

A Computational History Approach to Interpretation and Analysis of Moral European Values: the VAST Research Project

Silvana Castano¹, Alfio Ferrara¹, Giulia Giannini², Stefano Montanelli¹ and Francesco Periti¹

¹Università degli Studi di Milano
Department of Computer Science
Via Celoria, 18 - 20133 Milano, Italy

²Università degli Studi di Milano
Department of Historical Studies
Via Festa del Perdono, 7 - 20122 Milano, Italy

Abstract

The rapid development and diffusion of artificial intelligence and data science techniques enable research in the field of humanities and social sciences to become more and more computational. In this paper, we present the computational history approach featuring the VAST project aiming to study the transformation of moral values across space and time, by collecting, digitising, and analysing narratives and experiences, with particular emphasis on the core European Values, such as freedom, democracy, equality.

Keywords

Computational History, Knowledge Modeling, Ontology-based data management

1. Introduction


In the field of the Humanities, Social Sciences, and Cultural Heritage, demand for semantic text-analysis techniques as well as for intelligent discovery, linking, querying, and visualization of massive volumes of data is increasing [1]. On the one side, this leads to large-scale digitization projects and tools for producing, curating, and exploiting humanities and social-science data, by relying on text analysis techniques and new hybrid methodologies derived from the intersection/intertwining of the involved research communities [2]. On the other side, the need to keep humans “in-the-loop” leads to crowdsourcing solutions, to help in large-scale, human-intensive processes such as text tagging, commenting, rating, and reviewing, as well as in the creation and upload of content in a methodical, task-based fashion. Furthermore, the rapid development and diffusion of artificial intelligence techniques and data science approaches enable research in the field of humanities and social sciences to become more and more computational. For example, various studies are devoted to exploit artificial intelligence techniques to analyze

HistoInformatics 2021 – 6th International Workshop on Computational History, September 30, 2021, online

✉ silvana.castano@unimi.it (S. Castano); alfio.ferrara@unimi.it (A. Ferrara); giulia.giannini1@unimi.it (G. Giannini); stefano.montanelli@unimi.it (S. Montanelli); francesco.periti@unimi.it (F. Periti)



© 2021 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

literary texts, historic productions, or public opinions about political events for knowledge extraction and/or classification/analytics purposes. As a consequence, we are assisting to a shift from *digital humanities* to the so-called *computational humanities* research, where the role of artificial intelligence, data science and cutting edge digital technologies is fundamental to achieve research advances and results [3].

In this paper, we present VAST (*Values Across Space and Time*), an EU H2020 project started in December 2020 providing a concrete example of computational humanities and, more specifically, computational history research. Going beyond analysing the transformation of moral values in the past, the VAST project will study how moral values are communicated and perceived today, by collecting, digitising, and analysing narratives and experiences of both communicators of moral values, like for example artists, directors, culture and creative industry institutions, museum curators, storytellers, educators, and the respective audiences, like spectators, museum visitors, students, and pupils. Within VAST, an interdisciplinary consortium of scholars from both humanities and computer science has been created. VAST aims to study the transformation of moral values across space and time, with particular emphasis on the core European Values, such as freedom, democracy, equality, rule of law, tolerance, dialogue, and dignity, that are widely recognized as the essential pillars to constitute a society in which inclusion, tolerance, justice, solidarity, and non-discrimination prevail [4].

The project envisions to bring European Values to the forefront by using cutting edge technologies to create a digital platform and a knowledge base by including narratives from three main areas: i) *theater*, focusing on ancient Greek Drama, ii) *science*, focusing on Scientific Revolution and natural-philosophy documents of the 17th century, and iii) *folklore*, focusing on folktales and fairy-tales. Through advanced techniques and digital tools, the research teams in the project are going to study how the meaning of specific values has been expressed, transformed, and appropriated throughout time, going back to the stories that helped to shape part of the European culture. VAST will examine narratives and user experiences that represent significant moments of European culture and history such as the classical period, and the Scientific Revolution of the 17th century, when the conceptual, methodological and institutional foundations of modern science were first established, to the modern era. In this respect, the main VAST issues about knowledge modeling and architectural design are discussed in the paper both from methodological and technical point of view.

The paper is organized as follows. Section 2 provides an overview of the VAST goals and features. In Section 3, the architectural design of the VAST platform is presented. Ongoing and future research activities of the project are finally outlined in Section 4 focusing on knowledge modeling and computational aspects.

2. The VAST project overview

The aim of the VAST project is to investigate how the values at the basis of the European Union have been transformed over the ages. Across time, from antiquity to modernity, a value represents a message that is communicated through different mediums (e.g., text, visual art, drama, oral narration) and this message can change when the context and the society where citizens live change. As a first goal, VAST aims at representing values and associated messages

as they are extracted from sources of the past, such as for example literature texts and theatrical performances. As a further goal, VAST aims at collecting and digitizing the today messages associated with values as they are perceived by the general audience in the present days. To this end, the activities of the project are organized in three pilots that are characterized by i) different significant moments of the European history and ii) different types of sources that are exploited for extraction of the value messages. In particular, we focus on studying the transformations of values from ancient Greek tragedies to modern theatrical plays (i.e., *Pilot 1: Ancient Greek Drama*), from seventeenth century works of natural philosophy to science museum exhibits (i.e., *Pilot 2: Scientific Revolution Texts*), and from traditional European fairy tales to different types of storytelling (i.e., *Pilot 3: European Folktales*).

Pilot 1: Ancient Greek Drama. The Pilot 1 of VAST is related to values in ancient Greek tragedies and how they are perceived by contemporary theatrical plays and general audiences. It is mainly focused on Greek tragedies and their adaptations across Europe and the World along time. The goal is to analyze how the values of the antiquity, that are recognized to be discussed in specific tragedies (e.g., *Lysistrata*, Comedy, 411 BC), are revisited in the present through modern artistic reproductions, such as acting, music, and voice. The pilot aims at inspiring ideas and debates on what people find important in their own life and in their life with the others, like for example human rights, the right of political asylum, expansionism, genocide, the conflict between East and West, and the concept of the “other”. The main features of Pilot 1 are summarized in Table 1.

Table 1
The properties of Pilot 1

Feature	Pilot 1
General Context	Arts
Type of Narrative	Greek Tragedies
Time	Antiquity Present
Space	Western Europe
Communication Medium	Theatre
Theoretical background- Methodology	Philosophical, Philological, Theatrical
Communicators	Artists, Culture/Creative Industry Institutions
User Engagement	Theatrical Plays
Target Audiences	Spectators

Pilot 2: Scientific Revolution Texts. The Pilot 2 of VAST is related to values in texts of 17th century about natural philosophy and how they are perceived by experts in science museums and museum visitors like students and pupils. It is mainly focused on the early-modern era, the period known as the *Scientific Revolution* featured by great discoveries and inventions. The texts considered in the project are mostly about imaginary travel stories or fictional communities of ideal perfection in which the new intellectual achievements were embedded in an imaginary narrative context (e.g., *The Man in the Moone*, Francis Godwin, 1638). The goal is to analyze the shift in the message communicated by values concerned with the science of the past and

those concerned with the modern science. The pilot aims at promoting the organization of educational programs for museum visitors, focused events, such as talks and debates, where visitors are encouraged to share their experiences and visions about science-related values (e.g., freedom of research, science for public good). The main features of Pilot 2 are summarized in Table 2.

Table 2
The properties of Pilot 2

Feature	Pilot 2
General Context	Science
Type of Narrative	17 th Century Texts
Time	Early Modern Present
Space	Western Europe
Communication Medium	Museum Exhibitions, Educational Activities
Theoretical background- Methodology	Philosophical, Pedagogical, History & Communication of Science
Communicators	Curators
User Engagement	Exhibitions, Educational Programs, Special Events
Target Audiences	Visitors, Students, Pupils

Pilot 3: European Folktales. The Pilot 3 of VAST is related to values in folktales throughout the History of Europe and how they are perceived by storytelling experts in fairytale museums and museum visitors. Though fictitious, folktales are important simulations of the reality. Moreover, the variability of tales makes them the ideal case study for cross-cultural comparisons on social dynamics, including cooperation, competition, or decision making. The pilot is mainly focused on archetypical stories (e.g., the *Grimms' Fairy Tales*, 1812) and it includes texts from several European countries (i.e., Portugal, Italy, Slovenia, Greece, Cyprus). Folktale narratives are central to the construction of the self, embodied with memories, emotions, appetites, and culture-based values. The goal is to analyze the value dichotomies that are typically addressed in folktales, such as for example good/evil, right/wrong, punishment/reward, moral/immoral, trust/distrust, and male/female. The pilot aims at promoting events where a given folktale with a number of associated national adaptations are presented and the museum visitors can provide their feedback and emotions in the form of comments and/or storytelling with respect to the received value(s). The main features of Pilot 3 are summarized in Table 3.

3. The VAST architecture

The VAST architecture describes the modules and the tools that are employed to enforce the project activities and the pilot development (Figure 1). In particular, the VAST architecture is designed to support a computational-oriented approach focused on three main targets described in the following.

Table 3
The properties of Pilot 3

Feature	Pilot 3
General Context	Folklore
Type of Narrative	European Folktales
Time	Modern Present
Space	Western Europe
Communication Medium	Museum Exhibitions, Educational Activities
Theoretical background- Methodology	Philosophical, Pedagogical Psychological
Communicators	Curators, artists, storytellers
User Engagement	Exhibitions, Educational Programs, Special Events
Target Audiences	Visitors, Students, Pupils

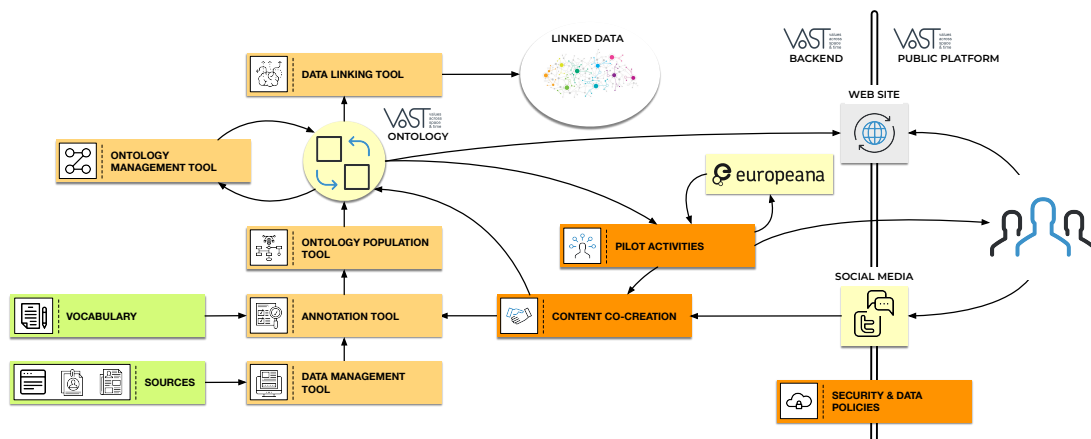


Figure 1: The VAST architecture

Historical content digitization. The VAST pilots require the acquisition of a number of heterogeneous historical sources that are selected to provide either explicit and implicit references to messages/interpretations associated with values. Document annotation modules and tools are designed in the VAST architecture to support historians and humanity scholars in the extraction of value-related knowledge from sources. A crucial aspect for annotation task is the specification of a reference vocabulary to support the work of human annotators and to avoid undisciplined proliferation of keywords. The idea is to go beyond conventional annotation approaches and to investigate the adoption of semi-automated solutions to the progressive enrichment of the reference vocabulary based on machine learning techniques (see for example [5]).

Multi-dimensional knowledge modeling and visualization. A key aspect in the VAST project is about the use of space and time as dimensions of analysis for observing the transformation of values throughout the considered historical sources. As a further issue, it is possible that a certain value is associated with multiple, different interpretations provided by distinct individuals in the same historical period. As a result, the VAST ontology meta-model needs to be defined not only around the specification of concepts featuring the values, but also around the association of each value with the keywords (i.e., the tags) used by individuals to denote the meaning of that value. In addition to the knowledge extracted from the historical sources through annotation, it is important to note that the VAST ontology has to contain the value interpretations provided by the final users (e.g., museum visitors, educators, students) as a feedback to the activities/events proposed within the pilots. The ontology-related modules and tools of the VAST architecture are designed to enforce modeling, population, and linking functionalities over the knowledge acquired through annotation. The VAST ontology aims at supporting i) dissemination tasks to the general audience through the VAST website, and ii) content creation activities to the final users involved in the pilots. In both scenarios the idea is to go beyond conventional visualization tools of the ontology contents and to provide interactive dashboards characterized by topic-driven organization of values aimed at highlighting the available value interpretations collected across space and time from sources and users (see [6] for a possible solution in this direction).

Collaborative content creation. Each VAST pilot aims at promoting focused events and activities to engage different target audiences on the project ambitions. The idea is to design interactive experiences where the final users are exposed to the value messages and interpretations coming from the Past, so that they can realize how these values have been transformed and differently perceived in the Present. The users involved in the pilot activities are encouraged to share their feelings according to modalities and mediums that are being defined within the project. A first modality is called *immediate* and it is based on the “synchronous” collection of user feedback during and/or at the end of the event/activity. In this modality, we expect to rely on interviews and questionnaires to enforce the user contributions. A further modality is called *remote* and it is based on the “asynchronous” collection of the user feedback some days/weeks after the participation to the event/activity. In this modality, social media channels represent a spontaneous medium that a user can exploit to share her personal feedback. As a further option, we plan to develop content-creation mechanisms where users are involved in collaborative writing experiences. The idea is to define a human-in-the-loop approach to *participatory storytelling* where users can contribute to the definition of a value interpretation by providing personal considerations as well as like/dislike reactions to the proposals of the other users (see for example [7]).

4. Ongoing and future work on knowledge modeling in VAST

One of the main VAST objectives is the development of methods and techniques capable of adding value to the historical collections of documents held by museums, archives, and other cultural institutions. The VAST architecture is designed to support multidimensional modeling

and visualization of knowledge extracted from the source documents and from the contents created by final users. Main ongoing project activities on knowledge extraction and modeling are devoted to address the following issues.

Document annotation. Historical documents are the main source of knowledge in VAST. One of the main goals of annotation is to tag and select the specific portions of documents that may be relevant for the understanding of values in time and space. In fact, each document in VAST is a historical source of information that is associated with a temporal dimension, typically the date of the document, and a spatial dimension, typically the country or the geopolitical entity where the document has been historically written and published. Annotations are performed by scholars and expert of the source documents under examination. The VAST scholars already defined the first version of the common vocabulary of VAST keywords to employ for document annotation. VAST keywords are mainly tags that have been defined with the goal of providing a first level of abstraction over the document contents but that still do not represent a conceptualization of the document contents. So, the goal of the VAST keywords is mainly related to the fact of overcoming the linguistic differences among documents (to this purpose all the VAST keywords are English words or short English sentences) and of providing a reference terminology that can be associated with document portions that are similar in terms of content but different in style and lexicon. The VAST keywords are integrated within the annotation tool and they have been associated with a set of shared annotation guidelines.


Ontology design. A reference meta-model has been defined for the VAST ontology in this first stage of the project, based on three main notions, namely i) *VAST annotation*, ii) *VAST interpretation* and iii) *VAST concept*. A **VAST annotation** represents an association between a source document, either a whole document or a portion of it, and one or more VAST keywords taken from the VAST controlled vocabulary. A **VAST concept** is the ontology representation of an entity, like a person or an institution, or a value considered in VAST, like freedom, democracy, and equality. The VAST concepts are interconnected by conceptual relations. In particular, a binary relationship among a pair of concepts can be specified to denote a semantic relation holding between them. A **VAST interpretation** is a relation between an annotation and a VAST concept. The association of concepts about entities with annotations is usually straightforward because such entities are mentioned directly in the source document or because there is a historical or philological evidence supporting that annotation. In case of values, the association of the concept representing the value and the document annotation is less straightforward because it derives from a specific interpretation that a scholar gives of the text at hand. For this reason, the design choice of the VAST ontology is to reify the notion of interpretation as a class of spatial ontology entities that represent the relation between a VAST concept, an annotation, and the scholars who propose or support that interpretation. Thanks to this design, the VAST ontology can model the existence of multiple interpretations of the same annotation. Moreover, we implicitly introduce a notion of *consensus* for interpretations in that different, and potentially controversial, interpretations might be supported by different scholars. This flexibility of design is motivated by the need of supporting different views of the VAST values across time and space. In fact, by selecting a specific time and space frame we can easily select the interpretations involved and - through those - the keywords that are associated with a value in that interpretations in order to study how the definition of values in terms of keywords and documents vary in the temporal and spatial dimensions. A more detailed description of the

VAST meta-model is provided in [8].

Ontology implementation. A number of conceptualization models exist that are already employed also in the field of cultural heritage, humanities, and social sciences [9, 10, 11]. In particular, for the VAST ontology implementation, we rely on the CIDOC Conceptual Reference Model (CRM) [12], a formal ontology specifically designed to support the integration, mediation, and interchange of heterogeneous cultural heritage information. The choice of CRM for implementing the VAST meta-model is based on multiple motivations. First, CRM is a standard that aims to offer a complete and relatively off-the-shelf solution that is readily applicable. In VAST we need to represent both temporal and spatial data, and CRM provides specific constructs to this end. In CRM, basic classes and properties are specified aimed at representing documentation of cultural heritage and scientific activities. The ontology is complemented by a number of modular extensions of the basic model. These extensions are designed to support different types of specialized research questions (e.g., representation of bibliographic data, geographical and archaeological data, data about social phenomena) and they are harmonized with the base ontology. In this respect, we mention i) CRMtex [13] for supporting knowledge representation of ancient documents, and ii) CRMsci [14] for archaeological excavations, scientific observations, and measurements. Furthermore, CRM supports an *event-oriented modeling* of knowledge, meaning that fact representation is articulated in a set of event-based relationships. This way, a value interpretation that we aim to model in the VAST ontology can be implemented as an instance of the class E5-Event to represent a specific association between a document text snippet, a keyword, and a scholar, thus providing an explicit description of a document annotation. Ontology implementation is in progress and details about modeling of the VAST entities through the CRM constructs are going to be investigated.

Ontology population and exploitation. As a subsequent activity, goal of future work, we will provide a semi-automated ontology population process, where text analysis and data linking techniques are exploited to aggregate keywords and contents, and to define mapping with the VAST ontology concepts in order to suggest possible interpretations to the scholars and experts. Moreover, we will also study how to integrate the knowledge extracted from documents, concerning the Past of values, with activities and user feedback and experiences, concerning the Present of values.

Acknowledgments

 This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101004949. This document reflects only the author's view and the European Commission is not responsible for any use that may be made of the information it contains.

This paper is partially funded by the RECON project within the UNIMI-SEED research programme.

References

- [1] F. Karsdorp, B. McGillivray, A. Nerghes, M. Wevers (Eds.), *Computational Humanities Research 2020*, volume 2723, Amsterdam, the Netherlands, 2020.
- [2] C. Biemann, G. Crane, C. Fellbaum, A. Mehler, *Computational Humanities - Bridging the Gap between Computer Science and Digital Humanities (Dagstuhl Seminar 14301)*, *Dagstuhl Reports* 4 (2014) 80–111.
- [3] G. Michael, *Agent-Based Modeling and Historical Simulation*, *DHQ: Digital Humanities Quarterly* 8 (2014).
- [4] The EU values, *The EU values*, <https://ec.europa.eu/component-library/eu/about/eu-values/>, 2020.
- [5] S. Castano, M. Falduti, A. Ferrara, S. Montanelli, *The LATO Knowledge Model for Automated Knowledge Extraction and Enrichment from Court Decisions Corpora*, in: *Proc. of the 1st Int. Workshop "CAiSE for Legal Documents" (COURT 2020)*, Grenoble, France, 2020, pp. 15–26.
- [6] S. Castano, A. Ferrara, S. Montanelli, *Topic Summary Views for Exploration of Large Scholarly Datasets*, *Journal on Data Semantics (JoDS)* 7 (2018) 155–170.
- [7] S. Castano, A. Ferrara, S. Montanelli, *Crowdsourcing task assignment with online profile learning*, in: *Proc. of the 26th Int. Conference on Cooperative Information Systems (CoopIS 2018)*, La Valletta, Malta, 2018.
- [8] S. Castano, A. Ferrara, S. Montanelli, F. Periti, *From Digital to Computational Humanities: The VAST Project Vision*, in: *Proc. of the 29th Italian Symposium on Advanced Database Systems (SEBD 2021)*, Pizzo Calabro (VV), Italy, 2021.
- [9] V. Mascardi, V. Cordi, P. Rosso, *A Comparison of Upper Ontologies*, Technical Report, Dipartimento di Informatica e Scienze dell'Informazione (DISI), Università degli Studi di Genova, 2007.
- [10] M. Doerr, *Ontologies for Cultural Heritage*, Springer, 2009, pp. 463–486.
- [11] C. Partridge, A. Mitchell, A. Cook, J. Sullivan, M. West, *A Survey of Top-Level Ontologies-to Inform the Ontological Choices for a Foundation Data Model*, Technical Report, Cambridge University, 2020.
- [12] M. Doerr, *The CIDOC conceptual Reference Module: an Ontological Approach to Semantic Interoperability of Metadata*, *AI magazine* 24 (2003) 75–75.
- [13] F. Murano, A. Felicetti, *CRMtex-An Ontological Model for Ancient Textual Entities*, in: *Decimo convegno annuale dell'Associazione per l'Informatica Umanistica e la Cultura Digitale*, Pisa, 2021.
- [14] M. Doerr, A. Kritsotaki, Y. Rousakis, G. Hiebel, M. Theodoridou, *CRMsci: The Scientific Observation Model-An Extension of CIDOC-CRM to Support Scientific Observation*, Technical Report, FORTH, 2014.