training for joining, creating and managing groups of users
- Training for accessing and managing digital content via mobile devices

This document is the second version of the ECLAP deliverable on use cases and user requirements, developing in more detail the needs of the project's target users. ECLAP's main aim is to create an online archive of performing arts in Europe and reach users who would like to browse, search, view and interact with the archive's collections, as well as to

join groups with the same interests and provide archive with new content. The users are separated into eleven groups and are clustered in three macro-categories:

- Education/Research
 - Student/Researcher in higher education
 - Teacher in higher education
 - Performing arts student
 - Performing arts teacher
 - Primary school teacher
 - Secondary school teacher
- Leisure/Entertainment and Tourism
 - Leisure user
 - Tourism operator
- Cultural heritage professionals
 - Performing arts practitioner
 - Content manager/provider
 - Media professional

User groups are also categorized with regard to their education and technological skills (low, medium, high). Admittedly, the users of the Education and Research domain are the main target group of ECLAP, since they are the most likely to be ‘heavy users’ of the portal and its content (Belice Baltussen et al. 2010, p.6). The document discusses also quite extensively use cases and user requirements for the Cultural Content Managers and Leisure Users. However, it does not provide a separate section on education and training needs of users. There are, though, some training requirements that are ‘hidden’ in the use cases and user requirements. More specifically:

- Students and Researchers in higher education probably need training in order to use content for a specific subject of study (i.e. academic publications).
- Cultural content managers would need training to create and contribute to virtual exhibitions.
- All users may need training for joining, creating and managing groups of similar users for exchanging knowledge and sharing interest and experiences.
- Training is also discussed for users who access ECLAP via mobile devices and would like to organize the downloaded content in their device. For them a specific application will be developed.

Finally, a discussion takes place about exporting data via an API. However, developing an API is considered a complex technical process and beyond the aims of this deliverable, in

spite of the fact that certain use cases can be a basis for inspiration for an API development process.

Agiatis Benardou, Panos Constantopoulos, Costis Dallas, Eliza Papaki, Christos Papa-theodorou: “***DARIAH Teach - WP2 User Requirements & Benchmarking of Key Competencies - Deliverable 7: Report on user Requirements: Reference Curricula to be Developed and Benchmarking Key Criteria***”

Key takeaway messages:

- An ideal platform for training material should provide a wide range of features focusing rather more on its easy, efficient and plural function than aesthetics, and encouraging interaction and communication among members of the community and social networking.
- Modules should come with clear goals, flexible structure, open to the community of DH, dynamic search and evaluation based on different levels of complexity, as well as dynamic curation of learning object metadata.
- Focus on the need of the community to share training material through a platform, whose sustainability would be guaranteed through DARIAH.
- Target group: Researchers that are trainers in the field of DH

This report has been conducted by the DARIAH Teach Erasmus Plus network and focuses explicitly on the user requirements of “teachers of Digital Humanities”, who are included in the general category of “researchers”, in regard with a tool/platform they would like to have in their modules. The report is based on 15 interviews with instructors of Digital Humanities holding academic-related positions, being at different career stages (ranging from doctoral candidates to full professors). The interviewees emphasized the potential of the DARIAH Teach platform as being user-friendly, a free-structure environment, dynamic instead of a repository-like platform, as well as its ability of building a community, making clear statements of what users can and cannot do with it, being user-tested during its development and including alternative ways to FAQs. According to the interviewees, the platform should also enable both synchronous and asynchronous collaboration and communication, encourage user interaction and offer various content features.

With respect to the Modules, interviewees expect to have a variety of content and features: theoretical courses on Digital Humanities and block courses on basic Computer Science skills including units that combine theory and practice and assignments that teach students how to do research in relation to (re)searching content (online archives and collections). According to the interviewees, modules and teaching material should be exportable to other learning platforms, open and freely available for students after the completion of their

studies, with a clear and shareable copyright, which would enable them to be adapted by other universities. Furthermore, modules should have a dynamic base for search, searching not only keywords but also formats and enabling serendipitous search.

Concerning the structure of the module, opinions differ. Some of the interviewees argue that the units should not be absolutely standalone, with at least the initial ones to be required, while others support that module units should be created as separate entities. The suggested module length ranges from 10 to 14 teaching weeks, following the format of semesters and thus being clear and familiar.

The learning outcomes are expected to be explicitly stated and complexity should be more intense as the module progresses with evaluation based on the different level of assignments' difficulty. Finally, modules should cater for dynamic curation of learning object metadata.

4.3. Experiences of the PARTHENOS community

In a brief survey, the PARTHENOS community was asked for their experiences with training and education requirements. The presentation of results follows the order of proposed questions. The full texts provided by the project representatives can be used for further PARTHENOS-internal analysis if needed. However, as the time for answering the questionnaire was very short, the authors did not intend to publish results in the deliverable in full text (and did not ask for permission) as they assumed, that this concern would complicate and extend the return rate extensively.

What kind of training or education services have been put in place within your project?

The offered training and education services range from peer-to-peer workshops, training camps and doctoral summer schools, PhD courses, written guidelines on specific topics and online tutorials and webinars. Especially in the CENDARI project, a sophisticated training programme based on modular training materials for online viewing and download is mentioned. It includes the publication of training materials on Basecamp at regular intervals. In the CENDARI project, Basecamp also serves as a forum for users to post queries if desired. The training material is presented in the format of written documentation (e.g. pdf form). The documents explain how to use a particular tool and are often illustrated by screenshots. This written information is accompanied by short training videos which were uploaded onto the project's YouTube channel. Some weeks after the release of these materials, webcasts are offered on the topic. The webcast consist of a short presentation, fol-

lowed by a Q&A session with members of the user group. A detailed online survey that queried both the usefulness and the ease of use of the tool in question followed. These results are then collated and shared with both the development team and the user group. Most of these training materials are later adapted and published on the project's website.

What are the main target groups for the training?

The experiences of the PARTHENOS community are based on training for the following target groups:

EHRI:

- Historians, archivists, other researchers from the humanities
- Courses aim at the graduate level

PERICLES

- Academic and scientific communities active in fields related to the project, such as Digital preservation, Computer science, Information science
- Individuals working with data (e.g. researchers, data creators, data users, data curators, archive managers, conservators, collection holders)

ICCU

- Experts, managers of museums, libraries and archives that deal with digital collections
- Undergraduate and graduate students doing research in the fields for which the institute is responsible

CENDARI

- Potential users of the infrastructure: historians, graduate students, digital humanists, archivists, and librarians

AthenaPlus

- Project partners
- Gradually expanded to outside stakeholders, and in the end they mainly offered training for outside stakeholders

DASISH

- Developers and managers of data archives and repositories, decision makers from research and education institutions as well as researchers in the field of SSH

IPERION

- Potential users of the research infrastructure

ARIADNE

- Individuals with a scientific interest and ability to benefit from training in archaeological research data management
- Priority is given to users who have not previously used the ARIADNE infrastructure, young researchers, and researchers working in countries where no such research facilities exist.

CLARIN

- Researchers applying the research infrastructure
- Researchers applying for research grants needing data management expertise
- Potential data providers and data centres
- Students

DigCurV

- Staff of cultural heritage institutions
- Regarding the “nestor Schools” the target groups are slightly different: cultural heritage staff, along with students of archive/museum/library studies, teachers and researchers from this area as well as attendees from private entities

In which way did you identify training needs?

The mentioned approaches for identifying training needs are mainly questionnaires and surveys to be filled in by the target communities. Some projects collect feedback from the work packages or from participants, organizers and lecturers of Summer Schools and workshops. Analysing support requests is also a mentioned method of collecting training needs.

How did you transform the mentioned needs into training material?

The transformation of user needs into training material is a challenge for all requested projects. Regarding the implementation of training offers the process typically starts from the own experiences of how to communicate skills.

A typical workflow mentioned by one project representative includes extensive test-driving of the infrastructure and new developments, often liaising with the creators or developers. Access to technical documents, prior presentations, and other sources of information is given beforehand. Based on that test-drive process information is tailored into user-friendly written guides and videos etc. and disseminated to the respective communities. After a period of time, feedback would be collected and presented to both the developers and members of the user group.

Another project representative mentioned the adjustment of existing workshops and training materials to their teaching experiences. The creation of user guides as a reference manual was mentioned as well.

What didn't work in terms of methods of training, and why (if applicable)?

Experiences of the PARTHENOS community showed that Summer Schools and other face-to-face meetings tend to take too much time. A three-week long Summer School is too intense and it is difficult for participants to take off so much time.

The participants indicated also that it is challenging to drum up regular enthusiasm in the user group, even though this group consisted of interested future users.

Distance learning tools like tutorials turned out to not be the best means to get messages across and to maintain user interest and attention. A better approach is to set up moderated tools like webinars or Skype training. The same applies to written documentation and materials. The learning curve climbed much faster if those materials are accompanied by personal training in a workshop or webinar.

Additionally, it was pointed out that attracting new user groups is much easier during a conference than at pre- or post-conference sessions. Thus training offers should be part of the regular conference schedule.

Is there anything else you would like to say to influence how we might provide training through PARTHENOS?

If recruiting a user group, the recommendation is to attempt to create a core group of users that are linked to project members (for example, postgraduate students at a participating institution) and offer incentives for them to participate regularly.

If the aim is just to create training materials it should be considered how and where they will be distributed. For example, if material is uploaded to the project's website, it needs to be made sure that due publicity is given to this. Digital Humanities projects should plan the creation of a user group from the very beginning of the project. This includes the very important development phase, in order to insure that the needs of the final users are met. In the CENDARI project, these phases did not include the same group of people, as the participatory workshops, outreach workshops and Trusted Users group worked with different users. The presence of a physical person who is acquainted with the material is definitely of great value. Online materials are a 'must' too but are best accompanied with a real-life Q&A and/or hands-on training time and should be accomplished with well-designed, yet easy to administer content management systems. Training events need to be planned and announced well in advance. The concept of hands-on exercises to consolidate theoretical

knowledge acquired in a lecture before Summer Schools or workshops resulted in positive participant feedback.

4.4. User study about digital approaches at University of Copenhagen

In 2016 the University of Copenhagen conducted a user study related to the CLARIN-DK network. The objective of this user study was to get an overview of the types of research conducted at the Danish universities and the research methods and tools applied in order to determine research and training needs which could potentially be addressed through the CLARIN project. Another objective was to get the users' feedback on the CLARIN-DK platform.

The meetings revealed that most researchers collect different kinds of materials during their careers and at a certain point they tend to wish that "somebody" will take the responsibility to store the materials for the future. So a general observation was that researchers are very interested in a platform for storage of data. Of course not all departments and researchers have language-based materials – some have photos from archaeological excavations, some have questionnaire survey data consisting mostly of yes/no answers or numbers. Requirements for storage and processing of these types of data have not been included in this user study as language data are our main focus. One of our conclusions was however, that researchers in the humanities generally have at their disposal a wealth of resources: historical texts, literary texts, old language texts, dialect materials etc.

Regarding researchers' use of digital methods in their research today, a general observation from the completed series of meetings is that the researchers have very different approaches to and experience with digital technologies. For some this is new territory and the usefulness of web-services such as those available through CLARIN-DK is still debatable. Others have an extensive insight into methods such as corpus work and the use of digital platforms.

The tools used by researchers range from quite simple tools, to commercial off-the-shelf tools (typically for statistics) and complex tools, either self-produced or developed by a third party for the specific purpose. Some researchers expressed a distinct need for highly specialized tools, such as a general language lemmatizer for 16th century Danish, a dictionary of different spelling variants of 17th century Danish place names or a viewer tool displaying facsimile, transliteration and translation aligned by user specified segment.

Tools with such a narrow scope are usually non-existent or they have to be modified for the particular operating system, format requirements, sublanguage, etc. One topic where many researchers seem to have common interests is within the subject of annotation; some work with it manually, some automatically; some work with text and others with speech. A few recent seminars reflect that this is a field in rapid progress.

Researchers' needs for highly specialized and sometimes also non-existent tools are not ground-breaking news, however the urgency of the situation became very clear. This certainly emphasizes the need for a platform that can inspire the users to new work methods and facilitate the sharing of existing self-produced tools.

The meetings with researchers demonstrated quite clearly that cross-institutional communication and cooperation in order to understand more about other researchers' research methods and to meet some of the researchers' needs for digital tools and methods are needed.

Research at the Department of History is an example of a research area with a high degree of readiness for digital methods. They take an interest in sophisticated data processing methods involving word/text statistics, named entity recognition and geospatial visualization methods and tools. These approaches allow researchers to extract or analyse information about places, people and events and find new relationships between them. History researchers already use different combinations of existing tools from Google and elsewhere, but the tools are often originally created for other purposes and there are copyright issues in connection with use of data in some tools from commercial platforms, such as Google. CLARIN-DK web services are also interesting, but many researchers commented that they found it difficult to get an overview of the different types of data and services that are offered in the CLARIN-DK platform. They expressed a need for more information about the results that can be obtained from the online tools, and some had only a vague idea about the general application or even the existence of CLARIN-DK. This is all in support of the idea that researchers need courses in the use of infrastructures and linguistic tools. Furthermore, the existing and rather generic tools included in CLARIN-DK are far from sufficient to cover the researchers' needs, despite their usefulness.

4.5. Inventory of existing platforms for training and education

The overview of existing training resources available openly to digital humanities researchers complements the presented analysis of project-generated reports, the survey results and the user study from University of Copenhagen.

ADHO - The Alliance of Digital Humanities Organizations <http://adho.org/> is not really an inventory of training materials. However, the website offers a collection of resources which includes a detailed list of relevant summer schools in the field of Digital Humanities. ADHO is an umbrella organisation whose goals are to promote and support digital research and teaching across arts and humanities disciplines and it will support excellence in research, publication, collaboration and training.

DARIAH Teach <http://dariah.eu/teach/> began in January 2015. Currently there is no content available via the website regarding its training outputs. However, in the near future it will become a valuable resource for open-source, high quality, multilingual teaching materials for Digital Humanities. Led by Maynooth University, DARIAH Teach aims to strengthen and foster innovative teaching and learning practices among the members of DARIAH.

The Digital Humanities Course Registry <https://dariah.eu/library/dh-course-registry.html> is an inventory of Digital Humanities courses and programmes. The service offers a search environment that combines a map of Europe with a database that contains information on Digital Humanities courses. Students as well as lecturers can search the database on the basis of topographical location, credits or degrees that are awarded, and keywords. The Digital Humanities Course Registry offers a basic documentation on scholarly education programmes throughout Europe, ranging from typical knowledge of Humanities subjects to expertise in data modelling and preparation for further digital use with emphasis on the appropriate presentation of research results. In the framework of Task 7.4, WP7 will analyse existing higher education curricula and deliver a report.

DiRT - Digital Research Tools <http://dirtdirectory.org/> is a registry that enables searching for digital research tools for scholarly use. DiRT is maintained by an international volunteer community of professors, students, and librarians and it is overseen by a steering/curatorial board, and supported by an editorial board. It can be used for discovering and comparing DH tools. It enables access to a variety of tools that range from software for analysis and visualization work to tools for annotating resources and managing bibliographies. The tool descriptions include information such as the platform (e.g. Windows, iOS, etc.), financial aspects and licensing. In addition one can find reviews, tips, and tricks for efficient use. The Directory uses a Creative Commons Attribution license.

Open Educational Resource Platform <https://www.oercommons.org/groups/dariah/229/> offers training material in various research areas. In the 'DARIAH-Group' a huge number of either self-generated learning and teaching materials or external open licensed material in

the field of Digital Humanities for higher education has been collected. The group is an effort from DARIAH-DE. Training material is divided into the categories:

- Computational Linguistics,
- Digital Humanities (general),
- Digital Libraries and Databases,
- Semantic Technologies,
- Software Engineering,
- Technical Applications, and
- Technical basics.

Zenodo zenodo.org/collection/user-dcc-rdm-training-materials is a platform that enables sharing, preserving and publishing multidisciplinary research results in form of data and publications that are not part of the existing institutional or subject-based repositories of the research communities. All research outputs from all fields of science are welcome. Types of files range from books and book sections to images, software and interactive materials such as lessons. Zenodo was launched within the EU funded OpenAIREPlus project.

The existing bibliography in WP2 covers first and foremost PARTHENOS related project reports and deliverables and excludes research publications about training and education. That is why not all reports are freely available. This additional bibliography was created to fulfill another function. The bibliography covers mainly research publications about training and education and is freely available to everyone on the internet. To enable low threshold access and easy export and usage, it was created in Zotero. It can be access via

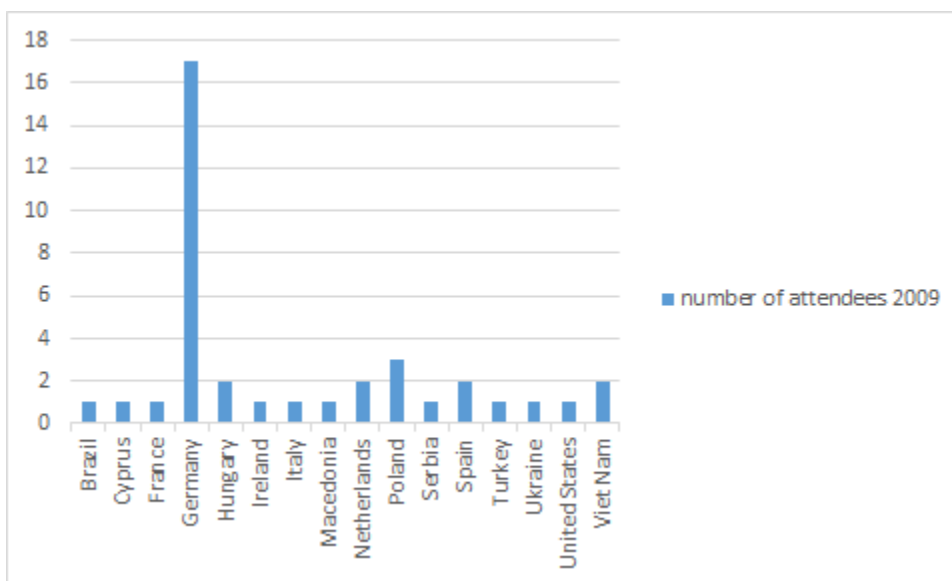
https://www.zotero.org/groups/training_and_education_in_dh

If people wish to add, edit, and / or remove items from the “Training and Education in DH” group's library, one can become a group member by a simple request via e-mail. The bibliography will be expanded in the duration of the PARTHENOS project and completed with tags. It can easily be included afterwards into concrete training and education plans or be used as a starting point for further research on training and education. If one wishes to get regular updates on new items in the library one can subscribe to a feed here:

<https://api.zotero.org/groups/593883/items/top?start=0&limit=25&format=atom&v=1>

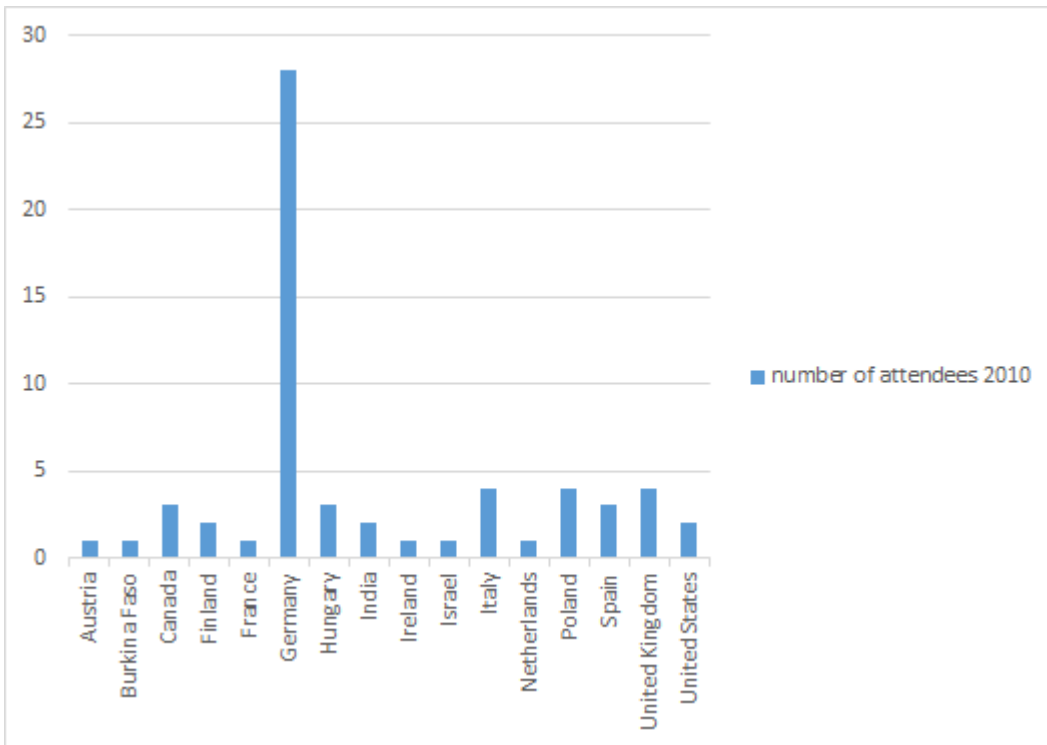
4.6. ESU - European Summer University in Digital Humanities (ESU)

The team from University of Leipzig collected data about the participants of the ESU and shared them with the T2.4 team. All data are made anonymous before the analysis. No conclusions can be drawn on the situation of an individual. Together with Stefanie Löpke from University of Leipzig the T2.4 team examined data about the attendees of the ESU. The data collected covers the years 2009, 2010, 2012, 2013, 2014 and 2015 and are derived from the conference management tool used to register attendees.⁶⁷ For the geographical distribution of attendees in the respective years see charts below.

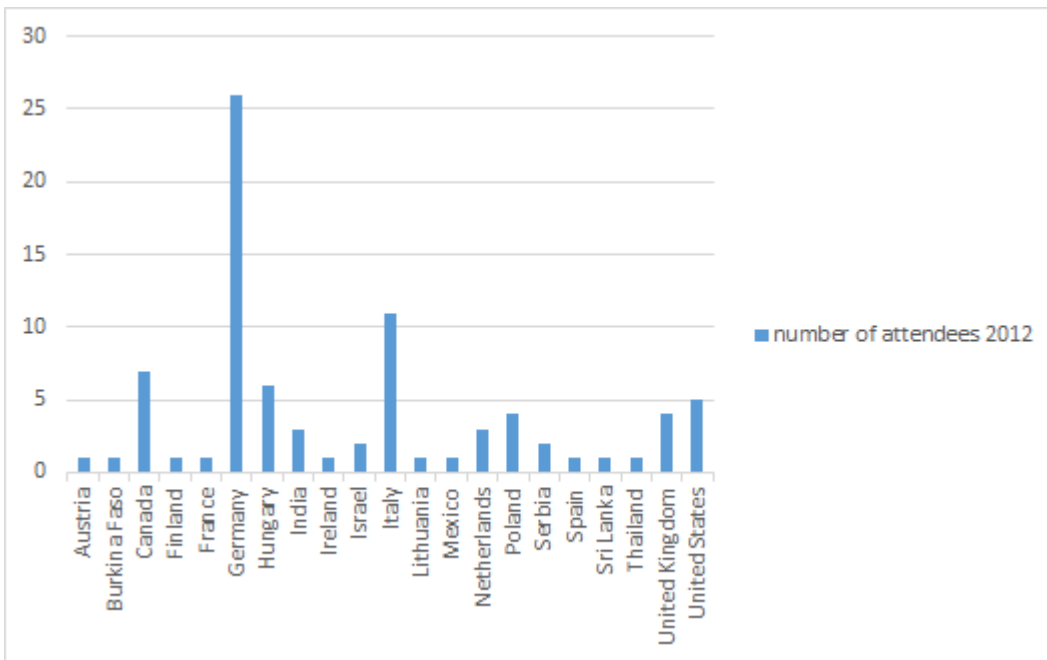


Number and geographical distribution of attendees in 2009 (n=38)

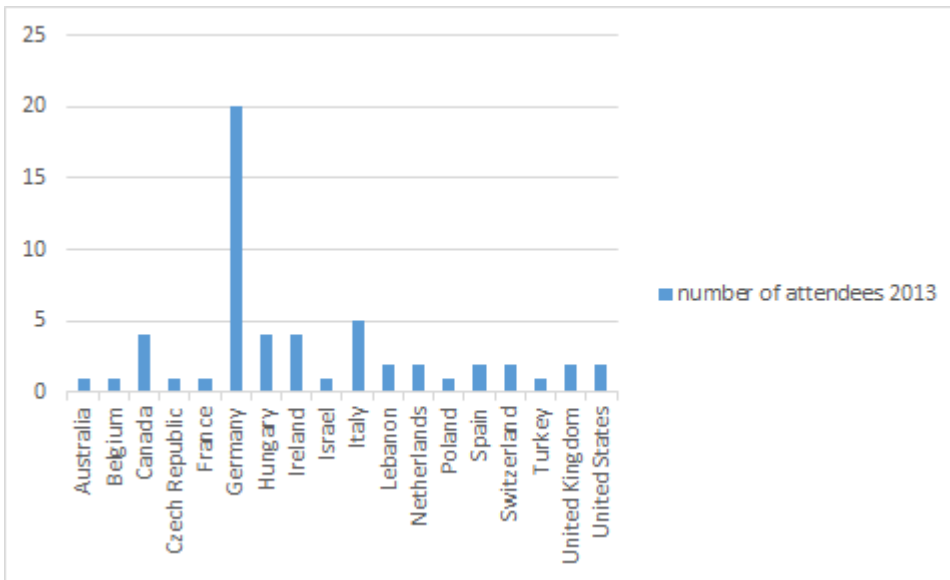
⁶⁷ There was no ESU in 2011 since another conference took place at the same time and place at Leipzig University.



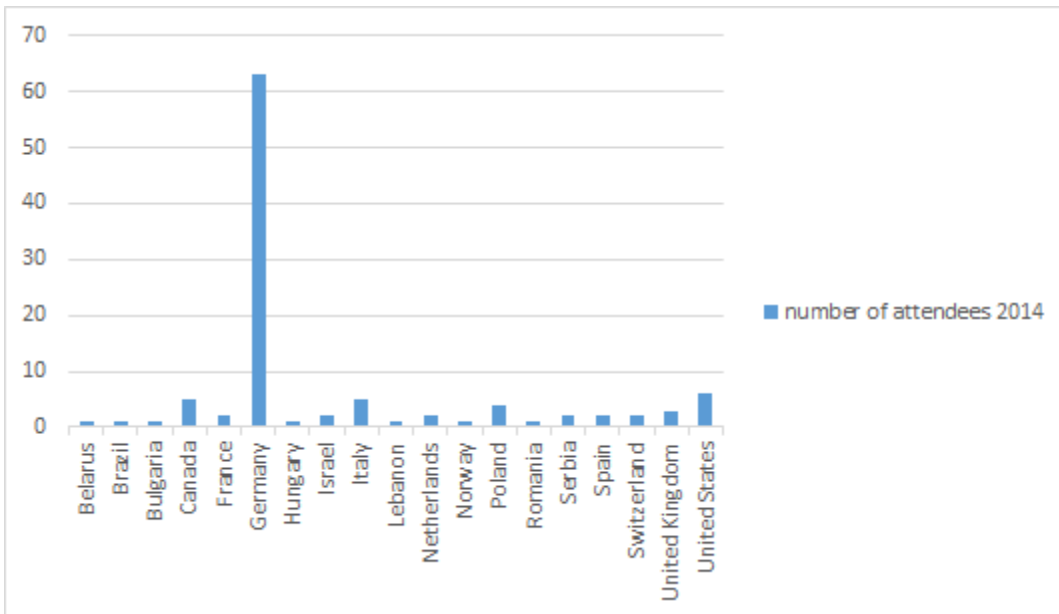
Number and geographical distribution of attendees in 2010 (n=61)



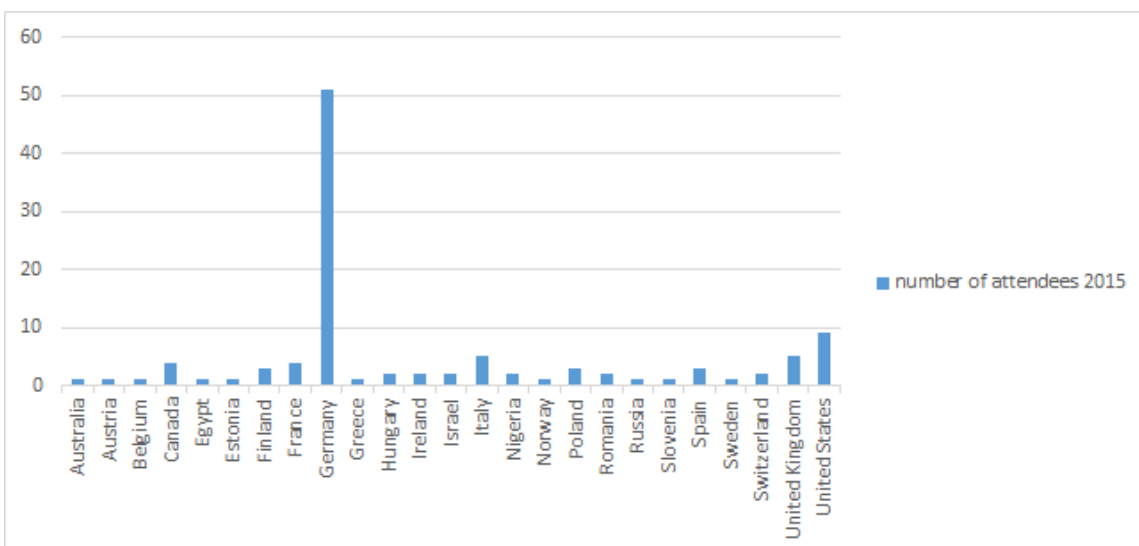
Number and geographical distribution of attendees in 2012 (n=83)



Number and geographical distribution of attendees in 2013 (n=56)

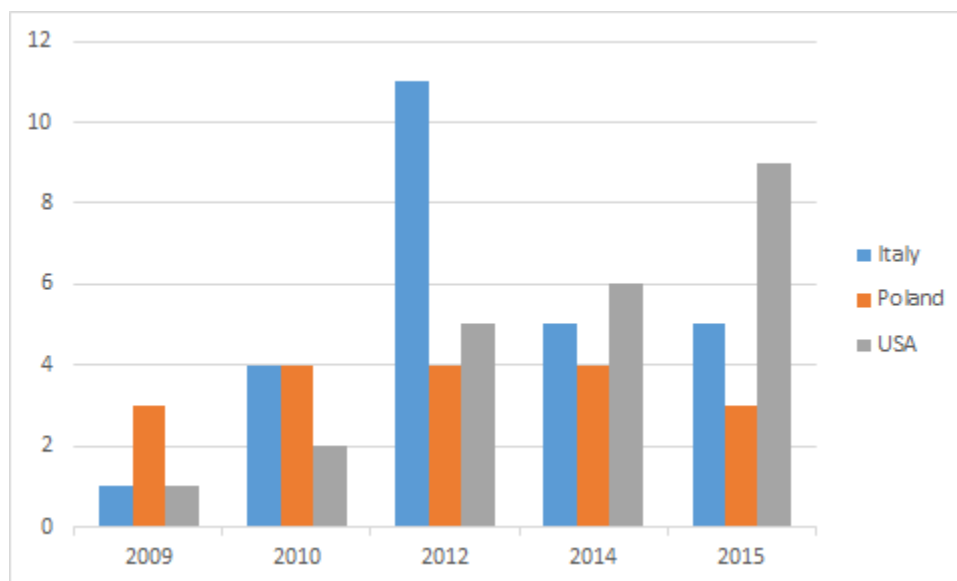


Number and geographical distribution of attendees in 2014 (n=105)



Number and geographical distribution of attendees in 2015 (n=109)

One can clearly see from the charts that the overall number of attendees as well as the geographical spread increases over time. Most participants came from Germany (n=185), followed by Italy (n=26), USA (n=23) and Poland (n=18). See chart below.



Even if the ESU is defined as having a focus on European researchers, people from the USA are obviously attracted to the workshops as well. The Summer University is well established and globally recognised. Thus it is a highly relevant cooperation partner when it comes to the test and implementation of developed training material.

Based on the organisational affiliation, one can draw first conclusions on the attendees' research areas. The whole spectrum ranges from Linguistics to Theology, and includes Library and Informations Sciences to Economics and Computer Science. For methodological reasons a detailed analysis is challenging. Data is only available for the years 2012, 2013, 2014 and 2015. Not all participants specified their affiliation, and institution names are listed as registered in the conference management tool by the respective attendees. Language may vary for that reason. Thus, no detailed analysis was conducted but a sample was taken for each year for some specific questions. This analysis revealed that the number of participants from Computer Science (or similar disciplines) is increasing over the years, whereas the number of people with an affiliation related to linguistics or literary studies is constantly high and forms the core of ESU participants. This means that for the other subject areas, the actual increase in the number of participants is due in a big part to participants from the field of Computer Sciences.

Regarding the career stages and/or academic degrees, data quality is also very heterogeneous. In 2009, 38 attendees took part in the ESU. Of those attendees, 28 provided information on their stage of career / academic degree. Not surprisingly the most dominant

group are the participants with a PhD or higher. The data shows that most attendees are researchers working within the humanities and cultural heritage disciplines. However, there is a small number of people working in the project management area as well as in cultural heritage institutions (CHI). Taking this into account, the presented target groups seem to fit that audience very well.

All this is closely related and dependant on the offered workshops. In general, people apply for a one or two week workshop. The workshop summary will then be published on the ESU website and people choose the workshops they are most interested in and apply for them. By analysing the offered workshops, one gets a first idea of topics that are “hot topics” and those that raise continuous interest and seem to be more fundamental in the field of training and education.

Workshops that generate most interest (at least 10 participants or more) are the following:

2009: Corpus and Corpus Analysis in Language (and Literary) Sciences

2010: From Document Engineering to Scholarly Web Projects; Digital History and Culture methods, sources and future looks

2012: Computing Methods applied to DH; XML Markup and Document Structuring Query
in

Text Corpora Stylometry; Computer-Assisted Analysis of Literary Texts

2013: Computing Methods applied to DH: TEI-XML Markup and CSS/XSLT Rendering;
Editing in the Digital Age: From Script, to Print, to Digital Page

2014: Advanced Topics in Humanities Programming with Python

2015: Methods and Tools for the Corpus Annotation of Historical and Contemporary Writ-
ten

Texts; Basic Statistics and Visualization with R; XML-TEI encoding, structuring and render-
ing; Comparing Corpora; Digital Editions and Editorial Theory: Historical
Texts and Documents

A statistically valid interpretation is hardly feasible due to the small data sample. However, if one considers this as a trend, the data reveals two trends:

- 1) There is a tendency towards more computer sciences orientated training (e.g. “Computing Methods applied to DH”; “Advanced Topics in Humanities Programming with Python”).
- 2) ESU attendees are interested in specifically linguistic related topics (e.g. “Corpus Analysis in Language and Literary” or “Sciences or Comparing Corpora”).

4.7. Conclusions

The feedback provided by the PARTHENOS community and the document analysis revealed a preference for face-to-face meetings. The combination of workshops, summer schools or Skype conferences with moderated distance-learning modules like webinars seem to be the most common and promising way of implementation. The experiences have shown that a human moderator/contact person to ask questions is one characteristic for a successful training module. Online tutorials or written documentations without a point of contact are classified as of minor effectiveness.

The topics for offered training courses and material mostly derive from surveys conducted within the projects. Thus training needs are mainly focused on concrete infrastructure or tools developed in the projects. This is not surprising since all projects, except DASISH, did not aim at a systematic development of training or education services. If work was conducted in the projects towards training and education needs, then this was mostly motivated by the intention to improve the developed tools / infrastructures towards usability. A general motivation to systematically develop, organize and set up training and education modules in a more comprehensive manner still exists among the people working in the projects but could not be realized due to the typical project characteristics like fixed project duration, less manpower for additional work, questions of sustainability of project results etc.

On the other hand, the experience of a PARTHENOS partner, namely ARIADNE, revealed that training offers on broad subjects and more general topics attract little interest from the community. Thus the opportunity for tailored training and expert input for researchers' own projects were highly appreciated. Consequently, the ARIADNE training will focus on more specific topics in the future.

One important point suggested from the community is the advertisement of training. Finding the appropriate channels for the focused target groups will be a crucial factor for the success of training events. Close cooperation between WP2 – T2.4 and T2.5 as well as a strong collaboration between WP7 and WP8 will help the project to cope with that challenge.

Training fees might also be considered as a limiting factor for the participation in summer schools etc. By setting up face-to-face events, financing for the target group needs to be kept in mind.

The projects of the PARTHENOS cluster have all created training materials and opportunities for their users. The range of practices is diverse, however: diverse in the focus of the training materials, diverse in the modalities deployed to deliver the training and diverse in the perceived value and response to these interventions. In addition, the training that has been delivered has been largely ad hoc, driven by the other activities of the project, rather than by an overt plan for training and education in the context of the research infrastructure project's goals.

The user study as well as the survey replies revealed that approaches need to be developed that raise the awareness about DH in general and research infrastructures (e.g. CLARIN - which was the use case in the Danish study) in particular. Insight into Humanities researchers epistemological practices are needed to successfully develop infrastructures and make digital tools and methods attractive and easy to use for the target groups.

There are a number of already existing inventories presenting and structuring training and education materials. To link activities to this platforms seems reasonable.

One main aspect that the ESU data exposed is the attractiveness of trainings that help to develop and maintain skills in the field of computer sciences. Understanding DH methods or digital tools in particular is not in the focus of main interest. However, highly discipline-specific training, for example in the area of linguistics, is the second thematic core area of interest. Surprisingly the ESU data indicate that apart from the targeted audience in Europe, researchers from the USA are among the top 3 nations (apart from Germany itself) that participate in the Summer University. If this trend can be followed up and be verified it might be interesting for the transnational activities in the development of the PARTHENOS training and education plan.

5. Communication requirements

Main authors: Juliane Stiller and Jenny Oltersdorf, responsible for the final edition: Claus Spiecker (all FHP)

5.0. Introduction

This chapter presents the work done by the members of Parthenos task 2.5 (T2.5). Their objective is the collection of the communication requirements expressed by the PARTHENOS community, both with regard to scientific communication and dissemination strategies. The authors (KNAW-NIOD, CNRS, and FHP) collected and analysed a preliminary sample of relevant scientific communication platforms such as e-journals and repositories to obtain the criteria for their evaluation. The overall goal was to gather information on relevant journals and repositories in the field and to enable their evaluation with regard to their attractiveness for Digital Humanities researchers for publishing their findings there.

Regarding dissemination strategies, a number of dissemination reports from PARTHENOS-related projects were gathered and scrutinized to get insight into the dissemination strategies. The results of both strands (analysis of scientific communication as well as the analysis of dissemination strategies) will support the work of WP 8.

The methodological approach, including the components for document analysis, can be found in section 5.1. In the first part of the section, the approach for scholarly communication is described. The second part of 5.1 deals with the methodology regarding dissemination strategies. It is followed by section 5.2, which presents the first set of relevant scholarly e-journals and repositories, their analysis and a set of criteria for their evaluation. The results of the analysed dissemination reports are presented in section 5.3, which is divided into three parts: 5.3.1 is about target groups, 5.3.2 deals with the dissemination activities and 5.3.3 is about success criteria for dissemination strategies. The last paragraph 5.4 concerns the next steps and work to be done in T2.5.

5.1. Method

One objective of T2.5 is the collection of relevant scientific publication platforms in the field of Digital Humanities and the provision of criteria for their evaluation. Task members started to gather information on significant journals and repositories. The sample is based on

the scholarly experiences of T2.5 and Parthenos WP8 members. Work in this field was undertaken in very close cooperation with WP8.

According to the PARTHENOS Description of Work, T8.3 will evaluate the need for the creation of a scientific e-journal in the Digital Humanities research area. Hence, it was agreed to develop a set of evaluation criteria in T2.5 first. These criteria aim to provide a means of establishing how desirable the publication of research output in a certain e-journal / repository is. The resulting criteria list is inspired by and grounded on the analysis of the proposed e-journals and repositories. The criteria will be refined, discussed and adjusted by the WP8 team.

To extract dissemination requirements, existing communication and dissemination plans from ESFRI projects and other relevant initiatives were reviewed. This approach is in line with the general decision of WP2 to base its analysis on documents rather than to conduct new surveys. The following projects and initiatives were reviewed with regard to their dissemination strategies: EHRI; DARIAH, TextGrid; EUDAT, DASISH, Europeana Cloud, AR-IADNE, DM2E, Apex, Cendari, DCH-RP, CLARIN. Each of the communication and dissemination plans were examined and a template for document analysis was developed accordingly. The template includes four main components that derive from the documents and are discussions within the task.

The components are:

- audience / target groups,
- message / purpose,
- methods / activities,
- impact / success criteria.

Audience is the target or stakeholder group. Typically this can be internal and external stakeholders, researchers in various career stages, policy makers, service and content providers, public.

Message / purpose is the goal of the dissemination activity that is particularly targeted on specific groups. It entails the dissemination of research results from the project group, dissemination of the offerings of the project groups in terms of software or training, etc., increasing traffic to website or collaboration with researchers beyond the project's context.

Methods / activities are the ways in which the message is delivered to the audience. This can be via different channels such as social media activities, e.g. Twitter, Facebook, Instagram, mailing lists, websites, blogs, collaborative events, workshops and conferences, dis-

tribution of printed material, publications, interviews, videos, training activities, e.g. webinars, summer schools, workshops and networking events.

Impact / success criteria determine the change that was induced by the activities. It also lists which metrics are used to measure such a change. This could be usage statistics, number of subscribers, number of participants at events, or collaboration requests.

Nine relevant reports from the projects EHRI, Europeana Cloud, DM2E, DARIAH-DE, EUDAT, CENDARI, CLARIN and DCH-RP were scrutinized. As the dissemination level for the reports is heterogeneous and not all are publicly available, the authors decided to give full references without links. If the dissemination reports are needed for further analysis, we recommend to request the respective authors of the dissemination reports for this directly.

- 1) Tellegen, Jan Willem. 2011. 'Publicity & Dissemination Strategy, Concept for Identity, Branding & Graphic Design'. Deliverable D8.1. EHRI.
- 2) Moyle, Martin, Marnix van Berchum, and Friedel Grant. 2013. 'Stakeholder Engagement Plan'. Deliverable D6.1. Europeana Cloud.
- 3) Sam Leon, and Violeta Trkulja. 2013. 'Dissemination and Engagement Plan'. Deliverable D4.4. DM2E.
- 4) Benardou, Agiatis, Sally Chambers, Nephelie Chatzidiakou, Jill Cousins, Alastair Dunning, Stefan Ekman, Vicky Garnett, et al. 2014. 'Researcher Communication Plan'. Deliverable 6.3. Europeana Cloud.
- 5) Mathias Göbel, Nadja Grupe, Christian Heise, Maren Köhlmann, Katharina Meyer, Markus Neuschäfer, Stefan Schmunk, and Sibylle Söring. 2014. 'DARIAH-DE und TextGrid. Disseminationsstrategie inklusive Marketingkonzept sowie DARIAH-DE Open Mission Statement und Publikationsstrategie'. Report 7.2 / 7.3.3. DARIAH-DE 2 / TextGrid 3.
- 6) Madeleine Gray, Hilary Hanahoe, and Adam Carter. 2015. 'Annual Dissemination and Outreach Report 3'. Deliverable D3.3.3. EUDAT.
- 7) O'Brien, Catherine. 2012. 'Dissemination Strategy'. Deliverable D2.2. CENDARI.
- 8) Elisa Sciotto, Luca Martinelli, Patrizia Martini, and Sara di Giorgio. 2014. 'Report on Dissemination Activities'. Deliverable 2.3.2. DCH-RP.
- 9) Maegaard, Bente, Hanne Fersøe, and Lina Henriksen. 2009. 'Requirements and Best Practice for Transnational Coordination and Collaboration with Third Parties'. D8S-3.1. CLARIN.

5.2. Scholarly communication

The following platforms (e-journals and repositories) for scientific communication were collected by the T2.5 team and colleagues from WP8. These have been mentioned as being

relevant for the Digital Humanities sector from the experience of the team. The list is not aiming at a complete coverage of all existing e-journals and repositories but as a first starting point for further analysis.

- *Digital Humanities Quarterly (DHQ)*, <http://www.digitalHumanities.org/dhq/>, is an open-access, peer-reviewed, digital journal covering all aspects of digital media in the Humanities. It is published by the Alliance of Digital Humanities Organizations.
- *Digital Studies*, http://www.digitalstudies.org/ojs/index.php/digital_studies, *Digital Studies / Le champ numérique* (ISSN 1918-3666) is a refereed academic journal serving as a formal arena for scholarly activity and as an academic resource for researchers in the Digital Humanities. DS/CN is published by the Société canadienne des humanités numériques (CSDH/SCHN), a partner in the Alliance of Digital Humanities Organisations (ADHO).
- *Digital Literary Studies*, <http://journals.psu.edu/dls>, is an international peer-reviewed interdisciplinary publication with a focus on those aspects of Digital Humanities primarily concerned with literary studies.
- *Journal of the Text Encoding Initiative*, <http://journal.tei-c.org/>, is the official journal of the Text Encoding Initiative Consortium. It publishes selected papers from the annual TEI Conference and Members' Meeting and special issues based on topics or themes of interest to the community or in conjunction with special events or meetings associated with TEI.
- *DHCommons*, <http://dhcommons.org/journal/issue-1>, overlays and interacts with the *DHCommons* project registry and will provide peer review for mid-stage digital projects. The most ambitious aim of *DHCommons* is to make visible the important developmental work that often goes unseen in the midst of a DH project and to help DH scholars claim departmental, disciplinary, and institutional credit for that labor. *DHCommons* will become the robust and recognizable system of academic credit that its practitioners require.
- *Journal of Digital Media and Literacy*, <http://www.jodml.org/>, is published by the James L. Knight School of Communication at Queens University of Charlotte with support from the John S. and James L. Knight Foundation. *JoDML* is an academic, peer-reviewed journal publishing traditional research articles alongside hybrid, mixed-media articles and creative digital projects. The goal is to examine the ways people use technology to create, sustain, and impact communities on local, national and global levels. Broadly defined, digital and media literacy refer to the ability to access, share, analyse, create, reflect upon, and act with media and digital information
- *Kairos: A Journal of Rhetoric, Technology, and Pedagogy*, <http://kairos.technorhetoric.net/>, is a refereed open-access online journal exploring the intersections of rhetoric, technology, and pedagogy. The journal reaches a wide audience - the interna-

tional readership typically runs about 4,000 readers per month. *Kairos* publishes bi-annually, in August and January, with regular special issues in May.

- *Hal Archive*, <https://hal.archives-ouvertes.fr/>, is an open archive where authors can deposit scholarly documents from all academic fields. It covers all academic fields but most documents are from Humanities and Social Sciences.
- International Journal of Humanities and Arts Computing – <http://www.eupublishing.com/journal/ijhac>, it focuses both on conceptual or theoretical approaches and on case studies or essays demonstrating how advanced information technologies further scholarly understanding of traditional topics in the arts and Humanities.
- Journal of Digital Humanities – <http://journalofdigitalhumanities.org/>, is a comprehensive, peer-reviewed, open access journal that features the best scholarship, tools, and conversations produced by the Digital Humanities community in the previous trimester.
- Digital Scholarship in the Humanities – <http://dsh.oxfordjournals.org/>, is an international, peer reviewed journal that publishes original contributions on all aspects of digital scholarship in the Humanities including, but not limited to, the field of what is currently called the Digital Humanities.
- Frontiers in Digital Humanities – <http://journal.frontiersin.org/journal/digital-Humanities>, publishes articles on the most outstanding discoveries in all the research areas where computer science and the Humanities intersect, with the aim to bring all relevant Digital Humanities areas together on a single, open-access platform.

The following criteria are derived from the analysis of the above-mentioned platforms for scholarly communication (e-journals and repositories). They are a suggestion for the evaluation of scholarly journals and / or repositories in WP8.

The most obvious and important criterion is the **domain**. For the envisaged target groups within PARTHENOS, a Digital Humanities relation or the openness to create one is mandatory.

Closely related with the domain is the criterion of **covered topics / subjects**. The attraction to publish in a journal or repository may vary with its coverage. This in turn can range from a wide scope containing all DH (Digital Humanities) topics (including methods and tools) to a more narrow focus on a specific topic such as the *Journal of the Text Encoding Initiative* which is the dissemination channel of a special initiative and consequently focused on TEI-related topics. Besides, there is a need to distinguish between journals that focus on specific Humanities subjects and where DH is just one aspect of research method among other topics and those that have a clear DH focus covering all the disciplines.

The criterion of **regional / international coverage** relates particularly with regard to topics – these can range from international importance to a highly regional-specific spotlight. This

is directly linked with the targeted group of authors and the audience determined by the favoured **languages** of the journal / repository.

The **formats / outputs accepted** are a fifth decisive factor that may be relevant for the evaluation of a journal / repository for the DH community. Relevant questions are for instance: Are submissions in a wide range of formats accepted? Are traditional academic formats like essays, articles, book reviews and so forth encoded in a TEI-compatible format for longevity and ease of management? Will the web platform support the infrastructure for ongoing blogging and commenting on publications or ad hoc reviews of the literature as in *Digital Humanities Quarterly*? All these are crucial questions when it comes to that criterion.

A further benchmark in the evaluation of e-Journals / repositories are **accepted publication types**. Typically, academic articles, literature and research reviews will be welcome. It might be of interest if working papers, field synopses, editorials and provocative opinion pieces or reviews of websites, new media art installations, Digital Humanities systems and tools are also accepted publication types.

The criterion of **open access** is relevant as well as a guaranteed quality management based on critical **peer review**.

A further point on the evaluation of a journal / repository is the **ability to be quantitatively analysed** by typical bibliometric and webometric methods. Thus, the availability in citation indices like Scopus or Web of Science or the provision of usage statistics and downloads is helpful.

5.3. Requirements for dissemination

This section identifies the target groups mentioned in the dissemination reports, the dissemination activities itself and the success criteria to measure the achievements of the activities.

5.3.1. Target groups

Since the extracted target groups from the documents vary widely in wording and context (target groups were not exclusive, e.g. parallel mentioning of users and researchers, even if researchers are seen as users), the final list of target groups is based on a process of three steps. Firstly, all mentioned target groups were extracted from the documents. Secondly, double mentions and synonyms were consolidated and thirdly, generic terms were

created and terms were grouped by them (whenever possible). The result appears as follows:

- user (including researchers and scholars at different stages of their career and different kinds of research institutions)
- content providers and content aggregators
- providers of research infrastructures
- project partners (within the same project)
- external projects
- media (including national broadcast and publishers)
- political decision makers and research funding agencies
- GLAM institutions (gallery, library, archive, museum)
- technical developers
- industry representatives
- private organizations
- the general public.

5.3.2. Dissemination activities

Even if there are numerous dissemination activities mentioned in the documents, a group of five emerged as the most evident. These are, first and foremost, dissemination activities via the project's website – announcements of new findings, etc. Secondly, partners' institutional websites are used for the dissemination of information. Thirdly, newsletter and fourthly press releases are common means when it comes to dissemination strategies. Finally, networking and consulting at conferences in various phases of the projects was also mentioned as one of the most important activities regarding the dissemination of project results.

Use of social media, journal publications, wiki and other collaborative tools, working groups, posters and booklets were mentioned as well. However, they seem to be of lesser importance and they are dedicated towards special target groups in contrast to the first mentioned group of five applying to nearly every target group and message.

5.3.3. Success criteria

Setting up success criteria turned out to be a challenge. Even if they are mentioned as being important, only a few criteria could be found in the Research Infrastructures (RIs) dissemination reports. The distribution among the analysed RIs dissemination plans is skewed since only three projects listed means for quantitative measurement of success. All success criteria use quantitative means based on the counting of a characteristic value.

Most often, website statistics (usage, links etc.), increased numbers of social media followers, likes (on Facebook for example), comments and number of presentations and workshops held were mentioned. Respondents to online questionnaires, the number of applications to transnational access training programmes as well as the number of people attending workshops and seminars were also listed. The quantitative measurement of written output such as the number of project brochures distributed, number of press releases / articles from the project itself as well as number of press releases / articles referencing the project constituted a third group of success criteria.

5.4. Next steps

The presented evaluation criteria are the result of the intensive cooperation of T2.5 and T8.3. It will be used by T8.3 as a basis to evaluate the need for the creation of a scientific e-journal in the Digital Humanities research area. If a journal already exists that answers to all the criteria, the next step might be to find a way to collaborate and support the efforts of that platform. If there is no such journal yet, and the decision will be not to create a new scientific e-journal, the collected information and the found criteria will help to develop alternate measures to support and improve scientific communication in the PARTHENOS community and beyond.

The dissemination strategy of PHARTHENOS as a project will be based on the findings, too. So participation at conferences, publications, press releases, and the use of social media, usage statistics etc. are on the target and will be closed analysed regarding to their impact and if possible in regard to their quantitative effects.

References

- '15cBOOKTRADE'. 2015. Accessed December 4 2016. <http://www.modlangs.ox.ac.uk/research/15cBooktrade/>.
- Abbagnano, Nicola. 1959. *Problemi di sociologia*. Torino: Taylor.
- ADS. 2016. 'Archaeology Data Service'. Accessed January 27 2016. <http://archaeologydataservice.ac.uk/>.
- AFNLP. 2011. 'Proceedings of the Fifth International Joint Conference on Natural Language Processing (IJCNLP 2011)'. Chiang Mai, Thailand: Asian Federation of Natural Language Processing. <https://aclweb.org/anthology/I/I11/I11-1000.pdf>.
- AGORA. 2016. 'Scholarly Open Access Research in European Philosophy'. Accessed January 27 2016 at http://cordis.europa.eu/project/rcn/191888_en.html.
- Amedeo, Enrico. 2007. *UML. Unified Modeling Language*. Pocket. Milan, Italy: Apogeo.
- Anderson, Sheila, Jakub Benes, Pavlina Bobic, Anna Bohn, Andrea Buchner, Valentine Charles, Emiliano Degl'Innocenti, et al. 2014. 'Archive Directory'. Deliverable 5.1. CENDARI. http://www.cendari.eu/sites/default/files/CENDARI-D5.1-Archive-Directory_final.pdf.
- Anderson, Sheila, and Tobias Blanke. 2013. 'Intermediating the Human and Digital: Researchers and the European Holocaust Research Infrastructure (EHRI)'. EHRI. http://www.ehri-project.eu/webfm_send/273.
- Anderson, Sheila, Reto Speck, Petra Links, Agiatis Benardou, Panos Constantopoulos, and Costis Dallas. 2013a. 'Data Requirements'. Deliverable D.16.5. EHRI. <https://goo.gl/KYn8Yn>.
- . 2013b. 'Functional Specifications'. Deliverable D.16.6. EHRI. <https://goo.gl/C33OT4>.
- Angelis, Stavros, Agiatis Benardou, Panos Constantopoulos, Costis Dallas, Aggeliki Fotopoulou, Dimitris Gavrilis, Natalia Manola, Sheila Anderson, Reto Speck, and Petra Links. 2013. 'Researcher Practices and User Requirements'. Deliverable D.16.4. EHRI. <https://goo.gl/HvDRra>.
- Angelis, Stavros, Agiatis Benardou, Nephelie Chatzidiakou, Panos Constantopoulos, Costis Dallas, Alastair Dunning, Jennifer Edmond, et al. 2015. 'User Requirements

Analysis and Case Studies Report. Content Strategy Report'. Joint Deliverable 1.3 and 1.6. Europeana Cloud.

http://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/Europeana_Cloud/Deliverables/D1.3%20D1.6%20User%20Requirements%20Analysis%20and%20Case%20Studies%20Report%20Content%20Strategy%20Report.pdf.

APEX. 2016. 'Archives Portal Europe Foundation'. Accessed January 28 2016 at <http://apex-project.eu/>.

ARIADNE. 2016. 'Advanced Research Infrastructure for Archaeological Dataset Networking in Europe'. Accessed January 27 2016 at <http://www.ariadne-infrastructure.eu/>.

Arjan Hogenaar, Heiko Tjalsma, and Mike Priddy. 2011. 'Research in the Humanities and Social Sciences'. In *Studies on Subject-Specific Requirements for Open Access Infrastructure*, edited by Christian Meier zu Verl and Wolfram Horstmann, 165–213. Bielefeld, Germany: Universitätsbibliothek. <http://pub.uni-bielefeld.de/publication/2458719>.

Arofan, Gregory. 2011. 'The Data Documentation Initiative (DDI): An Introduction for National Statistical Institutes'. Open Data Foundation. http://odaf.org/papers/DDI_Intro_forNSIs.pdf.

Arranz, Victoria, Daan Broeder, Bertrand Gaiffe, Maria Gavrilidou, Monica Monachini, and Thorsten Trippel, eds. 2012. 'Describing LRs with Metadata: Towards Flexibility and Interoperability in the Documentation of LR'. In *LREC 2012, Eighth International Conference on Language Resources and Evaluation*. Istanbul, Turkey. <http://www.lrec-conf.org/proceedings/lrec2012/workshops/11.LREC2012%20Metadata%20Proceedings.pdf>.

Assante, Massimiliano, Leonardo Candela, Donatella Castelli, Paolo Manghi, and Pasquale Pagano. 2015. 'Science 2.0 Repositories: Time for a Change in Scholarly Communication'. *D-Lib Magazine* 21 (1/2). <http://www.dlib.org/dlib/january15/assante/01assante.html>.

ATHENA. 2010. 'IPR Guide | Athena Europe IPR'. *ATHENA*. http://athena.iprguide.org/lang_en/page/home-page.

———. 2013. 'Access to Cultural Heritage Networks across Europe'. September 4 2016 at <http://www.athenaeurope.org/>.

- . 2013b. 'ATHENA Step-by-step IPR Guide'. 2013. Accessed July 22 2016.
<http://www.athenaeurope.org/index.php?en/192/step-by-step-ipr-guide>.
- AthenaPlus. 2016. 'Access to Cultural Heritage Networks for Europeana'. Accessed January 26 2016 at <http://www.athenaplus.eu/>.
- Attwood, T. K., D. B. Kell, P. McDermott, J. Marsh, S. R. Pettifer, and D. Thorne. 2010. 'Utopia Documents: Linking Scholarly Literature with Research Data'. *Bioinformatics* 26 (18): i568–74. doi:10.1093/bioinformatics/btq383.
- "Audit and Certification of Trustworthy Digital Repositories. Recommended Practice." 2011. CCSDS 652.0-M-1. CCSDS.
<http://public.ccsds.org/publications/archive/652x0m1.pdf>.
- 'Available Rights Statements'. 2016. Accessed January 28 2016 at
<http://pro.europeana.eu/share-your-data/rights-statement-guidelines/available-rights-statements>.
- Bardi, Alessia, and Paolo Manghi. 2014. 'Enhanced Publications: Data Models and Information Systems'. *LIBER Quarterly* 23 (4): 240–73.
- . 2015a. 'Enhanced Publication Management Systems: A Systemic Approach Towards Modern Scientific Communication'. In *Proceedings of WWW '15, the 24th International Conference on World Wide Web*, 1051–52. Geneva, Switzerland.
<http://dl.acm.org/citation.cfm?id=2742026>.
- . 2015b. 'A Framework Supporting the Shift from Traditional Digital Publications to Enhanced Publications'. *D-Lib Magazine* 21 (1/2). doi:10.1045/january2015-bardi.
- Belice Baltussen, Lotte, Maia Borelli, Irene Scaturro, Ferruccio Marotti, Emanuele Bellini, Jaap Blom, Johan Oomen, et al. 2010. 'User Requirements and Use Cases'. DE2.1.1. ECLAP.
<http://cordis.europa.eu/docs/projects/cnect/1/250481/080/deliverables/001-ECLAPDE211UserRequirementsandUseCasesv10.pdf>.
- Benardou, Agiatis, Sally Chambers, Nephelie Chatzidiakou, Jill Cousins, Alastair Dunning, Stefan Ekman, Vicky Garnett, et al. 2014. 'Researcher Communication Plan'. Deliverable 6.3. Europeana Cloud.
http://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/Europeana_Cloud/Deliverables/D6.3%20European%20Research%20Communications%20Plan.pdf.

- Benes, Jakub, Pavlina Bobic, Nadia Boukhelifa, Emiliano Degl'Innocenti, Jean-Daniel Fekete, Jonathan Gumz, Mark Hedges, et al. 2013. 'Functional Description: Visualization'. Deliverable D8.2. CENDARI.
- Beneš, Jakub, Kathleen Smith, Andrea Buchner, Klaus Richter, and Pavlina Bobič. 2014. 'Report on Archival Research Practices'. D4.1. CENDARI.
http://www.cendari.eu/sites/default/files/CENDARI_D4.1-Report-on-Archival-Practices%20%281%29_0.pdf.
- Ben Kaden, and Michael Kleineberg. 2015. 'Erweiterte Publikationen in Den Geisteswissenschaften. Zwischenergebnisse Des DFG-Projektes Fu-Push'. In *Proceedings of DHd 2015, 2. Jahrestagung Des Verbandes Digital Humanities Im Deutschsprachigen Raum*. Graz, Austria. doi:10.5281/zenodo.15432.
- Biblioteca nazionale centrale di Firenze. 2016. 'Nuovo soggettari'. Accessed August 8 2016 at <http://thes.bncf.firenze.sbn.it/>.
- Biblissima. 2016. 'Bibliotheca bibliothecarum novissima'. Accessed August 8 2016 at <http://biblissima-condorcet.fr>.
- Bird, Steven, and Gary Simons. 2003. 'Project Muse: Seven Dimensions of Portability for Language Documentation and Description'. *Language* 79 (3): 557–82. Pre-print available via <http://www-01.sil.org/~simonsg/preprint/Seven%20dimensions.pdf>.
- Bøe, Marianne, Rød, Linn-Merethe, Parra, Carla, De Smedt, Koenraad, Kvalheim, Vigdis, Utaaker Segadal, Katrine, Kvamme, Trond, Dione, Bamba, Lyse Samdal, Gunn Inger. 2014. 'Handbook on legal and ethical issues for SSH data in Europe', Part II. Accessed July 22 2016 at http://dasish.eu/publications/projectreports/DASISH_D6.5_februar_2015.pdf
- Boonstra, Onno, Leen Breure, and Peter Doorn. 2004. *Past, Present and Future of Historical Information Science*. Amsterdam: NIWI-KNAW.
<http://www.ahc.ac.uk/docs/pastpresentfuture.pdf>.
- British Library. 2016. 'Incunabula Short Title Catalogue'. Accessed August 8 2016 at <http://www.bl.uk/catalogues/istc/>.
- Broeder, Daan, Ineke Schuurman, and Menzo Windhouwer. 2014. 'Experiences with the ISOcat Data Category Registry'. In *Proceedings of LREC 2014, Ninth International Conference on Language Resources and Evaluation*. Reykjavik, Iceland.
http://www.lrec-conf.org/proceedings/lrec2014/pdf/153_Paper.pdf.

- Brouillard, Julien, Claire Loucopoulos, and Corinne Szeitszneider. 2014. 'Analysis, Scenarios Use Cases, Opportunities of Innovative Services for DCH, and Future Development'. Deliverable D7.2. AthenaPlus.
<http://www.athenaplus.eu/getFile.php?id=365>.
- Burnard, Lou. 2011. 'Editorial Guidelines'. Deliverable D4.1. AGORA.
<https://goo.gl/nJyR4V>.
- Calzolari, Nicoletta, Riccardo Del Gratta, Gil Francopoulo, Joseph Mariani, Francesco Rubino, Irene Russo, and Claudia Soria. 2012. 'The LRE Map. Harmonising Community Descriptions of Resources.' In *Proceedings of LREC 2012, Eighth International Conference on Language Resources and Evaluation*, 1084–89. Istanbul, Turkey. http://lrec.elra.info/proceedings/lrec2012/pdf/769_Paper.pdf.
- Calzolari, Nicoletta, Monica Monachini, and Valeria Quochi. 2011. 'Interoperability Framework: The FLaReNet Action Plan Proposal'. In *Proceedings of Workshop on Language Resources, Technology and Services in the Sharing Paradigm*. Chiang Mai, Thailand. http://www.aclweb.org/website/old_anthology/W/W11/W11-33.pdf#page=57.
- Candela, Leonardo, Donatella Castelli, Paolo Manghi, and Alice Tani. 2015a. 'Data Journals: A Survey'. *Journal of the Association for Information Science and Technology* 66 (9): 1747–62. doi:10.1002/asi.23358.
- . 2015b. 'Data Journals: A Survey'. *Journal of the Association for Information Science and Technology* 66 (9): 1747–62. doi:10.1002/asi.23358.
- CARMEN. n.d. 'The Worldwide Medieval Network'. <http://www.carmen-medieval.net/>.
- Carr, Nicholas. 2009. 'Is Google Making Us Stupid'. *The Atlantic*.
<http://www.theatlantic.com/magazine/archive/2008/07/is-google-making-us-stupid/306868/>.
- Carusi, Annamaria, and Torsten Reimer. 2010. 'Virtual Research Environment - Collaborative Landscape Study'. UK: JISC.
<http://www.webarchive.org.uk/wayback/archive/20140615234259/http://www.jisc.ac.uk/media/documents/publications/vrelandscapeporeport.pdf>.
- CENDARI. 2013a. 'Functional Description, Portal and VRE'. Deliverable D8.1. CENDARI.

- . 2013b. 'Domain Use Cases'. Deliverable D4.2. CENDARI. <https://goo.gl/y3xU8A>. http://www.cendari.eu/sites/default/files/CENDARI_D4.2%20Domain%20Use%20Cases%20final.pdf.
- . 2016. 'Collaborative European Digital Archival Research Infrastructure'. Accessed January 26 2016 at <http://www.cendari.eu/>.
- . 2016. 'TRAME. Text and manuscript transmission of the Middle Ages in Europe'. Accessed August 8 2016 at <http://git-trame.fefonlus.it>.
- CEOS / WGISS / DSIG. 2011. 'Data Lifecycle Models and Concepts'. Committee on Earth Observation Satellites (CEOS), Working Group on Information Systems and Services (WGISS, Data Stewardship Interest Group (DSIG). <http://wgiss.ceos.org/dsig/whitepapers/Data%20Lifecycle%20Models%20and%20Concepts%20v8.docx>.
- CESSDA. 2015. "CESSDA Data Archives and Digital Preservation." CESSDA User Guide for Digital Preservation. <http://cessda.net/CESSDA-Training/Data-Archives-and-Digital-Preservation>.
- CHARISMA. 2016. 'Cultural Heritage Advanced Research Infrastructures: Synergy for a Multidisciplinary Approach to Conservation/Restoration'. Accessed January 28 2016 at http://cordis.europa.eu/project/rcn/92569_en.html.
- Choukri, Khalid, Stelios Piperidis, Prodromos Tsiavos, and John Hendrik Weitzmann. 2011. 'META-SHARE: Licenses, Legal, IPR and Licensing Issues'. D6.1.1. T4ME Net (META - NET). http://www.meta-net.eu/public_documents/t4me/META-NET-D6.1.1-Final.pdf.
- CLARIN ERIC. 2016. 'Common Language and Technology Research Infrastructure (European Research Infrastructure Consortium)'. Accessed January 26. <http://clarin.eu/>.
- CNRS. 2016. 'Centre National de La Recherche Scientifique'. Accessed January 28. <http://www.cnrs.fr/accueil.php>.
- Cockburn, Alistair. 2000. *Writing Effective Use Cases*. Boston, USA: Addison-Wesley Professional.
- Consortium of European Research Libraries. 2006-2007 'The CERL Portal. Manuscripts and Early Printed Material'. Accessed August 8 2016 at <http://cerl.epc.ub.uu.se/sportal/>.

- Consortium of European Research Libraries. 2016. 'Material Evidence in Incunabula'. Accessed August 8. <http://data.cerl.org/mei/search>.
- COST. 2015. 'Annual Report 2014. A Story of Diversity'. Brussels: COST. <http://www.cost.eu/download/AnnualReport2014>.
- . n.d. 'Annual Report 2013'. COST. http://www.cost.eu/download/COST_Annual_Report_2013.
- . 2016. 'European Cooperation in Science and Technology'. Accessed January 26 2016 at <http://www.cost.eu/>.
- Cultura Italia. 2013. 'Cultura Italia. Un Patrimonio da Esplorare'. Accessed August 8 2016 at <http://www.culturaitalia.it/>.
- DANS. n.d. "Data Archiving and Networked Services (DANS)." Data Archiving and Networked Services (DANS). <https://dans.knaw.nl/en>.
- DANS. 2016. 'E-Depot Dutch Archaeology (EDNA)'. Accessed January 27 2016 at <http://www.dans.knaw.nl/en/about/services/archiving-and-reusing-data/easy/edna>.
- DARIAH-DE. 2016. 'Digital Research Infrastructure for the Arts and Humanities - Germany'. Accessed January 26 2016 at <https://de.dariah.eu/>.
- DARIAH EU. 2016. 'Digital Research Infrastructure for the Arts and Humanities'. Accessed January 28 2016 at <http://dariah.eu/>.
- DARIAH-IT. n.d. 'Digital Research Infrastructure for the Arts and Humanities - Italy'.
- DASISH. 2012a. 'Digital Services Infrastructure for Social Sciences and Humanities'. January 25 2016 at <http://dasish.eu/>.
- . 2012b. 'Report about Preservation Service Offers'. Deliverable D4.2. DASISH. [http://dasish.eu/publications/projectreports/D4.2 -
_Report_about_Preservation_Service_Offers.pdf](http://dasish.eu/publications/projectreports/D4.2_-_Report_about_Preservation_Service_Offers.pdf).
- . 2012c. 'Roadmap for Preservation and Curation in the SSH'. Deliverable D4.1. DASISH. [http://dasish.eu/publications/projectreports/D4.1 -
_Roadmap_for_Preservation_and_Curation_in_the_SSH.pdf](http://dasish.eu/publications/projectreports/D4.1_-_Roadmap_for_Preservation_and_Curation_in_the_SSH.pdf).
- . 2014a. 'Course Modules'. Deliverable D7.1. DASISH. <http://dasish.eu/publications/projectreports/DASISH-D7.1-final.pdf>.

- . 2014b. 'Metadata Quality Improvement & Portal Progress Report'. Deliverable D5.2A & D5.2B. DASISH. http://dasish.eu/publications/projectreports/DASISH-D5.2_AB_final_25nov-R.PDF.
- . 2014c. 'Report about Preservation Policy-Rules (Preservation Challenges)'. Deliverable D6.6. DASISH. http://dasish.eu/publications/projectreports/DASISH_D6.6_september_2014.pdf.
- . 2014d. 'List of Recommended Deposit Services for SSH'. D4.3. DASISH. http://dasish.eu/publications/projectreports/DASISH_D4.3_081214-final.pdf.
- . n.d. (2014?). 'DASISH Portal Progress Report'. D5.2B. http://dasish.eu/publications/projectreports/DASISH-D5.2_AB_final_25nov-R.PDF.
- "Data Preservation Alliance for the Social Science." n.d. Data-PASS. <http://www.data-pass.org>.
- Data Seal of Approval. Quality Guidelines for Digital Research Data.* 2009. http://www.data-archive.ac.uk/media/57319/dsa_booklet.pdf.
- Day, Michael. 1997. 'Mapping Dublin Core to UNMARC. Interoperability between metadata formats'. Accessed August 8 2016 at http://www.ukoln.ac.uk/metadata/interoperability/dc_unimarc.html.
- DC. 2016. 'Dublin Core'. *Dublin Core® Metadata Initiative*. Accessed January 27 2016 at <http://www.dublincore.org/>.
- DCC. 2016. 'Digital Curation Centre'. Accessed January 27. <http://www.dcc.ac.uk/>.
- DCH-RP. 2014. *A Roadmap for Preservation of Digital Cultural Heritage Content*. Rome: DCH-RP. <http://www.dch-rp.eu/getFile.php?id=440>.
- . 2016. 'Digital Cultural Heritage Roadmap for Preservation'. Accessed January 26. <http://www.dch-rp.eu/>.
- DELOS. n.d. 'Interoperability Concepts'. https://workinggroups.wiki.dlorg.eu/index.php/Interoperability_Concepts.
- Desipri, Elina, Maria Gavrilidou, Penny Labropoulou, Stelios Piperidis, Francesca Frontini, Monica Monachini, Victoria Arranz, Valérie Mapelli, Gil Francopoulo, and Thierry Declerck. 2012. 'Documentation and User Manual of the META-SHARE Metadata

- Model'. Deliverable D7.2.4. META-NET. <https://goo.gl/DnTvsZ>. http://www.meta-net.eu/public_documents/t4me/META-NET-D7.2.4-Final.pdf.
- Dierickx, Barbara, and Maria Teresa Natale. 2013. 'Report on User Needs and Requirements in Relation to the Creative Applications for the (re)use of Digital Cultural Heritage Content'. D5.1. AthenaPlus. <http://www.athenaplus.eu/getFile.php?id=361>.
- DigCurV. 2016. 'Digital Curator Vocational Education Europe'. Accessed January 27. <http://www.digcur-education.org/>.
- Dillo, Ingrid, Mike Priddy, Linda Reijnhoudt, Ben Companjen, Tim Veken, Kepa Rodriguez, and Yael Gherman. 2015. 'Workshop'. Deliverable D.19.4. EHRI. <https://goo.gl/1hSwBF>.
- "DIN 31644: Criteria for Trustworthy Digital." 2012. <http://www.din.de/en/getting-involved/standards-committees/nid/standards/wdc-beuth:din21:147058907>.
- D'Iorio, Paolo. 2009. 'Final Report'. Deliverable D1.8. Discovery. <https://goo.gl/1SWWmb>. <http://www.discovery-project.eu/reports/discovery-final-report.pdf>.
- Discovery. 2016. 'Philosophy in the Digital Era?' Accessed January 27 2016 at <http://www.discovery-project.eu/home.html>.
- DM. n.d. 'Digital Medievalist'. <https://digitalmedievalist.wordpress.com/>.
- DM2E. 2016. 'Digitised Manuscripts to Europeana'. Accessed January 26 2016 at <http://dm2e.eu/>.
- "DRAMBORA: Digital Repository Audit Method Based on Risk Assessment." 2007. <http://www.repositoryaudit.eu/>.
- DYAS (DARIAH GR). 2016. 'Greek Research Infrastructure Network for the Humanities'. Accessed January 28 2016 at <http://www.dyas-net.gr/?lang=en>.
- E.C.C.O. 2013. 'European Recommendation for the Conservation and Restoration of Cultural Heritage'. E.C.C.O. <https://goo.gl/DT4NpL>. <http://www.ecco-eu.org/documents/ecco-documentation/european-recommendation-for-the-conservation-and-restoration-of-cultural-heritage/download.html>.
- . 2016. 'European Confederation of Conservator-Restorers' Organisations'. Accessed January 27 2016 at <http://www.ecco-eu.org/>.

- ECLAP. 2016. 'European Collected Library of Artistic Performance'. Accessed January 26 2016 at <http://www.eclap.eu/>.
- EHRI. 2016. 'European Holocaust Research Infrastructure'. Accessed January 27 2016 at <http://ehri-project.eu/>.
- Elisa Sciotto, Luca Martinelli, Patrizia Martini, and Sara di Giorgio. 2014. 'Report on Dissemination Activities'. Deliverable 2.3.2. DCH-RP. <http://www.dch-rp.eu/getFile.php?id=445>.
- Engelhardt, Claudia, Stefan Strathmann, and Katie McCadden. n.d. 'Report and Analysis of the Survey of Training Needs'. DigCurV. <https://goo.gl/ANpolt>.
<http://www.digcur-educa-tion.org/index.php/eng/content/download/3322/45927/file/Report%20and%20analysis%20of%20the%20survey%20of%20Training%20Needs.pdf>.
- Escuela Técnica Superior de Arquitectura de la Universidad Politécnica de Madrid. 2012. 'New Structure of Standardisation Advance of CEN/TC 346 - Conservation of Cultural Heritage'. *Recopar*, December.
<http://polired.upm.es/index.php/recopar/article/view/2226/2308>.
- ESFRI. 2008. Social Sciences and Humanities Roadmap Working Group Report. 2008. Accessed July 22 2016 at http://ec.europa.eu/research/infrastructures/pdf/esfri/esfri_roadmap/roadmap_2008/ssh_report_2008_en.pdf.
- ESFRI. 2016. 'European Strategy Forum on Research Infrastructures'. Accessed January 27 2016 at http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri.
- EUDAT. 2016. 'Research Data Services, Expertise & Technology Solutions'. Accessed January 26 2016 at <http://www.eudat.eu/>.
- "European Framework for Audit and Certification of Digital Repositories." 2010.
<http://www.trusteddigitalrepository.eu/Welcome.html>.
- Europeana Cloud. 2014. 'Unlocking Europe's Research via The Cloud'. September 24 2016 at <http://pro.europeana.eu/structure/europeana-cloud>.
- Fallon, Julia, Pavel Kats, Alastair Dunning, and Marcin Werla. 2015. 'Product & Services Requirements for Implementing Europeana Cloud Services'. Deliverable D5.7. Eu-

ropeana Cloud.

http://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/Europeana_Cloud/Deliverables/D5.7%20Product%20and%20Service%20Requirements.pdf.

Fassina, Vasco. 2008. 'European Technical Committee 346 - Conservation of Cultural Property - Updating of the Activity After a Three Year Period'. In *Proceedings of the 9th International Conference on NDT of Art*, 9. Jerusalem, Israel: NDT.net.
<http://www.ndt.net/article/art2008/papers/174Fassina.pdf>.

———. 2012. 'European Technical Committee - CEN TC 346 - Conservation of Cultural Heritage'. presented at the European Heritage Heads Forum. Public Engagment with Cultural Heritage, Potsdam/Berlin, May 23 2016 at
http://ehhf.eu/sites/default/files/201407/Session_6_Fassina.pdf.

Feijen, Martin. 2011. *What Researchers Want. A Literature Study of Researchers' Requirements with Respect to Storage and Access to Research Data*. Edited by Paul Gretton and Keith Russell. Stichting: SURF.
http://www.surf.nl/binaries/content/assets/surf/en/knowledgebase/2011/What_researchers_want.pdf.

Fernie, Kate. 2014. 'Report on Data Sharing Policies'. Deliverable D3.3. ARIADNE.
http://www.ariadne-infrastructu-re.eu/content/download/2106/11888/version/2/file/D3.3+Report+on+data+sharing+policies_final.pdf.

FLaReNet. 2016. 'Fostering Language Resources Network'. Accessed January 28 2016 at
<http://www.flarenet.eu/>.

Fleming, Arlene K. 2008. 'Standards of International Cultural and Financial Institutions for Cultural Heritage Protection and Management'. In *'IAIA08 Conference Proceedings', The Art and Science of Impact Assessment*. Perth, Australia: International association for impact assessment.
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.501.9482&rep=rep1&type=pdf>.

Fowler, Martin. 2004. *UML Distilled (Third Edition)*. Addison-Wesley Professional.

- Geser, Guntram, and Hannes Selhofer. 2015. "Preliminary Innovation Agenda and Action Plan." ARIADNE D2.3. SRFG. <http://www.ariadne-infrastructure.eu/Resources/D2.3-Preliminary-Innovation-Agenda-and-Action-Plan>.
- Gnadt, Timo, and Claudia Engelhardt. n.d. 'DASISH - Data Service Infrastructure for the Social Sciences and Humanities'. 7.1. Goettingen.
- Gordon McKenna, Chris De Loof, and Chris De Loof. 2009. 'Digitisation: Standards Landscape for European Museums, Archives, Libraries'. ATHENA. <https://goo.gl/ehbZ8K>. <http://www.athenaeurope.org/getFile.php?id=435>.
- 'Handbook on Legal and Ethical Issues for SSH Data in Europe, Part II'. 2014. D6.5. DASISH. http://dasish.eu/publications/projectreports/DASISH_D6.5_februar_2015.pdf.
- Hansen, Dorte Haltrup, Lene Offersgaard, and Sussi Olsen. 2014. 'Using TEI, CMDI and ISocat in CLARIN-DK'. In *Proceedings of LREC 2015, the Ninth International Conference on Language Resources and Evaluation*, 613–18. Reykjavik, Island. http://www.lrec-conf.org/proceedings/lrec2014/pdf/325_Paper.pdf.
- Harnad, Stevan. 2005. 'OA Impact Advantage = EA + (AA) + (QB) + QA + (CA) + UA Open Access Archivangelism'. *Eprints*. September 17. <http://eprints.soton.ac.uk/262085/1/OAA.html>.
- Hart, Penny. 2015. 'Recognising Influences on Attitudes to Knowledge Sharing in a Research Establishment:: An Interpretivist Investigation'. *International Journal of Systems and Society* 2 (2): 68–87. doi:10.4018/IJSS.2015070105.
- Hennicke, Steffen, Stefan Gradmann, Kristin Dill, Gerold Tschumpel, Klaus Thoden, Christian Morbindoni, and Alois Pichler. 2015. 'Research Report on DH Scholarly Primitives'. Deliverable D3.4. DM2E. <https://goo.gl/bp39Bq>. http://dm2e.eu/files/D3.4_2.0_Research_Report_on_DH_Scholarly_Primitives_150402.pdf.
- Henriksen, Lina, Dorte H. Hansen, Bente Maegaard, Bolette S. Pedersen, and Claus Povlsen. 2014. 'Encompassing a Spectrum of LT Users in the CLARIN-DK Infrastructure'. In *Proceedings of LREC 2014, Ninth International Conference on Language Resources and Evaluation*. Reykjavik, Iceland. http://www.lrec-conf.org/proceedings/lrec2014/pdf/814_Paper.pdf.
- HHS. n.d. 'U.S. Dept. of Health and Human Services'. <http://www.usability.gov/>.

- . n.d. 'Use Cases'. <http://www.usability.gov/how-to-and-tools/methods/use-cases.html>.
- Höckendorff, Mareike, Stefan Pernes, and Marcus Held. 2015. 'Konzept Dissemination Und Lehrmittelsammlung Cluster 5 - Big Data in Den Geisteswissenschaften'. R 5.4.1. DARIAH-DE. <https://dev2.dariah.eu/wiki/pages/viewpage.action?pageId=26150061#Meilensteine,Reports,Arbeitspl%C3%A4ne-Cluster5>.
- Hollander, Hella, and Maarten Hoogerwerf. 2014. 'Service Design'. Deliverable D13.1. Den Haag, Netherlands: ARIADNE. http://www.ariadne-infrastructu-re.eu/content/download/4974/29046/version/2/file/D13.1_Ariadne_Service_Design.pdf.
- Hogenaar Arjan, Tjalsma Heiko, Priddy Mike. 2011. 'OpenAIRE Research in the Humanities and Social Sciences'. Accessed July 22 2016 at <http://pub.unibielefeld.de/publication/2458719>
- Hoogerwerf, Maarten. 2009. 'Durable Enhanced Publications'. In *Proceedings of African Digital Scholarship and Curation 2009*. Pretoria, South Africa. http://www.ais.up.ac.za/digi/docs/hoogerwerf_paper.pdf.
- Horsman, Peter, Petra Links, Karsten Kühnel, Mike Priddy, Linda Reijnhoudt, and Markus Merenmies. 2013. 'Guidelines for Description'. Deliverable D.17.3. EHRI. <https://goo.gl/ZNkEwB>.
- Horsman, Peter, Petra Links, Karsten Kühnel, Mike Priddy, Linda Reijnhoudt, Laurents Sesink, and Markus Merenmies. 2012. 'Metadata Schema for the Portal Site and Report on Terminology'. Deliverable D17.2. EHRI. http://data.d4science.org/uri-resolver/id?fileName=EHRI_D17.2_Metadata_schema_for_the_portal_site.pdf&smp-id=55e6cdcf4b0a7a3b54f1468&contentType=application%2Fpdf.
- Huma-Num. 2015. 'La TGIR Des Humanités Numériques'. Web page. March 24. <http://www.huma-num.fr/>.

- Hunter, Jane. 2008. 'Scientific Publication Packages – A Selective Approach to the Communication and Archival of Scientific Output'. *International Journal of Digital Curation* 1 (1): 33–52. doi:10.2218/ijdc.v1i1.4.
- ICCU. 2010. 'Istituto Centrale per il Catalogo Unico'. Accessed August 8 2016 at <http://www.iccu.sbn.it/opencms/opencms/it/>.
- IFLA Working Group on Functional Requirements and Numbering of Authority Records (FRANAR). 2013. Functional Requirements for Authority Data A Conceptual Model (2013). Accessed August 8 2016 at http://www.ifla.org/files/assets/cataloguing/frad/frad_2013.pdf.
- INDIGO DataCloud. 2016. 'Towards a Sustainable European PaaS-Based Cloud Solution for E-Science'. Accessed January 28 2016 at <https://www.indigo-datacloud.eu/>.
- International Federation of Library Associations and Institutions. 2011. International Standard Bibliographic Description. Consolidated Edition. Accessed August 8 2016 at http://www.ifla.org/files/assets/cataloguing/isbd/isbd-cons_20110321.pdf.
- IPERION CH. 2016. 'Integrated Platform for the European Research Infrastructure ON Culture Heritage'. Accessed January 28 2016 at <http://www.iperionch.eu/>.
- ISCH COST Action IS1005. 2016. 'Medieval Europe - Medieval Cultures and Technological Resources (Medioevo Europeo)'. Accessed January 28 2016 at http://www.cost.eu/COST_Actions/isch/IS1005.
- ISIDORE. 2011. 'Portal for Digital Humanities by French National Research Center'. April 4 2016 at <http://www.antidot.net/en/our-customers/public-services/isidore/>.
- ISO. 2009. "ISO Data Category Registry." *ISO 12620:2009*.
http://www.iso.org/iso/catalogue_detail.htm?csnumber=37243.
- ISO. 2012. "Audit and Certification of Trustworthy Digital Repositories." *ISO 16363:2012*.
http://www.iso.org/iso/catalogue_detail.htm?csnumber=56510.
- Jankowski, Nicholas Warren, Clifford Tatum, Zuotian Tatum, and Andrea Scharnhorst. 2011. 'Enhancing Scholarly Publishing in the Humanities and Social Sciences: Innovation through Hybrid Forms of Publication'. In *Proceedings of PKS Scholarly Publishing Conference*.
http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1929687.

- Jisc. 2016. 'Joint Information Systems Committee'. Accessed January 26 2016 at <https://www.jisc.ac.uk/>.
- Kate Fernie. 2013. 'Initial Dissemination Plan'. D4.2. ARIADNE. http://www.ariadne-infrastruc-tu-re.eu/ita/content/download/4019/23217/file/ARIADNE_D4.2_Initial_dissemination_plan.pdf.
- Kathleen Smith, Sibylle Söring, Ubbo Veenster, and Felix Lohmeier. 2012. 'The TextGrid Repository: Supporting the Data Curation Needs of Humanities Researchers.' In *Digital Diversity: Cultures, Languages and Methods (Proceedings of DH 2012, Digital Humanities Conference)*. Hamburg, Germany. http://www.dh2012.uni-hamburg.de/wp-content/uploads/2012/08/TextGrid_DH2012_Poster.pdf.
- Kenworthy, Edward. 1997. *Use Case Modelling – Capturing User Requirements*. available online (last visited on 06/01/2016): http://www.zoo.co.uk/~z0001039/PracGuides/pg_use_cases.htm.
- Kingston, Jeff. 2015. 'Japanese University Humanities and Social Sciences Programs Under Attack'. *The Asia-Pacific Journal* 13 (September). <http://japanfocus.org/-Jeff-Kingston/4381/article.html>.
- Krauwier, Steven. 2014. 'The Knowledge Sharing Infrastructure [KSI]'. CE - 2013 - 0149. CLARIN. <https://www.clarin.eu/sites/default/files/CE-2013-0149-KSI-v8.0.pdf>.
- Kuipers, Tom, and Jeffrey van der Hoeven. 2009. 'Insight into Digital Preservation of Research Output in Europe'. D3.4. PARSE.Insight. <http://www.parse-insight.eu/wp-content/plugins/download-monitor/download.php?id=21>.
- L'Hours, Hervé, Lene Offersgaard, Marion Wittenberg, Bartholomäus Wloka, and Mike Priddy. n.d. 'DASISH Metadata Quality Improvement'. D5.2A. http://dasish.eu/publications/projectreports/DASISH-D5.2_AB_final_25nov-R.PDF.
- Library of Congress. 2015. 'Metadata Object Description Schema: MODS'. <http://www.loc.gov/standards/mods/>.
- Links, Petra, Sheila Anderson, Reto Speck, and Agiatis Benardou. 2012. 'Overview of Use at Partner Sites'. D.16.3. EHRI. (*not available on ehri website*).

- Links, Petra, Peter Horsman, Karsten Kühnel, Mike Priddy, Linda Reijnhoudt, Laurents Sesink, and Markus Merenmies. 2012. 'Report on Standards Including Survey of Existing Approaches'. Deliverable D17.1. EHRI. (*not available on ehri website*).
- Luyten, Dirk, and Hans Boers. 2013. 'Digital Handbook on Privacy and Access'. Deliverable D.3.2. EHRI. <https://goo.gl/5kyJnV>. http://www.ehri-project.eu/webfm_send/471.
- Maarten Zeinstra, Lisette Kalshoven, Alastair Dunning, Julia Fallon, and Louise Edwards. 2013. 'Minimum Requirements for Europeana Cloud'. D5.1. Europeana Cloud. http://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/Europeana_Cloud/Deliverables/D5.1%20Minimum%20requirements%20for%20the%20cloud.pdf.
- Madeleine Gray, Hilary Hanahoe, and Adam Carter. 2015. 'Annual Dissemination and Outreach Report 3'. Deliverable D3.3.3. EUDAT. <http://hdl.handle.net/11304/09532e76-f406-11e4-ac7e-860aa0063d1f>.
- Maegaard, Bente, Hanne Fersøe, and Lina Henriksen. 2009. 'Requirements and Best Practice for Transnational Coordination and Collaboration with Third Parties'. D8S-3.1. CLARIN. <http://www-sk.let.uu.nl/u/D8S-3.1.pdf>.
- Marras, Cristina, and Giovanni De Grandis. 2014. 'Good Practice Report'. Deliverable D1.9. AGORA. <https://goo.gl/oZNP40>.
- Mathias Göbel, Nadja Grupe, Christian Heise, Maren Köhlmann, Katharina Meyer, Markus Neuschäfer, Stefan Schmunk, and Sibylle Söring. 2014. 'DARIAH-DE und TextGrid. Disseminationsstrategie inklusive Marketingkonzept sowie DARIAH-DE Open Mission Statement und Publikationsstrategie'. Report 7.2 / 7.3.3. DARIAH-DE 2 / TextGrid 3. <https://textgrid.de/fileadmin/TextGrid/reports/DARIAH-TextGrid-Disseminationskonzept.pdf>.
- McCrae, John, Jorge Gracia, Roberto Navigli, Paul Buitelaar, Philipp Cimiano, Luca Matteis, Victor Rodriguez Donce, et al. 2015. 'Reconciling Heterogeneous Descriptions of Language Resources'. In *Proceedings of LDL-2015, the 4th Workshop on Linked Data in Linguistics*, 39–48. Beijing, China. <http://www.aclweb.org/anthology/W/W15/W15-4205.pdf>.
- MESO DARIAH WG. n.d. 'Medievalist Sources (DARIAH Working Group)'. <http://www.medievalistsources.eu>.

- META-NET. 2016. 'META-NET - META Multilingual Europe Technology Alliance'. Accessed January 27 2016 at <http://www.meta-net.eu/>.
- META-SHARE. 2016. 'META-SHARE - a Project of META-NET'. Accessed January 28 2016 at <http://www.meta-share.eu/>.
- Minerva Project. 2009. 'MINERVA Technical Guidelines for Digital Cultural Content Creation Programmes: Version 2.0, 2008'. Accessed on August 8 2016 at <http://www.minervaeurope.org/interoperability/technicalguidelines.htm>.
- Minerva Working Group. 2008. 'Intellectual Property Guidelines'. 1.0. MINERVA. http://www.minervaeurope.org/publications/MINERVAeC%20IPR%20Guide_final1.pdf.
- Monachini, Monica, Valeria Quochi, Nicoletta Calzolari, Núria Bel, Gerhard Budin, P. Caselli, Khalid Choukri, et al. 2011. *The Standards' Landscape Towards an Interoperability Framework. The FLaReNet Proposal: Building on the CLARIN Standardisation Action Plan*. Pisa, Italy: CNR. <http://dspace.library.uu.nl/handle/1874/285299>.
- Moore, Kristen R., and Timothy J. Elliott. 2015. 'From Participatory Design to a Listening Infrastructure. A Case of Urban Planning and Participation'. *Journal of Business and Technical Communication*, September. doi:10.1177/1050651915602294.
- Moti, Nissani. 1995. 'Fruits, Salads, and Smoothies: A Working Definition of Interdisciplinarity'. *Journal of Educational Thought* 29 (2): 119–26.
- Moyle, Martin, Marnix van Berchum, and Friedel Grant. 2013. 'Stakeholder Engagement Plan'. Deliverable D6.1. Europeana Cloud. http://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/Europeana_Cloud/Deliverables/D6.1%20Stakeholder%20Engagement%20&%20Infrastructure%20Plan.pdf.
- National Information Standards Organization (NISO). 2007. *A Framework of Guidance for Building Good Digital Collections*. 3rd ed. A NISO Recommended Practice. Baltimore, MD: National Information Standards Organization (NISO). <http://www.niso.org/publications/rp/framework3.pdf>.
- NeDiMAH. 2016. 'Network for Digital Methods in the Arts and Humanities'. Accessed January 28 2016 at <http://www.nedimah.eu/>.

- (N)ERD. 2016. '(Named) Entity Recognition and Disambiguation'. Accessed August 8 2016 at <http://http://cloud.science-miner.com/nerd/>.
- "Nestor Criteria. Catalogue of Criteria for Trusted Digital Repositories. Version 2." 2009. nestor. http://files.dnb.de/nestor/materialien/nestor_mat_08_eng.pdf.
- Neuman, Yrsa, Hugo Strandberg, Martin Gustafsson, and Alois Pichler. 2014. 'Report on Future Strategy of Open Access'. D7.5. AGORA. <http://www.project-agera.org/wp-content/uploads/2012/09/D7.5-Report-D7.5-Open-Access-Business-Models-for-Publishers-FINAL.pdf>.
- NISO. 2016. 'National Information Standards Organization'. Accessed January 26 2016 at <http://www.niso.org/home/>.
- OAPEN Library. n.d. <http://www.oapen.org>.
- O'Brien, Catherine. 2012. 'Dissemination Strategy'. Deliverable D2.2. CENDARI. <https://goo.gl/s76R3E>. http://www.cendari.eu/sites/default/files/CENDARI_D2.2-Communications-Strategy-final.pdf.
- O'Connor, Alexander, Natasa Bulatovic, and Owen Conlan. 2014. 'Position Paper on Data Integration Architecture (CENDARI)'. CENDARI. (*we could not find it online*).
- OER (DARIAH). n.d. 'Open Educational Resources'. <https://www.oercommons.org/groups/dariah/229/>.
- Offersgaard, Lene, and Dorte Haltrup Hansen. 2014. 'Reusing CMDI Components for a textCorpusProfile - towards a Generic textCorpusProfile'. In . Soesterberg, The Netherlands. http://www.clarin.eu/sites/default/files/cac2014_submission_18_0.pdf.
- Palmer, L. Carole, Nicholas M. Weber, and Melissa H. Cragin. 2011. "The Analytic Potential of Scientific Data: Understanding Re-Use Value." *Proceedings of ASIST 2011, Annual Meeting, New Orleans*. http://www.asis.org/asist2011/proceedings/submissions/174_FINAL_SUBMISSION.pdf.
- Pampel, Heinz, Hans Pfeiffenberger, Angela Schäfer, Eefke Smit, Stefan Pröll, and Christoph Bruch. 2012. 'Report on Peer Review of Research Data in Scholarly Communication'. D3.3. APARSEN. Accessed April 30 2016 at

<http://www.alliancepermanentaccess.org/wp-content/plugins/download-monitor/download.php?id=83>

- Papatheodorou, Christos, Dimitris Gavriliis, Kate Fernie, Holly Wright, Julian Richards, Paola Ronzino, and Carlo Meghini. 2013. 'Initial Report on Standards and on the Project Registry'. Deliverable D3.1. ARIADNE. <http://www.ariadne-infrastruc-tu-re.eu/content/download/1781/9956/version/3/file/D3.1+Initial+report+on+standards+and+on+the+project+registry.pdf>.
- Park, Jung-Ran. 2009. 'Metadata Quality in Digital Repositories: A Survey of the Current State of the Art'. *Cataloging & Classification Quarterly* 47 (3-4): 213–28. doi:10.1080/01639370902737240.
- PARSE.Insight. 2016. 'Permanent Access to the Records of Science in Europe'. Accessed January 27 2016 at <http://www.parse-insight.eu/>.
- PARTHENOS. 2015. 'Grant Agreement. Number - 654119 - PARTHENOS'.
- . 2016. 'PARTHENOS Entities - Categorical Description - V1.12'. (*at this point an internal working document*)
- . 2016. 'Pooling Activities, Resources and Tools for Heritage E-Research Networking, Optimization and Synergies'. Accessed January 28 2016 at <http://www.parthenos-project.eu>.
- PCDK project team. 2012. 'Guidelines on Cultural Heritage. Technical Tools for Heritage Conservation and Management'. Council of Europe. <https://www.coe.int/t/dg4/cultureheritage/cooperation/Kosovo/Publications/Guidelines-ENG.pdf>.
- PERICLES. 2016. 'Promoting and Enhancing Reuse of Information throughout the Content Lifecycle Taking Account of Evolving Semantics'. Accessed January 28. <http://pericles-project.eu/>.
- Puhl, Johanna, Peter Andorfer, Mareike Höckendorff, Stefan Schmunk, Juliane Stiller, and Klaus Thoden. 2015. 'Diskussion und Definition eines Research Data LifeCycle für die digitalen Geisteswissenschaften'. 11. DARIAH-DE Working Papers. Göttingen: DARIAH-DE. <http://nbn-resolving.de/urn:nbn:de:gbv:7-dariah-2015-4-4>.

- Quochi, Valeria, Lothar Lemnitzer, and Marc Kemp-Snijders. 2009. 'Usage and Workflow Scenarios'. Deliverable D5R-2. CLARIN.
<https://services.d4science.org/workspace-6.10.1-3.10.0/workspace/DownloadService?id=52aba59c-9063-41f2-9336-aceb65cf1ea8&viewContent=true&redirectonerror=true>.
<http://hdl.handle.net/1839/00-DOCS.CLARIN.EU-54>.
- Rehm, Georg, and Hans Uszkoreit. 2013. *META-NET Strategic Research Agenda for Multilingual Europe 2020*. Springer Publishing Company, Incorporated.
<http://dl.acm.org/citation.cfm?id=2462570>.
- Rehm, Georg, Hans Uszkoreit, and META Technology Council, eds. 2013. *Strategic Research Agenda for Multilingual Europe 2020*. White Paper Series. Heidelberg: Springer.
- Reijnhoudt, Linda, Ben Companjen, Mike Priddy, Tim Veken, Mike Bryant, Kepa Rodriguez, and Yael Gherman. 2015. 'Filled Metadata Registry'. Deliverable D.19.5. EHRI. <https://goo.gl/bkCzOY>.
- RIN. 2008. 'To Share or Not to Share: Publication and Quality Assurance of Research Data Outputs'. RIN. <http://www.rin.ac.uk/system/files/attachments/To-share-data-outputs-report.pdf>.
- . 2016. 'Research Information Network'. Accessed January 27 2016 at <http://www.rin.ac.uk/>.
- Ronzino, Paola, Kate Fernie, Christos Papatheodorou, Holly Wright, and Julian Richards. 2013. 'Report on Project Standards'. Deliverable D3.2. ARIADNE.
<http://www.ariadne-infrastructure.eu/content/download/1782/9961/version/2/file/D3.2+Report+on+project+standards.pdf>.
- Rouchon, Olivier, Philippe Prat, and Marc Batllo. 2011. "Guide Méthodologique." ADO-NIS/SIAF/CINES-GM-0.5. <http://www.huma-num.fr/sites/default/files/guide-formats-numeriques.pdf>.
- Sahle, Patrick, ed. 2011. *Digitale Geisteswissenschaften*. Digital Humanities Curriculum. Cologne: Cologne Center for eHumanities. <http://www.cceh.uni-koeln.de/Dokumente/BroschuereWeb.pdf>.

- . 2013. 'DH studieren! Auf dem Weg zu einem Kern- und Referenzcurriculum'. Milestone 2.3.3. DARIAH-DE. <https://goo.gl/DImOCU>. http://cceh.uni-koeln.de/files/DARIAH-M2-3-3_DH-programs_1_2_0.pdf.
- Sam Leon, and Violeta Trkulja. 2013. 'Dissemination and Engagement Plan'. Deliverable D4.4. DM2E. http://dm2e.eu/files/D4.4_1.1_WP4_Dissemination_and_Engagement_Plan_131029.pdf.
- Schmidutz, Daniel, Ryan, Lorna, Müller Gjesdal, Anje, De Smedt, Koenraad. 2013 'DASISH Report about new IPR Challenges'. Accessed July 22 2016 at http://dasish.eu/publications/projectreports/D6.1_final.pdf.
- Schreibman, Susan, Raymond George Siemens, and John Unsworth, eds. 2004. *A Companion to Digital Humanities*. Blackwell Companions to Literature and Culture 26. Malden, MA: Blackwell Pub. <http://www.digitalHumanities.org/companion/>.
- Scuola Normale Superiore di Pisa. 2005. 'Linee Guida per Lo Sviluppo Di Sistemi Informatici Interoperabili Con CulturalItalia'. Scuola Normale Superiore. <http://www.culturalitalia.it/opencms/export/sites/culturalitalia/attachments/documenti/lineeguida/LineeguidaintegrazioneCulturalItalia.pdf>.
- Selhofer, Hannes, and Guntram Geser. 2014. 'First Report on Users' Needs'. Deliverable ARIADNE D2.1. ARIADNE. <https://goo.gl/wAwjUZ>. http://www.ariadne-infrastructure.eu/content/download/2870/16435/version/2/file/ARIADNE_D2-1+First+report+on+users+needs.pdf.
- . 2015. 'Second Report on User Needs'. Deliverable D2.2. ARIADNE. <http://www.ariadne-infrastructure.eu/content/download/5616/32917/version/3/file/D2.2+Second+report+on+users+needs.pdf>.
- Simukovic, Elena. 2012. 'Enhanced publications—Integration von Forschungsdaten Beim Wissenschaftlichen Publizieren'. Berlin, Germany: Humboldt- Universität zu Berlin. http://www.researchgate.net/profile/Elena_Simukovic/publication/257651610_Enhanced_publications_Integration_von_Forschungsdaten_beim_wissenschaftlichen_Publizieren/links/00b7d52595ff5c712e000000.pdf.
- Smith, Kathleen, and Nadia Boukhelifa. 2012. 'Participatory Design Workshop Report 1: WWI Researchers'. CENDARI. (*we could not find this in the WWW.*)

- SSH RWG. 2008. 'Social Sciences and Humanities Roadmap Working Group Report 2008'. ESFRI.
http://ec.europa.eu/research/infrastructures/pdf/esfri/esfri_roadmap/roadmap_2008/ssh_report_2008_en.pdf.
- Soria, Claudia, Núria Bel, Khalid Choukri, Joseph Mariani, Monica Monachini, Jan Odijk, Stelios Piperidis, Valeria Quochi, Nicoletta Calzolari, and others. 2012. 'The FLReNet Strategic Language Resource Agenda'. In *Proceedings of LREC 2012, Eighth International Conference on Language Resources and Evaluation*, 1379–86. Istanbul, Turkey. http://lrec.elra.info/proceedings/lrec2012/pdf/777_Paper.pdf.
- Speck, Reto. 2011. 'Stakeholder Report'. D.16.2. EHRI. (*we could not find this on the WWW*).
- T4ME (META-NET). 2016. 'Technologies for the Multilingual European Information Society (A Network for Excellence Forging the Multilingual Europe Technology Alliance)'. Accessed January 26 2016 at <http://www.meta-net.eu/projects/t4me/>.
- Tellegen, Jan Willem. 2011. 'Publicity & Dissemination Strategy, Concept for Identity, Branding & Graphic Design'. Deliverable D8.1. EHRI. <https://goo.gl/qQUC1m>.
- TextGrid. 2016. 'Virtuelle Forschungsumgebung Für Die Geisteswissenschaften'. Accessed January 26 2016 at <https://textgrid.de/>.
- "Trustworthy Repositories Audit & Certification: Criteria and Checklist." 2007. OCLC & CRL. https://www.crl.edu/sites/default/files/d6/attachments/pages/trac_0.pdf.
- Tsolis, Dimitrios. 2013. 'IPR Guidebook'. Europeana Photography D6.2.
<http://www.europeana-photography.eu/getFile.php?id=298>.
- UK Data Archive. 2016. 'Research Data Lifecycle'. Accessed July 22 2016 at <http://www.data-archive.ac.uk/create-manage/lifecycle>.
- University of Oxford. 2015. 'Medieval Libraries of Great Britain'. Accessed August 8.
<http://mlgb3.bodleian.ox.ac.uk/>.
- University of Oxford. 2016. 'TEXT-inc. A corpus of texts printed in the 15th century'. Accessed August 8. <http://textinc.bodleian.ox.ac.uk/>.
- Van den Eynden, Veerle, Louise Corti, Matthew Woollard, Libby Bishop, and Laurence Horton. 2009. *Managing and Sharing Data*. Third edition. Wivenhoe Park Colchester Essex CO4 3SQ: UK Data Archive University of Essex.

<https://uk.sagepub.com/en-gb/eur/managing-and-sharing-research-data/book240297>.

- Van Uytvanck, Dieter, Herman Stehouwer, and Lari Lampen. 2012. 'Semantic Metadata Mapping in Practice: The Virtual Language Observatory'. In *Proceedings of LREC 2012, Eighth International Conference on Language Resources and Evaluation*. Istanbul, Turkey.
http://pubman.mpg.de/pubman/item/escidoc:1454694:11/component/escidoc:1478393/VanUytvanck_LREC_2012.pdf.
- Váradi, Tamás, and Piroska Lendvai. 2011. 'Integrated Strategic Plan for Supporting HSS Research'. Deliverable 3C-6.1. CLARIN. <https://www.clarin.eu/content/reports>.
<http://hdl.handle.net/1839/00-DOCS.CLARIN.EU-48>.
- WDL. 2016. 'World Digital Library'. Accessed January 26. <https://www.wdl.org/en/>.
- WDL Content Selection Committee. 2015. 'World Digital Library - Content Workflow'. WDL. Accessed July 12. <http://project.wdl.org/content/>.
- Webb, Sharon and Charlene McGoohan. 2015. Digital Repository of Ireland: Requirements (DRI). National University of Ireland Maynooth. DOI: <http://dx.doi.org/10.3318/DRI.2015.6> Accessed July 22 2016 at <http://dri.ie/sites/default/files/files/dri-requirements-specification.pdf>.
- Weber, Nicholas M., Karen S. Baker, Andrea K. Thomer, Tiffany C. Chao, and Carole L. Palmer. 2012. "Value and Context in Data Use. Domain Analysis Revisited." *Proceedings of ASIS&T 2012, Baltimore, 12-30 October 2012 at* <https://www.asis.org/asist2012/proceedings/Submissions/168.pdf>.
- Wright, Holly, Julian Richards, Kate Fernie, Hannes Selhofer, Franco Niccolucci, Carlo Meghini, Paola Ronzino, Christos Papatheodorou, and Hella Hollander. 2014. 'Use Requirements'. Deliverable D12.1. Florence, Italy: ARIADNE. <http://www.ariadne-infrastructure.eu/Resources/D12.1-Use-Requirements-EU-Reporting>.
- Wright, Sue Ellen, Menzo Windhouwer, Ineke Schuurman, and Daan Broeder. 2014. 'Segueing from a Data Category Registry to a Data Concept Registry'. In *Proceedings of Terminology and Knowledge Engineering 2014*. Berlin, Germany.
<https://hal.archives-ouvertes.fr/hal-01005840/document>.