@neurIST: Infrastructure for Advanced Disease Management Through Integration of Heterogeneous Data, Computing, and Complex Processing Services

Siegfried Benkner, Antonio Arbona, Guntram Berti, Alessandro Chiarini, Robert Dunlop, Gerhard Engelbrecht, Alejandro F. Frangi, Christoph M. Friedrich, Susanne Hanser, Peer Hasselmeyer, Rod D. Hose, Jimison Iavindrasana, Martin Köhler, Luigi Lo Iacono, Guy Lonsdale, Rodolphe Meyer, Bob Moore, Hariharan Rajasekaran, Paul E. Summers, Alexander Wöhrer, and Steven Wood

***Abstract*—The increasing volume of data describing human disease processes and the growing complexity of understanding, managing, and sharing such data presents a huge challenge for clinicians and medical researchers. This paper presents the @neurIST system, which provides an infrastructure for biomedical research while aiding clinical care, by bringing together heterogeneous data and complex processing and computing services. Although @neurIST targets the investigation and treatment of cerebral aneurysms, the system's architecture is generic enough that it could be adapted to the treatment of other diseases. Innovations in @neurIST include confining the patient data pertaining to aneurysms inside a single environment that offers clinicians the tools to analyze and interpret patient data and make use of knowledge-based guidance in planning their treatment. Medical researchers gain access to a critical mass of aneurysm related data due to the system's ability to federate distributed information sources. A semantically mediated grid infrastructure ensures that both clinicians and researchers are able to seamlessly access and work on data that is distributed across multiple sites in a secure way in addition to providing computing resources on demand for performing computationally intensive simulations for treatment planning and research.**

***Index Terms*—Aneurysms, architecture, biomedical grid, biomechanical simulation, ontology.**

I. INTRODUCTION

THE VOLUME of data describing human disease processes, including their understanding, diagnosis, and management, is growing exponentially. The data are increasingly heterogeneous, comprising text, images, and other structures. Data are also diverse in context, from global guidelines based on broad epidemiological studies, through disease-specific knowledge gained from both *in vitro* and *in vivo* scientific studies, to individual patient-specific data. The data span scales from molecular, through cellular, to tissue, organ, and patient representations, and are increasingly interrelated in part due to recent breakthroughs in our understanding of disease processes through functional genomics studies. The huge volume of this information and its rate of growth represent an unprecedented data management challenge. In particular, it is often impossible for an individual, whether a clinician responsible for patient management, a scientist involved in research, or an engineer developing a new generation of imaging or interventional devices, to understand and assimilate this knowledge, and equally a challenge to dissect from it the information of use to them. This situation is encountered in the research and treatment of almost all diseases.

It is increasingly clear that new methods are required to manage, integrate, and interrogate the data in a manner that is accessible to a range of end users. A number of projects in the biomedical field are tackling the challenges posed by data fragmentation with the development of advanced Grid-based IT infrastructures. The European ACGT Project [1] developed a Grid-based platform for supporting multicentric postgenomic

Manuscript received October 30, 2009; revised March 12, 2010; accepted April 15, 2010. Date of publication April 29, 2010; date of current version November 5, 2010. This work was supported by the framework of the @neurIST Integrated Project, which is cofinanced by the European Commission under Contract IST-027703.

S. Benkner, G. Engelbrecht, M. Köhler, and A. Wöhrer are with the Department of Scientific Computing, University of Vienna, Vienna 1090, Austria (e-mail: sigi@par.univie.a.at; gerry@par.univie.a.at; koehler@par.univie.a.at; woehrer@par.univie.a.at).

A. Arbona is with GridSystems S.A., Palma de Mallorca 07121, Spain (e-mail: arbona@gridsystems.com).

G. Berti, P. Hasselmeyer, L. Lo Iacono, G. Lonsdale, and H. Rajasekaran are with the NEC Laboratories Europe, IT Research Division, NEC Europe Ltd., St. Augustin 53757, Germany (e-mail: berti@it.neclab.eu; lo_iacono@it.neclab.eu; lonsdale@it.neclab.eu; hasselmeyer@it.neclab.eu; rajasekaranit.neclab.eu).

A. Chiarini is with the B3C BioComputing Competence Centre, Bologna 40136, Italy (e-mail: a.chiarini@scsolutions.it).

R. Dunlop is with Infermed Ltd., London WC1B 3HW, U.K. (e-mail: robert.dunlop@infermed.com).

A. F. Frangi is with the Information and Communication Technologies Department, Center for Computational Imaging and Simulation Technologies in Biomedicine, Universitat Pompeu Fabra, and Networking Center for Biomedical Research in Bioengineering, Biomaterials and Nanomedicine, and Institució Catalana de Recerca I Estudis Avançats, Barcelona E08003, Spain (e-mail: alejandro.frangi@upf.edu).

C. M. Friedrich is with the Department of Bioinformatics, Fraunhofer Institute for Algorithms and Scientific Computing, St. Augustin 53754, Germany (e-mail: friedrich@scai.fraunhofer.de).

S. Hanser is with the Department of Medical Informatics, University Freiburg, Freiburg 79104, Germany (e-mail: susanne.hanser@uniklinik-freiburg.de).

B. Moore is with IDAC Ireland Limited, Dublin 2, Ireland (e-mail: bmoore@idacireland.com).

R. D. Hose and S. Wood are with the Department of Medical Physics and Clinical Engineering, University of Sheffield, Sheffield S10 2JF, U.K. (e-mail: d.r.hose@sheffield.ac.uk; steven.wood@sth.nhs.ukis).

P. E. Summers is with the Nuffield Department of Surgery, University of Oxford, Oxford OX3 9DU, U.K. (e-mail: pefs44@yahoo.co.uk).

J. Iavindrasana and R. Meyer are with the Geneva University Hospitals, Geneva CH1211, Switzerland (e-mail: jimison.iavindrasana@sim.hcuge.ch; rodolphe.meyer@sim.hcuge.ch).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TITB.2010.2049268

clinical trials on cancer. ACGT addresses the challenges associated with data integration with a semantic mediator for resolving semantic heterogeneities and data access wrappers to cope with syntactic heterogeneities. The Health-e-Child Project [2] developed a Grid-based healthcare platform for European pediatrics, aiming at a seamless integration of traditional and emerging sources of biomedical information. In the same vein, the U.S. caBIG initiative [3] developed a Grid-based collaborative information network for sharing of data and knowledge that aims at accelerating the discovery of new approaches for the detection, diagnosis, treatment, and prevention of cancer. The U.S. Biomedical Informatics Research Network [4] initiative fosters distributed collaborations in biomedical science focusing on brain imaging of human neurological disorders and associated animal models.

This paper presents the IT infrastructure developed through the @neurIST European Integrated Project [5], which aims to improve research and treatment of cerebral aneurysms by vertical integration of diverse biomedical information sources, whereas the aforementioned projects focus mainly on integration of biomedical data, the @neurIST infrastructure is unique in additionally providing data and computational services that support advanced end-user applications for risk assessment, for linking genetic information to the disease, for multimodal image processing, and for virtual endovascular treatment planning. @neurIST relies on a flexible, generic service framework, and advanced security mechanisms addressing the stringent privacy and security concerns of medical application domains. The service infrastructure has been augmented with semantic information based on a new ontology, which formalizes the conceptual space of @neurIST and provides the basis for semantic service discovery and data mediation.

The remainder of this paper is structured as follows. Section II provides an overview of the @neurIST project. Section III discusses collection and categorization of data within the @neurIST clinical study. Section IV describes the requirements of the @neurIST. Sections V and VI present the system and security architecture. Section VII provides insight into the @neurIST system in action followed by concluding remarks.

II. @NEURIST PROJECT

A. Project Overview

@neurIST is an information society technologies (IST) integrated project funded within the European Commission's Sixth Framework Programme. The @neurIST system focuses on supporting the research and treatment of cerebral aneurysms. The project developed a *generic* service-oriented IT infrastructure that supports transparent access to and integration of heterogeneous data from diverse sources together with on-demand computational services for supporting complex simulations. The infrastructure supports personalized patient management (i.e., data capture, referral, decision support, and treatment planning), as well as clinical research in cerebral aneurysms [6], [7]. The consortium brings together 30 multisectorial partners representing hospitals, universities, research institutes, and the industry across Europe.

At the heart of the project is the clinical problem of managing unruptured cerebral aneurysms and associated research into risk factors. Currently, aneurysm diagnosis and treatment relies on the interpretation of numerous pieces of information, which are recognized to give an incomplete picture of the disease. In part, this information comes from the patients themselves in the form of radiographic images, family history, and physiological measurements, and in part, it is derived from the understanding of the available medical literature and the past experience of the patients' clinicians. One shortcoming is that the current management process is subjective, and at the same time, the scales of risk involved are difficult to communicate. Moreover, the data is inherently incomplete. This places the clinicians and the patient in the situation of taking decisions involving significant risk while having an appreciation of the situation that is far from optimal. To cover all contingencies, patients may be subjected to tests and undergo treatments, which are of minimal real benefit or opt against potentially life-saving treatment simply due to a lack of knowledge, information, or understanding.

The same issues apply to many other diseases. The end result is that the quality of care is suboptimal, and its cost (both human and financial) is increased. @neurIST aims at keeping this cost low and raising the quality of diagnostics and treatment by building a distributed IT infrastructure, following a service-oriented architecture (SOA) model, which integrates Grid service technologies to bring together the different stakeholders involved in the research and treatment of a particular disease. Though the infrastructure developed by the project is used in relation to just one disease, i.e., cerebral aneurysm, the infrastructure's core technologies are transferable to meet the needs of other medical areas.

B. Treatment and Research of Cerebral Aneurysms

An aneurysm is a localized, blood-filled dilation of a blood vessel caused by disease or weakening of the vessel wall. If left untreated, the aneurysm can burst leading to severe hemorrhage and sudden death. Currently treatment of cerebral aneurysm is offered to almost all patients because, despite the misgivings of the clinical practitioners, there is insufficient evidence to support a nonintervention decision. Taking into account the prevalence of this disease [1–5%], the annual rupture rate [0.2–1%], and the average treatment and first-year follow-up costs of patients [50 000 Euro], it is estimated that, in Europe alone, unnecessary interventional or surgical procedures cost several hundred million Euros per annum [8], [9]. An integrated IT-system that provides a means to comprehensively understand and manage cerebral aneurysms together with a personalized risk-assessment strategy would lead to better patient care and reduced costs.

Debate regarding the optimal treatment of patients harboring unruptured intracranial aneurysms has developed in recent years, stemming largely from the influence of three main factors

1) Increased diagnosis through use of noninvasive imaging, such as computed tomography (CT) and MRI scans.
2) The results of the International Subarachnoid Aneurysm Trial (ISAT) indicating that endovascular therapy using

coils inside the blood vessels is safer than surgical intervention in some situations [10]–[12].

3) The International Study of Unruptured Intracranial Aneurysms (ISUIA) indicating that the chance of an aneurysm rupturing, in particular, patient populations to be very low (6 in 100 000 people) [13].

In combination, these three factors have led to considerable uncertainty with regard to the optimal treatment for patients harboring unruptured intracranial aneurysms. Most importantly, unanswered questions remain as to which particular lesions carry a hemorrhage risk significant enough to justify intervention, and when aneurysm treatment is warranted, which method should be used.

The natural history of cerebrovascular aneurysms is complex and, to a large extent, still unknown though genetic, physical, and environmental factors are thought to play a key role in pathogenesis and progression. The identification of at-risk individuals with higher chances for developing brain aneurysms and their rupture are two extreme decision-making situations in the screening and/or handling of this disease, where integrative decision support tools are essential for health professionals.

C. @neurIST-Specific Challenges

The treatment of intracranial aneurysms is typically performed in tertiary referral centers. These are hospitals that receive patients for treatment on the basis of an initial diagnosis made elsewhere. The imaging data from the initial center is used as a starting point along with rechecking of the clinical history with patient prior to further imaging and laboratory tests.

Radiological images from digital subtraction X-ray angiography, CT, and MR play a central role in the course of a patient's care. The diagnostic and treatment use of these images is almost exclusively based on visual inspection and measurements of the spatial dimensions of the lesion. Angiographic images provide little or no information about the conditions of blood flow within the aneurysm and its related vessels, albeit it is known that they play an important role in the initiation, growth, and rupture of cerebral aneurysms. A more advanced use of these images comprises modeling and computing of aneurysmal flow patterns and shape models. This requires considerable computation power and interaction software that are not frequently found within clinical centers. Outsourcing of such simulation services is a meaningful option, but raises many security and privacy concerns that need to be addressed. Furthermore, though most scanners use the Digital Imaging and Communications in Medicine (DICOM) Standard [14] for storage and communication of these images, some proprietary standards still exist for imaging data that would be of use in hemodynamic modeling, and hence, require data integration.

The @neurIST project also encompasses active research undertakings in the areas of genetics, transcriptomics, and biomechanics of intracranial aneurysms. In the research, investigations are typically performed by external institutions requiring access to patient-specific anonymous information and subject to approval from the responsible ethical board, and the consent of the patients. These investigations may one day yield consolidated, clinically proven, new information about the diseases, and therefore, there exists the further goal of being able to return clinically significant results generated outside the hospital context for incorporation into the patient record and care management.

D. Innovations in @neurIST

The @neurIST project has developed innovative solutions that address the earlier clinical, research, and informatics challenges. On the clinical side, the project has developed a single environment (the @neurIST platform) to manage all patient data with an emphasis on that related to cerebral aneurysms (although the platform itself is generic). The platform offers clinicians the ability to analyze, interpret, and make use of knowledge-based guidance in their treatment process. The platform also offers them the facility to simulate their treatment plans using computational models and compare multiple options before deciding on the personalized treatment solution for the patient.

For researchers, the @neurIST platform provides integrated access to clinical data from a distributed network of medical centers, making it feasible to obtain a critical mass of data related to aneurysms. It is emphasized that only those records for which the patient has provided consent are accessible through the system. The clinical sites acquired such data in the course of the @neurIST study conducted as part of this project (see Section III), designed to mirror the course of routine clinical practice. The researchers are further able to perform large-scale simulations using the computational infrastructure provided by the platform, as well as to relate the information in the clinical records to the academic literature. It also offers stent developers the means to simulate the behavior of different stent geometries and/or different aneurysm geometries using the @neuCompute infrastructure for running series of flow simulations.

On the informatics side, the platform brings together cutting edge technologies in the fields of distributed data access and integration, data mining, semantic technologies, on-demand high-performance and high-throughput computing, advanced visualization techniques, and security. The services that make use of such a diverse array of technologies are tied together using the SOA approach offering the possibility for a multitude of interested entities to take part in the system with minimal technical obstacles.

III. @NEURIST CLINICAL STUDY

A. Collection of Patient Data in @neurIST

The @neurIST project requires extensive clinical data in order to implement, test, and support the tool and system development as well as @neurIST-based research. The initial vehicle for data collection has been the @neurIST Study. In this study, clinical history collected by a research nurse, or drawn directly from hospital electronic records were used to support genetic and transcriptomic investigations requiring specifically collected blood samples, along with morphohemodynamic studies based on 3-D imaging data collected in the course of routine care of the patients. The process of data collection was intended to populate the database with a core of high-quality clinical data acquired under research conditions. Seven clinical centers, located in five European countries were involved in the @neurIST

Study, collecting data over a period of two and a half years after obtaining approval from their local ethical committees. The targets for clinical history and blood sample collection within the study were 400 patients with ruptured intracranial aneurysms, 400 with unruptured intracranial aneurysms, and 400 age and gender-matched controls.

B. Data Categories Inside @neurIST

Managing and processing patient-related data is a core activity in @neurIST. This section provides an overview of the types of data used, together with associated issues, including privacy requirements, users and producers, storage and access mechanisms, as well as data quality, and provenance. In @neurIST, it is useful to distinguish between data, which is either

1) *primary* or *derived* (i.e., produced from primary data directly or via a computational process within @neurIST);
2) *personal* (i.e., related to a specific individual) or *nonpersonal* (e.g., averaged data or publications);
3) *structured* (i.e., represented in a relational data model) or *bulk* (i.e., mass data not commonly stored in a database, such as images or simulation results). Typically, the metadata describing bulk data is structured data.

In addition, data has a certain level of confidentiality, which entails appropriate access restrictions. Personal data has a high level of confidentiality, but nonpersonal data may also be confidential (e.g., new stent designs).

The primary personal data of the patient includes as follows.

1) *Patient clinical data* includes the data contained in the patient record or clinical information systems (CIS) of the clinical centers. Within @neurIST, patient data is structured according to the clinical reference information model (*CRIM*), which encompasses the clinical data relevant to a patient with an aneurysm. It includes data, such as patient administrative data, personal history—including known and suspected risk factors, signs and symptoms, vital signs and laboratory exams, reports summarizing imaging data, and annotations of the features of interest like aneurysm size and location, proteomics and genetics data, treatment, and follow-up evaluations.
2) *Radiological data* including images from digital subtraction X-ray angiography, CT, and MR.
3) *Blood analytics and genetic data* from each patient. They are used to obtain genetic information in the form of single nucleotide polymorphism (SNP) data and RNA for transcriptomic assessment.

This data is deidentified and pseudonymized before it is used for research purposes in the @neurIST system leading to the generation of derived data.

The derived personal data consists of the following.

1) *Biomechanical data* includes all data obtained during data processing of aneurysm geometries from the medical images, in particular, candidate biomechanical risk factors.
2) *Quality assurance annotations:* The earlier clinical primary data, collected in the @neurIST study, is subject to a quality assurance process, resulting in annotations, e.g., on image quality and completeness of data in the CRIM.

The @neurIST platform ensures that such diverse categories of data distributed across multiple sites are made available to the users of the platform in a seamless and transparent way via appropriate services. The terms relating to the medical description of a patient with an intracranial aneurysm are compliant with the HL7 standard, facilitating future interrelation of data from sources not explicitly covered in the project.

IV. REQUIREMENTS FOR @NEURIST SYSTEM ARCHITECTURE

Though the principal focus of @neurIST is cerebral aneurysms, the implications of the @neurIST solution are far wider in scope. The point-of-care integration of radiological, genomic, and clinical data with epidemiological and other data is applicable in all fields of medicine. The generic processes that occur repeatedly in most studies and treatment processes of diseases are as following.

1) *Obtaining relevant clinical information about each patient's clinical problems:* Clinical data is often widely distributed, particularly if a patient has undergone a healthcare journey involving diagnostic and treatment procedures in multiple medical institutions. The clinician may require additional data at successive stages in the patients' treatment process making the data itself a dynamic entity.
2) *Providing clinical decision support and guidance on best practice:* Informed decisions about further testing, treatment options and procedures are required at every step. Simply consolidating the numerous pieces of information, and recording their interpretation is one aspect of this. Providing context, so that appropriate best practice and recent findings are adequately considered and can be reviewed by all involved in the clinical management is the second aspect.
3) *Offering simulation services,* which allow clinicians and patients to assess complex treatment options. These services can allow for more efficient procedures by allowing rehearsal and ensuring the treatment device of choice is in stock for the procedure. These services can also enable new treatment solutions to be devised and tested by medical device manufacturers.
4) *Creating normalized population-based datasets,* where clinical data is gathered during the process of care, resulting in the reduction of the cost of population studies and an increase in their scope. At the same time, the information can provide new insights into management, as well as an audit mechanism for the quality of care.
5) *Providing knowledge discovery services,* by facilitating the acquisition and processing of complex data sources, such as the vast array of biomedical literature or publicly accessible bioinformatics databases.
6) *Local responsibility for local data*, with the clinical data collected and stored at the site of origin as a Biological Information System (BioIS) that acts local as a database of clinical relevance, and via @neurIST as a component in a federated database (see Section V-C for more details).

Implementing these requirements involves addressing issues, which are as follows.

1) *Combining heterogeneous data sources,* which are distributed geographically, encoded in different formats and may involve different languages and nomenclatures. Inte-

gration of heterogeneous data is key to supporting earlier processes 1–2 and 3–4.

2) *Providing sufficient computing power* for complex simulations and other data processing tasks both in terms of capability and capacity (cf., process 3).
3) *Maintaining the privacy of patient data and securing the entire IT infrastructure* to ensure compliance with legal regulations and to ward off malicious attacks.

The @neurIST system architecture as described next has been devised around these core requirements, comprising service-oriented subsystems for data integration, @neuInfo, and for the provision of simulation services, @neuCompute, with a common security and privacy layer as described in Section VI.

V. @NEURIST SYSTEM ARCHITECTURE

The @neurIST system aims to satisfy the aforementioned requirements through an infrastructure that provides seamless and secure access to distributed medical data and computation resources to the end users of the system. The @neurIST system has essentially two modes or cycles of operation. In the first mode, it operates as an integrative decision support system to help in diagnosis, risk assessment, treatment planning integrating complex, and multifactorial information. In a second mode of operation, it enables *in silico* research and the linking of genetic and phenotypic evidence available in large-scale text databases, so that new knowledge can be extracted, structured, and transposed for its exploitation in the decision support operation cycle. While the decision support system largely deals with identified patient data, the research system has access only to anonymized datasets. These two cycles are underpinned by several integrative applications, which are organized into suites that help in data collection, simulation, knowledge discovery, risk assessment, etc., along with system and middleware components that connect databases and services distributed at different sites. The system and middleware components take care of security, data access, data transport, and computing resources.

The @neurIST system and infrastructure has been developed using multiple technologies and deployed across a wide geographic distribution. @neurIST adopts the SOA approach to integrate the diverse components of the system, and uses open standards and technologies from the Internet, Grid, and medical domains. Fig. 1 shows the layered view of the reference architecture of @neurIST with their constituent components. The system can be logically divided into three layers—application, middleware, and resource.

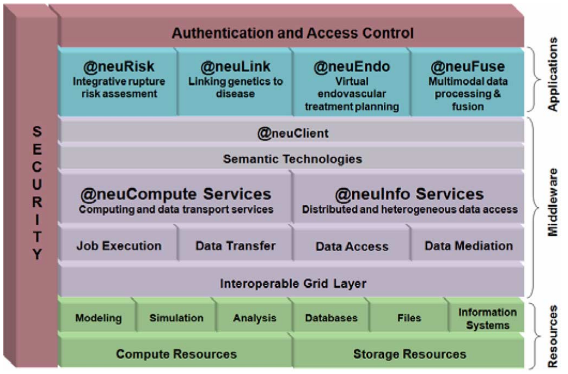

Fig. 1. @neurIST reference architecture—layered view.

A. Application Layer

The application layer consists of the application suites that provide end users with functionalities for research and clinical treatment of aneurysms.

The *@neuLink* application suite [15] is targeted at basic research users within @neurIST, essentially geneticists and epidemiologists. Its objective is to link genetic information with disease information. The information is gathered from public bioinformatics databases and from the BioIS. In this way, it supports the clinical decision process with evidence from analyzed experiments and knowledge discovery from other data sources. The aggregated information is provided via a Web-application to a human user, as well as via the @neurIST data access infrastructure, to support other application suites. The main modules support the search for candidate genes and associated genetic variations (SNPs) from text and from the aggregated protein–protein interaction database PIANA [16]. The suite supports the interpretation of microarray experiments and provides results from applying data-mining methods for SNP association studies and risk assessments from the anonymized patient population. The data-mining process is implemented using the KNime workflow engine [17] to support better maintenance possibilities and to give a visualization of the workflow to the user. Initial results on knowledge discovery were presented in [18].

@neuLink consists of research-oriented modules including text mining and biomedical networking for expression analysis: a data-mining module and a probabilistic risk modeling module. Together, these modules support the following tasks:

1) finding disease-related candidate genes and variations [19];
2) analyzing and visualizing protein–protein interaction networks to find putative disease associated proteins;
3) expression analysis and integration of results;
4) data mining for genetic and general risk factors in the @neurIST databases.

Due to the amount of data and the complexity of algorithms in use, knowledge discovery processes are time-consuming tasks. Therefore, mining tasks in @neuLink are not performed online on user request, but executed at regular intervals, currently on a daily basis, with aggregated results stored in databases/indexes for later retrieval.

The primary use of *@neuFuse* is in the directing the processing and visualization of imaging data leading to, and resulting from computation fluid dynamics simulations. The @neuFuse suite typically receives raw 3-D angiographic data that is processed using an embedded tool chain. Through this tool chain, the angiographic scan is transformed into an annotated mesh, which is ready for use in a partial differential equation (PDE) simulation, including boundary conditions adapted to the specific arteries related to the problem. Moreover, the application enables medical imaging, biomedical instrumentation, and derived datasets to be organized in time and space, and to be inter-

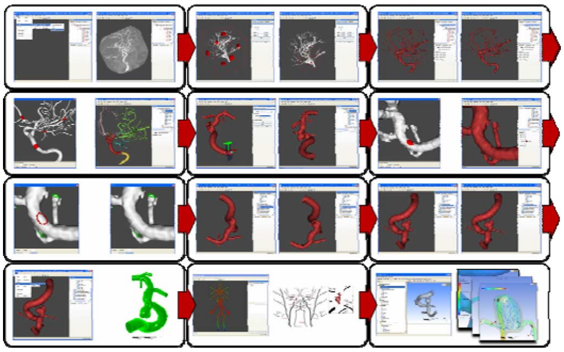

Fig. 2. Hemodynamics analysis, preparation, and visualization with @neuFuse: medical image view and preliminary model (upper row), clipping to the region of interest (ROI) (second row), correction of topological artifacts (second and third row), assignment of boundary conditions from the 1-D model, submission to @neuCompute (not shown), and visualization of simulation results (bottom row).

actively visualized independently or simultaneously. @neuFuse provides information to clinical users as well as researchers via the @neuInfo middleware. This information is also accessed by other application suites. The @neuFuse *front end* is a high-end workstation application aimed at visualization and manipulation of image and *derived data.* The front-end component of @neuFuse is a stand-alone C++ application based on the openMAF framework [20] that can also be launched from other suites, for example, from @neuEndo, on the local workstation, where @neuEndo is accessed to perform a preplanning session and get the stent position or from @neuRisk, in order to perform a hemodynamic analysis on the patient dataset and to extract a complex rupture risk index to populate for derived data in the derived data store (DDS, see Section V-C). From the front end, it is possible to retrieve patient data, by launching the *data client*, which is an external application that will receive the user query from the front end, authenticate the user, then query @neuInfo and ultimately manage the retrieval of the needed images to be analyzed by the user.

Fig. 2 shows the hemodynamic tool chain implemented in @neuFuse: starting from medical images retrieved from hospital anonymized databases, a raw geometrical model is extracted, then it is topologically healed and an automatic centerline is extracted. Boundary conditions can be extracted with a model that uses patient-specific data contained in the clinical record. Then, the simulation of the hemodynamic model is run on @neuCompute, and finally, results are stored using DDS services. Details of the tool chain and its validation with over 35 users have been published elsewhere [21], [22].

@neuEndo is a tool that uses cutting-edge finite-element analysis (FEA) and computational fluid dynamics (CFD) technology based on ANSYS Mechanical and CFX [23] to assess mechanical and flow performance of stents used for treatment of aneurysms. It is aimed at both industrial and clinical end users. For industrial users, it is a design tool during stent development [24], while clinicians can use it to assess the suitability of commercially available stents for particular patient conditions [25]. @neuEndo currently has two distinct applications. One looks at the stent from a structural perspective and the other at the aneurysm from a flow perspective. @neuEndo uses the computational services provided by @neuCompute to carry out its simulations and accesses stent geometry data via corresponding @neuInfo services. These services are used when @neuEndo needs to carry out computationally intensive calculations.

@neuRisk forms the core of the integrative decision support system by providing individualized risk assessment, clinical recommendations, and care planning for patients with unruptured cerebral aneurysms. @neuRisk is based on computer-interpretable guidelines (CIGs) developed in the PROforma guideline representation language [26]. @neuRisk utilizes, via corresponding @neuInfo services, all available information for identifying patients, who are at high risk for rupturing asymptomatic cerebral aneurysms. It makes use of the functions or knowledge generated in other application suites to include genetic and simulation information in the risk assessment process. A key innovative element in @neuRisk is its ability to integrate information from a wide range of origins through an underlying formalism called argumentation [27], which allows reasoning in the presence of uncertainty and provide supporting and nonsupporting evidence behind alternative treatment options. Therefore, in addition to being useful for clinicians, this system is particularly useful for counseling patients and facilitating their comprehension of benefits and risks in deciding for their own treatment path. @neuRisk will eventually improve the decision-making processes and reduce current overtreatment. The reasoning process of @neuRisk is based on the AREZZO inference engine [28], for which a distributed version based on @neuCompute has been developed.

B. Middleware Layer

The middleware layer in @neurIST is designed to decouple the application layer from the resource layer. The @neurIST middleware constitutes a service-oriented infrastructure offering generic services for data access, data staging, semantic mediation, and Grid computing. As a consequence, application suites do not need to be aware of the exact physical or logical sources of the data or computing resources, significantly reducing the complexity while enhancing maintainability.

The resulting middleware layer consists of the @neuInfo and @neuCompute infrastructures, offering data and compute services to the application suites. Data and compute services are based on standard Web Services technologies; they are defined via the Web Services Description Language (WSDL) and are securely accessed through the exchange of SOAP messages. The @neurIST middleware has been developed on top of different Grid middleware infrastructures including Fura [29], Grid-enabled Medical Simulation Services (GEMSS) [30] and Vienna Grid Environment (VGE) [31], [32].

To facilitate access to data and compute services from the @neurIST application that suites a client-side service invocation framework called @neuClient has been developed. @neuClient provides a high-level abstraction layer offering a unified application programming interface for accessing @neuCompute and @neuInfo services from C++ and Java application code.

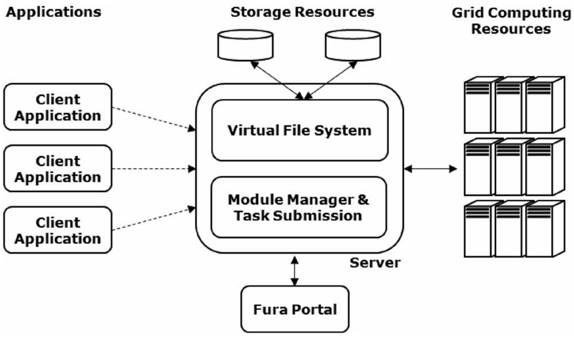

Fig. 3. @neuCompute infrastructure.

This includes methods for executing simulation jobs based on remote @neuCompute services, for accessing different kinds of @neuInfo data services, and for transferring (potentially large) files directly between different services.

@neuCompute is a self-contained Grid middleware, developed mainly on top of Fura [29] that allows the Grid enablement and distribution of applications on heterogeneous computational platforms. @neuCompute features a Web-based GUI, wizard-guided installation, and is Web Service compliant. @neuCompute is focused on distributed computing and provides services for dispatching and managing jobs, Grid data access, resource monitoring, and job parallelization. The architecture of @neuCompute follows the service-oriented paradigm. A deployment of @neuCompute, as shown in Fig. 3, is typically composed of one or more servers, a portal acting as a GUI, and a number of resource agents (numbering from a handful to thousands). The three components have a modular structure. The server and agents are extensible through a plug-in system, and the portal through new Web applications. WSDL interfaces provide external access to the Grid and also permit the internal communication among the components.

Regarding resource support, @neuCompute can be extended over heterogeneous resources: desktops, servers, and clusters running Windows, Linux, MacOSX, or UNIX. @neuCompute can distribute computations over them, on both dedicated and scavenging (CPU cycle stealing) modes.

@neuCompute features the concept of *modular* or abstract Grid-enabled application. The modules describe the command-line application interface and provisioning details (binaries), and they provide built-in facilities for parametric sweeps. Data staging and caching policies are also defined at the module level. Jobs (and modules) can be exported and imported between Grids. Module design and job management can be achieved either through a programmatic software development kit (SDK) or interactively through the portal. Three module families have been developed as compute services, including the ANSYS mechanical model, used primarily by the @neuEndo suite, the ANSYS-CFX module, used by @neuFuse, and the Arezzo module, servicing the computational engine used by @neuRisk.

The administrative tools in Fura include an installation and configuration wizard, a built-in role-based user system, authorization via access control lists, as a local refinement to the @neurIST role authorization system, resource usage policies, and error detection and recovery facilities.

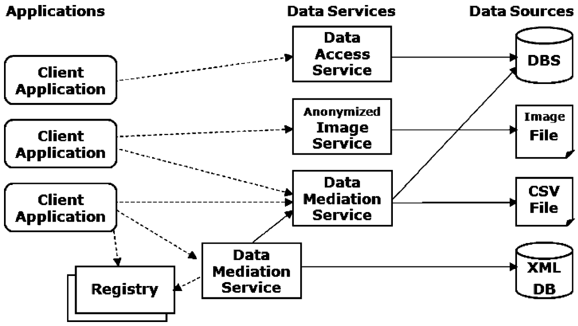

Fig. 4. @neuInfo data service categories.

@neuInfo offers a generic data management and integration framework that supports the provision and deployment of data services. Data services virtualize heterogeneous scientific databases and information sources as Web Services, enabling transparent access to and integration of relational databases, Extensible Markup Language (XML) databases and flat files. @neuInfo data services have been developed based on the VGE [31], [32] and utilize advanced data mediation and distributed query processing techniques based on Grid Data Mediation Service (GDMS) [33], OGSA-DAI [34], and OGSA-DQP [35] (see Fig. 4).

Data services hide the details of distributed data sources, resolving heterogeneities with respect to access language, data model, and schema. Since preserving the autonomy of data sources and ensuring access to live data are key requirements, data integration is based on a mediator approach, where local data sources are integrated bottom-up by mapping local database schemes into a virtual global schema. Internally, data services utilize the defacto standard for Grid-based data access and integration OGSA-DAI for accessing actual data sources. Client applications usually access data services by submitting Structured Query Language (SQL) queries and downloading query results in the form of OGSA-DAI-compliant XML documents.

Data services can be set up in different configurations all providing the same interface to clients. *Data access services* (DAS) provide access to a single data source, *anonymized image services* (AIS) support the transfer of (potentially) large image data of different modalities, and *data mediation services* (DMS) offer transparent access to multiple data sources via a global virtual schema. The virtual schema of a DMS provides an integrated, global view of the underlying local data sources. DMS translate queries with respect to the global schema into local queries, manage the access to the local data sources, and integrate results from local queries according to the global schema. Different tailor-made views of distributed data sources may be provided for specific usage scenarios, by setting up different variants of DMS. DMS can be composed recursively, i.e., a DMS can act as a local data source to another DMS, resulting in a tree-structured data integration scenario, where leaves represent physical data sources and nodes represent DMS global virtual schema.

A versatile mediation engine for automatically translating queries against the global schema into queries against local data sources constitutes the core component of mediation services. The mediation engine relies on a flexible mapping schema that

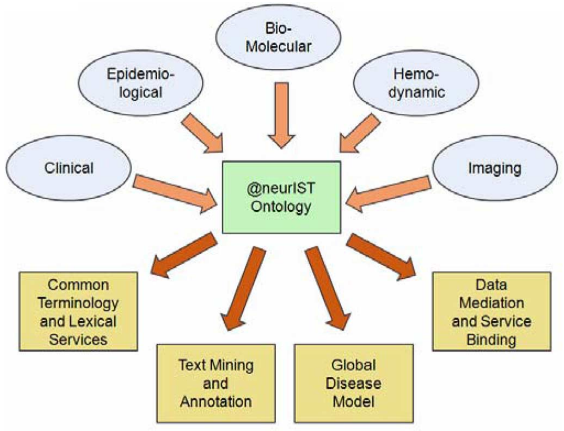

Fig. 5. @neurIST ontology.

controls the translation of global queries into local queries and the associated transformation of data formats, if any. Mapping schemes are based on the Global-as-View (GaV) [36] data integration approach, where the global schema is described in terms of the local schemes. The generation of mapping schemes is supported by semantic technologies as described later.

In order to optimize complex data integration scenarios, DMS may be configured to support distributed query processing. For distributed query processing, the data mediation engine relies on OGSA-DQP [35], which has been integrated to coordinate the distributed execution of queries to local data sources utilizing a set of additional server machines, as specified by the service provider.

Data mediation, which was initially based on handwritten mapping schemas [33], is facilitated through semantic technologies in order to reduce the integration effort. Support for semantic data mediation [37] relies on using the @neurIST ontology [38] (see in the following), a semantic broker [39], and a semantic query resolver. Individual data sources have been semantically annotated by linking their data schemas (tables and columns) to ontology concepts. The semantic broker utilizes these annotations to determine from a list of ontology concepts all data services that may be of relevance to these concepts. The semantic query resolver accepts queries at the level of ontology concepts, interacts with the semantic broker to receive a list of relevant data services, and then, attempts to automatically derive a mapping schema for integrating all relevant data sources. The derived mapping schema, which may have to be further refined manually, can then be utilized for setting up a new DMS.

Semantic data mediation within @neurIST leverages the W3C standards Web Ontology Language (OWL) and Resource Description Framework (RDF) for representing the ontology and data source annotations, and relies on various open source tools for semantic brokering and the semantic query resolver, including Sesame [40], OWLIM [41], and UNITY [42].

The *@neurIST ontology* [38] captures the conceptual space of @neurIST by modeling the relevant conceptual entities from clinical medicine, molecular biology, epidemiology, simulation, and disease and risk factors (see Fig. 5). It is intended to serve as a module in a complex architecture aiming to provide a better understanding and management of intracranial aneurysms and subarachnoid hemorrhages. Due to the integrative structure of the project, the ontology needs to represent entities from various disciplines on a large spatial and temporal scale. Initial term acquisition was performed by exploiting a database scaffold, literature analysis, and communications with domain experts. Uses of the ontology in the project, include terminological control, text mining, annotation, and data mediation. Even within one knowledge domain, distributed data and their respective data models may vary on different semantic and syntactic levels. A common terminology with well-defined relationships among the terms is the major tool for bringing together various data models on the semantic level while relying on underlying functionality for unification of data access. In @neuInfo, the domain ontology is employed as the semantic Data Grid glue, describing both data schemas from individual databases as well as the wrapping data services—both employing the same semantic markup. Ontological markup or annotation on the data schema level thus serves 1) to identify data services based on certain queries; 2) to identify attributes in database tables; and 3) to bind together equal attributes from different data schemas using SQL join and union operations.

The @neurIST ontology is based on the OWL. Following the current state-of-the-art in Semantic Web technology, the Protégé [43] editor is used with the OWL-DL [44] as the language for modeling the ontology. Both the RacerPro [45] and Pellet [46] reasoners are continuously employed during the editing process to ensure logical consistency of the ontology.

Besides supporting semantic data mediation, the @neurIST ontology serves the following additional purposes.

1) Representation of the project's conceptual space by describing a specific, well-delineated domain and introducing the proper definitions and descriptions for all relevant concepts and relations in this domain. It identifies the allowed terms and their respective meanings in the clinical and experimental documentation as well as in the development of databases and GUIs.
2) *Provision of definitions:* The concepts are arranged in a taxonomy and associated via relations (properties) by applying formal axiomatic definitions to constrain their interpretation and well-formed usage. The ontology gives both textual and formal (i.e., description logic) definitions for all required entities and is used as a standardizing instrument throughout the project, where ambiguities are frequent (e.g., clinical terms).
3) *Links to other terminologies:* The modeling approach conforms to the principle of reusing widely acknowledged existing terminologies/ontologies (Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) [47] and Foundational Model of Anatomy (FMA) [48]) and involves a mapping to the Unified Medical Language System [49], where possible.
4) *Provision of user-defined views:* The @neurIST ontology classifies entity types according to formal principles, i.e., according to the top-level categories of the DOLCE top-level ontology. We introduced a second hierarchy intended to represent human knowledge on the disease in different

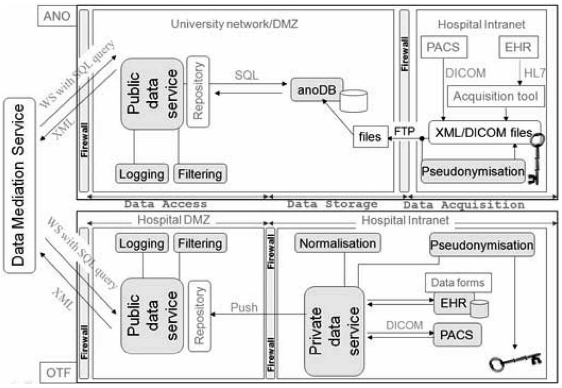

Fig. 6. ANO and OTF architecture of the BioIS.

contexts, which refer to the scientific areas involved. We integrated the possibility to generate user-defined subsets of ontology types, for example, clinical, epidemiological or simulation-related types. These entities and the corresponding synonyms build up the "dictionaries" used in the development of automatic text mining strategies in the @neuLink application suite.

C. *Resource Layer*

The resource layer consists of the modeling, analysis, and simulation applications, the associated computational resources (e.g., clusters and other health professional cards (HPC) systems) and the diverse data and information resources (e.g., databases and other information sources). The compute resources are offered by service providers to run computation intensive processes required by the different application suites of @neurIST. The database offer storage resources for simulation results and patient data.

In particular, the databases in the clinical centers that store patient information form a part of the BioIS, which connects the clinical centers within the @neurIST system. It is an @neurIST-specific implementation, driven by the need to gather clinical data from multiple clinical centers for research and treatment purposes.

The BioIS exchanges information with the application suites via the @neuInfo system (distributed queries and updates). This includes requests for clinical data, as well as patient-specific data outputs from other @neurIST services. Clinical information will remain at its source within the CIS of participating centers. The BioIS subsystem interfaces with the CIS databases.

As shown in Fig. 6, the BioIS supports two architectural styles: the anonymized database (ANO) model and the on-the-fly (OTF) model. The ANO model, which is based on a dedicated database containing a pseudonymized copy of the patient records (anoDB). The OTF model is characterized by a direct communication with the CIS via a private data service. Access to the BioIS from outside is realized in both models by means of an @neuInfo DAS deployed in the demilitarized zone (DMZ). In order to provide integrated access to the different BioIS systems in different hospitals and clinical centers, an @neuInfo DMS is provided, which hides the physical location of data from the application suites.

For exploration of pseudonymized patient data in the CRIM-based databases, a special user interface called @neuBrowser has been developed, which can be used in clinical as well as in research environments.

While the BioIS deals with clinical (primary) data, the DDS is an infrastructure for research (derived and secondary) data. Again, @neuInfo services are used to exchange data between the DDS and application suites.

VI. SECURITY ARCHITECTURE

A. *General Overview*

The security architecture developed for @neurIST focuses on the specific security requirements of a distributed and federated service environment for conducting research and supporting treatment in the health domain. It provides a security system, which ensures that patient data is made available only to authorized personnel, and that the privacy of the patient is preserved at all times.

Whereas these requirements relate to the flow of data from the CIS to the research setting, there is a similar necessity to strictly control the data flow between the research domain and the clinical decision support system of @neurIST. This is in part to prevent inappropriate treatment of patients on the basis of preliminary research results, but also to further ensure the risk of a breach in privacy is minimized. Here as elsewhere, the policies are consistent with the recommendations of a legal and ethical advisory board within the @neurIST consortium that provides guidance on issues of privacy, research ethics, including the return of research results to the clinicians and patients, and for providing guidance on the requirements for security. An example was the determination of which fields can be considered as valid in a pseudonymized file (e.g., ethnicity or age).

An important consideration is the demand for an efficient infrastructure to authenticate and authorize @neurIST stakeholders across multiple domains and borders. The security architecture achieves this by providing the following:

1) virtual organization (VO) and trust model underpinning the security of the system;
2) security credentials used for authenticating users (local and remote);
3) policies for authorizing users based on attributes, such as roles and location (local and remote);
4) policy enforcement systems (local and remote);
5) fine-grained access control at the data level.

Moreover, @neurIST utilizes Public Key Infrastructure (PKI) for message-layer security and https for transport-layer security.

B. *VO and Trust Model*

The @neurIST system gathers all the participating entities, such as clinics, research institutes, device manufacturers, and service providers into one large VO. This is important because, a credential issued to an @neurIST user should be recognized (either directly or through the construction of adequate trust paths) by all participating entities in the system to grant appropriate access to services and data in the system. To provide the

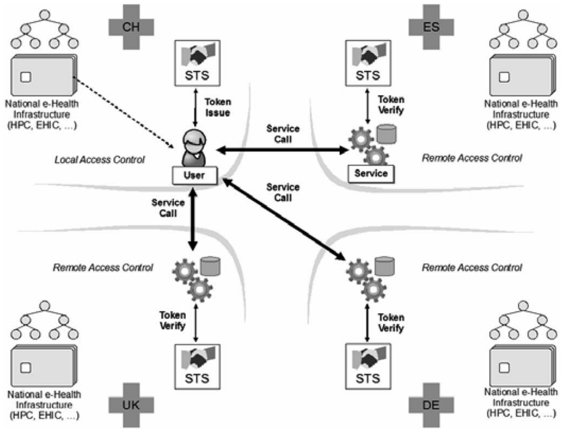

Fig. 7. Distributed access control.

required flexibility and manageability of such a large distributed service environment, the establishment and management of the VO should not rely on any centralized component. In @neurIST, this is realized by introducing a separation between the local and remote VO management, which does not require a system-wide harmonization of local identification and authentication policies and schemes.

The medical institutions and service providers run through a contracting phase to setup a research project. An additional step introduced into this common process is the exchange of security tokens. The security tokens here are the certificates of each partner's *Security Token Service* (STS). The exchanged tokens are, henceforth, trusted by each participating site to (remotely) authenticate system participants in the scope of the @neurIST study.

Even though the patient data available for secondary uses does not contain personally identifiable information (PII) and should not allow the linkage to a particular patient without disproportionate effort, the usage of this data is bound to the specific purposes of the associated study and must therefore be protected against unauthorized access. Hence, suitable access control policies need to be implemented and enforced.

C. Distributed Access Control

A distributed architecture consistent with the health application domain is shown in Fig. 7. Here, the local security model relies on national e-health infrastructures, such as HPC and local authorization systems to obtain a security claim from the local STS. This claim can be used to access the distributed services residing in a distinct security domain or even in a distinct country.

The attributes contained within the security claims have a system-wide scope, since the claims are accepted as a valid measure for user authentication and the attributes are recognized throughout the system and are used in the attribute-based access control to resources. The end users' attributes currently consist of their role inside @neurIST (e.g., clinician, researcher, nurse, etc.). The authentication and authorization process includes the following steps at the client-side (locally):

1) request a token from the local STS;
2) incorporate the token into the service request;
3) call the service.

On the service-side, the following steps are performed:

1) call the remote STS for @neurIST token verification;
2) verify the service request;
3) check the authorizations;
4) invoke the service.

The client requests a security claim from the local STS using mechanisms defined in the WS-Trust [50] standard specification. This request contains the originator's username and the desired relationship for the claim. The STS retrieves, which attributes the user has been assigned (for the given relationship) and issues an Security Assertion Markup Language (SAML) [51] attribute assertion containing these attributes and the user's certificate. The certificate is the one used by the originator inside the respective institution for local access control. Finally, the whole token is signed by the STS.

D. Privacy in @neurIST

The privacy of patient data is ensured by enforcing that 1) data used for research purposes is pseudonymized by removing PII from the medical data required for the study; 2) filtering checks on the data leaving the data services are performed; and 3) only authorized personnel are allowed to access the @neurIST system and can only perform tasks consistent with their role.

The pseudonym generator developed for @neurIST generates reversible pseudonyms from patient ID, as well as reversible study and series unique identifiers (UIDs) for DICOM files. Two implementations of the pseudonym generator are available for the BioIS: a stand-alone application, which can be executed from the command line for the ANO model and a Web Service implementation for the OTF model. A feature of the Web Service implementation is its ability to provide universal pseudonyms (avoiding possibility of subjects from different sites receiving the same pseudonym) on the basis of a locally generated intermediate pseudonym, such that the transmission of easily traceable PII to the Web Service is avoided. A DICOM anonymizer kit is provided for deidentifying DICOM files: it removes all identifying information in the DICOM headers and it hides the face of the patient when present. The anonymizer kit can be used either via a Web Service interface, via the command-line or via a GUI.

VII. @NEURIST SYSTEM IN ACTION

At the time of writing this paper, an operational test bed of the @neurIST system has been set up comprising a variety of services at different sites in Europe, as shown in Fig. 8.

At the clinical partner sites in Barcelona, Geneva, Oxford, Rotterdam, and Sheffield, patient data (CRIM) and images are made available via a corresponding @neuInfo data access service (DAS-CRIM) and an AIS. In order to provide an integrated view of all patient data, a data mediation service (DMS-CRIM) is provided at the University of Vienna. Access to SNP data derived by the @neuLink application suite is provided by corresponding @neuInfo services in Barcelona, Bonn, Paris, and Vienna (DAS-LINK, DAS-LINKSNP, DAS-SNP, and DMS-

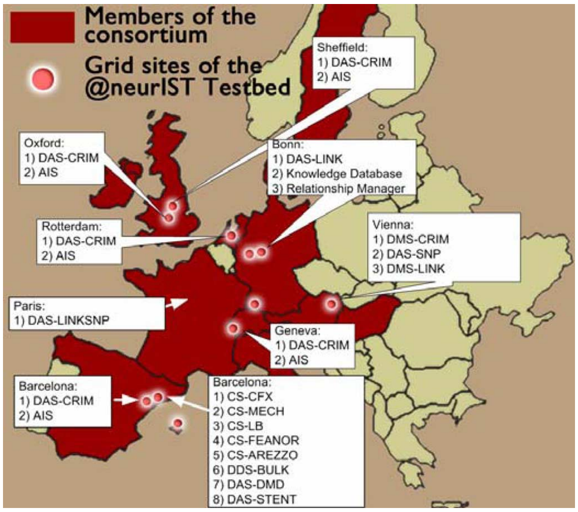

Fig. 8. @neurIST test bed.

LINK). Current @neuCompute services are hosted at Barcelona, including CS-CFX for CFD simulations, CS-FEANOR for structural analysis of vessel walls, and CS-AREZZO for processing clinical guidelines for @neuRisk. Barcelona also hosts @neuInfo services for managing derived data (DDS-BULK), for derived metadata (DAS-DMD), and for stent geometry data (DAS-STENT), which together form the DDS. Based on these services, a number of different workflows have been realized, covering a broad range of use case scenarios within @neurIST. In the following, two of these workflows are described in more detail.

A. *Integrated Risk Assessment Workflow*

@neuRisk is used in combination with @neuCompute to show how personalized guidelines can be distributed for massively parallel execution of individualized risk assessments. The workflow demonstrates how the various parameters computed by @neurIST tools plus general risk factors identified by @neuLink can be integrated in a single user interface. Clinicians use @neuRisk for analyzing rupture risk on a per-patient basis and for evaluating the pros and cons of several treatment options, including no treatment. @neuRisk uses @neuLink results in the form of decision rules and associated evidence. It connects to @neuInfo to retrieve indexes and other distributed patient-related data, such as clinical history and findings, family history, genomics profile, etc.

The integrative risk assessment workflow comprises the following interactions, as shown in Fig. 9. After the user has been authenticated (1) with the @neuRisk application suite, data about a single patient, possibly coming from different sources, is requested via an @neuInfo DMS (2) that integrates CRIM patient data from all involved clinical centers (DMS-CRIM). The mediation service fetches the required data by accessing the individual CRIM data access services (DAS-CRIM) (3–5) and (6) returns a consistent, consolidated virtual medical record (VMR). (7) @neuRisk submits the retrieved VMR and guideline to CS-AREZZO compute service for processing. (8) Additional data items, e.g., from new examinations, or secondary analysis, and an updated guideline are returned by CS-AREZZO and (9) @neuRisk presents the list of data items to the user. (10) The user provides values for the new items, which are submitted to CS-AREZZO for processing (11). A personalized result is produced and presented to the user (12, 13).

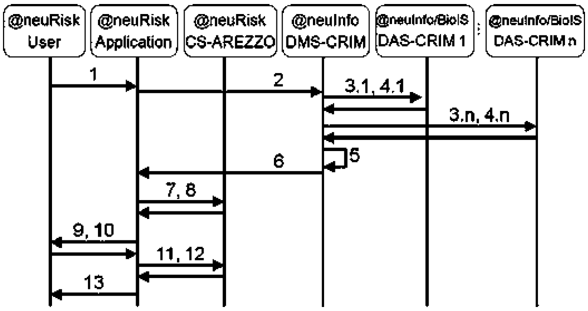

Fig. 9. Integrated risk assessment workflow.

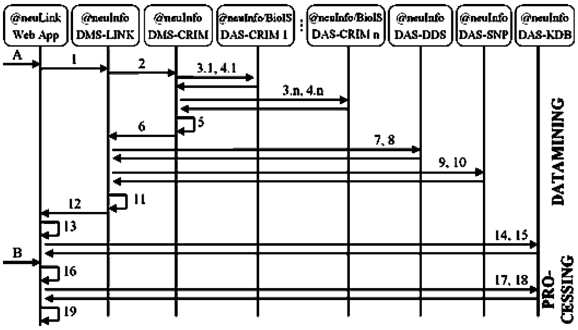

Fig. 10. Knowledge discovery with @neuLink.

B. *Knowledge Discovery With @neuLink*

This second workflow demonstrates how risk factors might be revealed by analyzing data from heterogeneous data sources. The @neuLink application suite is used to analyze structured and unstructured data from the @neurIST databases. It uses @neuInfo services to gather a data-mining matrix (consisting of CRIM data, DDS data, and SNP data) and presents results of an association analysis. This workflow helps to understand the relationships between factors held in heterogeneous databases. The workflow consists of a data-mining part and a processing part, as shown in Fig. 10.

The data-mining part (A) is an automated process (started on a regular base, e.g., once a week) in order to generate a data-mining matrix from structured and unstructured data gained from patient data via a DMS-CRIM @neuInfo service, SNP-data and derived data, accessed via the DAS-SNP, and DAS-DDS @neuInfo services, respectively (steps 1–12). The results of the data-mining process are stored via the DAS-KDB service in the @neuLink knowledge database (14, 15).

The processing step (B) is initiated by a user through the Web-portal of the @neuLink application suite. The user accesses and visualizes aggregated data from the knowledge database and potentially obtains new or confirms existing risk factors through this aggregated data (16–19).

TABLE I
PERFORMANCE RESULTS

Service	Model	Patient records	Time(sec)
DAS-CRIM-USFD	ANO	142	293
DAS-CRIM-PECS	ANO	162	259
DAS-CRIM-MIEMC	ANO	144	319
DAS-CRIM-UOXF	ANO	242	191
DAS-CRIM-HCPB	ANO	236	125
DAS-CRIM-UNIGE	OTF	435	1451
DMS-CRIM-UNIWIEN		1361	1864

C. *Performance Results*

A performance evaluation of data services for clinical data (i.e., CRIM data) is shown in Table I. Table I shows the time required to retrieve all data for all patients from different hospital databases (containing 90 tables each) using individual DAS as well as via a single DMS (last row). As can be seen, accessing the data via the OTF model takes significantly more time than accessing data services that are based on the ANO model. This is due to the fact that for OTF 1) access to the underlying CIS is required and 2) complex processing for pseudonymization and normalization is required at access time for each patient record, while for the ANO model this is done offline. Retrieving all data concurrently via the DMS is significantly faster than individual accesses combined, and mainly dominated by the time required for accessing the slower ANO service. In typical query scenarios, where only a few tables are accessed, the time for accesses via the DMS is in the range of a few seconds, which is usually acceptable even in interactive mode.

D. *Lessons Learned*

As a multidisciplinary research effort, @neurIST successfully faced the challenge of bringing together people from diverse scientific domains. An incremental approach to overall system development, as pursued by @neurIST, was well adapted to the time and effort required to reach a common understanding amongst stakeholders and helped in coping with the complexity of the system. Equally, adopting the SOA framework that allowed developers of end-user applications the mandatory shielding from the details of the underlying distributed IT infrastructure and resources to focus on their applications. Nonetheless, there are points, where early misunderstandings have had knock on effects. The potential of the ontology to play a central role in the system, for example, was one particularly difficult area for nonontology experts to grasp and fully bring to fruition, thus slowing progress. Another challenge in developing the ontology was finding an appropriate granularity and level of detail, in particular for supporting annotation and integration of heterogeneous data sources. The incorporation of the ANO data model for CRIM data in the @neurIST architecture was a response to the policy decisions in several of the clinical centers, which prohibited the level of connectivity required for the originally envisaged OTF-only architecture. Despite the minor shortcomings, the end architecture satisfies its design aims as demonstrated by the functioning of data and compute service provision to the end-user tools.

VIII. CONCLUSION

The @neurIST system provides an advanced service-oriented IT infrastructure that supports seamless access to distributed medical data and computational resources in an easy to use and secure way. The @neurIST system constitutes both an integrative decision support system to help in diagnosis, risk assessment, and treatment planning by integrating complex and multifactorial information, as well as an *in silico* research system for investigating and understanding the underlying causes of disease. Although @neurIST focuses on analysis and treatment of cerebral aneurysms, the SOA, incorporating appropriate security, privacy, and trust management components, is applicable to other diseases and makes it an appropriate architecture for future analysis and decision support systems in a clinical context.

Besides significant contributions to the understanding of aneurysms, which are beyond the scope of this paper, the @neurIST project has designed and implemented a novel SOA that supports the seamless integration of a variety of data access and integration services with computational services for advanced simulation and modeling tasks. The project has developed an ontology that formally captures the conceptual space of @neurIST required for a better understanding and management of intracranial aneurysms and subarachnoid hemorrhages, and serving as the basis for semantic annotation and integration of heterogeneous data sources.

Future work will aim at providing parts of the @neurIST service infrastructure via cloud computing technologies and at further enhancing the mechanisms for semantic data mediation.

REFERENCES

[1] M. Tsiknakis, M. Brochhausen, J. Nabrzyski, J. Pucacki, S. G. Sfakianakis, G. Potamias, C. Desmedt, and D. Kafetzopoulos, "A semantic grid infrastructure enabling integrated access and analysis of multilevel biomedical data in support of postgenomic clinical trials on cancer," *IEEE Trans. Inf. Technol. Biomed.*, vol. 12, no. 2, pp. 205–217, Mar. 2008.

[2] J. Freund, D. Comaniciu, Y. Ioannis, P. Liu, R. McClatchey, E. Morley-Fletcher, X. Pennec, G. Pongiglione, and X. Zhou, "Health-e-Child: An integrated biomedical platform for grid-based paediatrics," in *Studies in Health Technology and Informatics*. Amsterdam, The Netherlands: IOS Press, 2006, pp. 259–270.

[3] S. Oster, S. L. Hastings, S. Langella, T. Kurc, and J. H. Saltz, "Data management and exchange in the cancer grid," presented at the Advancing Practice, Instruction and Innovation Through Informatics (APIII), Frontiers in Oncology and Pathology Informatics, Pittsburgh, PA, 2004.

[4] Biomedical Informatics Research Network. [Online]. Available: http://www.nbirn.net/

[5] @neurIST: European initiative to integrate biomedical informatics in the management of cerebral aneurysms, Tirana, Albania. [Online]. Available: http://www.aneurist.org

[6] A. Arbona, S. Benkner, G. Engelbrecht, J. Fingberg, M. Hofmann, K. Kumpf, G. Lonsdale, and A. Woehrer, "A service-oriented grid infrastructure for biomedical data and compute services," *IEEE Trans. Nanobiosci.*, vol. 6, no. 2, pp. 136–141, Jun. 2007.

[7] H. Rajasekaran, L. L. Iacono, P. Hasselmeyer, J. Fingberg, P. Summers, S. Benkner, G. Engelbrecht, A. Arbona, A. Chiarini, C. M. Friedrich, M. Hofmann-Apitius, K. Kumpf, B. Moore, P. Bijlenga, J. Iavindrasana, H. Mueller, R. D. Hose, R. Dunlop, and A. F. Frangi, "@neurIST – Towards a system architecture for advanced disease management through integration of heterogeneous data, computing, and complex processing services," in *Proc. 21st IEEE Int. Symp. Comput.-based Med. Syst.*, 2008, pp. 361–366.

[8] S. C. Johnston, C. B. Wilson, V. V. Halbach, R. T. Higashida, C. F. Dowd, M. W. McDermott, C. B. Applebury, T. L. Farley, and D. R. Gress, "En-

dovascular and surgical treatment of unruptured cerebral aneurysms: Comparison of risks," *Ann. Neurol.*, vol. 48, no. 1, pp. 11–19, Jul. 2000.

[9] Y. B. Roos, M. G. Dijkgraaf, K. W. Albrecht, L. F. Beenen, R. J. Groen, R. J. de Haan, and M. Vermeulen, "Direct costs of modern treatment of aneurysmal subarachnoid hemorrhage in the first year after diagnosis," *Stroke*, vol. 33, no. 6, pp. 1595–1599, Jun. 2002.

[10] International Subarachnoid Aneurysm Trial (ISAT) Collaborative Group. "International Subarachnoid Aneurysm Trial (ISAT) of neurosurgical clipping versus endovascular coiling in 2143 patients with ruptured intracranial aneurysms: A randomised trial," *Lancet*, vol. 360, pp. 1267–1274, 2002.

[11] A. J. Molyneux, R. S. C. Kerr, L. M. Yu, M. Clarke, M. Sneade, J. A. Yarnold, and P. Sandercock, "International subarachnoid aneurysm trial (ISAT) of neurosurgical clipping versus endovascular coiling in 2143 patients with ruptured intracranial aneurysms: A randomised comparison of effects on survival, dependency, seizures, rebleeding, subgroups and aneurysm occlusion," *Lancet*, vol. 366, pp. 809–817, 2005.

[12] A. Camp, N. Ramzi, A. J. Molyneux, P. E. Summers, R. S. C. Kerr, M. Snead, J. A. Yarnold, J. Rischmiller, and J. V. Byrne, "Retreatment of ruptured cerebral aneurysms in patients randomized by coiling or clipping in the International Subarachnoid Aneurysm Trial (ISAT)," *Stroke*, vol. 38, no. 5, pp. 1538–1544, May 2007. [Epub 29 March, 2007].

[13] D. O. Wiebers, J. P. Whisnant, J. Huston, III, I. Meissner, R. D. Brown, Jr, D. G. Piepgras, G. S. Forbes, K. Thielen, D. Nichols, W. M. O'Fallon, J. Peacock, L. Jaege, N. F. Kassell, G. L. Kongable-Beckma, and J. C. Torner, "Unruptured intracranial aneurysms–Risk of rupture and risks of surgical intervention. International Study of Unruptured Intracranial Aneurysms Investigators," *N. Engl. J. Med.*, vol. 339, no. 24, pp. 1725–1733, Dec. 1998.

[14] Digital Imaging and Communications in Medicine (DICOM). [Online]. Available: http://medical.nema.org/

[15] C. M. Friedrich, H. Dach, T. Gattermayer, G. Engelbrecht, S. Benkner, and M. Hofmann-Apitius, "@neuLink: A service-oriented application for biomedical knowledge discovery," in *Proc. HealthGrid 2008*. Amsterdam, The Netherlands: IOS Press, pp. 165–172.

[16] R. Aragues, D. Jaeggi, and B. Oliva, "PIANA: Protein interactions and network analysis," *Bioinformatics*, vol. 22, no. 8, pp. 1015–1017, 2006.

[17] M. Berthold., (2009). KNIME (Konstanz Information Miner) Webpage: last accessed 2009-02-26 [Online]. Available: http://www.knime.org/

[18] C. M. Friedrich, M. Hofmann-Apitius, R. Dunlop, I. Chronakis, M. C. J. M. Sturkenboom, R. Risselada, B. Oliva, and F. Sanz, "Initial results on knowledge discovery and decision support for intracranial aneurysms," in *Proc. HEALTHINF 2008, INSTICC*, pp. 265–272.

[19] M. Hofmann-Apitius, J. Fluck, L. I. Furlong, O. Fornes, C. Kolarik, S. Hanser, M. Boeker, S. Schulz, F. Sanz, R. Klinger, H.-T. Mevissen, T. Gattermayer, B. Oliva, and C. M. Friedrich, "Knowledge environments representing molecular entities for the virtual physiological human," *Philos. Trans. R. Soc. A*, vol. 1878, no. 366, pp. 3091–3110, 2008.

[20] openMAF [Online]. Available: http://www.openmaf.org/maf/doc/maf-2-0-documentation/

[21] P. K. Singh, A. Marzo, S. Coley, G. Berti, P. Bijlenga, P. Lawford, M. C. Villa-Uriol, D. A. Rufenacht, K. M. McCormack, A. F. Frangi, U. J. Patel, and D. R. Hose, "The role of computational fluid dynamics in the management of unruptured intracranial aneurysms: A clinicians' view," *Comput. Intell. Neurosci.,* vol. 2009, p. 2, 2009, Article ID 760364, 1. DOI: 10.1155/2009/760364.

[22] A. Geers, I. Larrabide, A. G. Radaelli, H. Bogunovic, H. A. F. Gratama van Andel, C. B. Majoie, and A. F. Frangi, "Reproducibility of image-based computational hemodynamics in intracranial aneurysms: Comparison of CTA and 3DRA," in *Proc IEEE Int. Symp. Biomed. Imag.*, 2009, pp. 610–613.

[23] *Simulation Software version 12.0,* ANSYS, Inc., Canonsburg, PA. [Online]. Available: www.ansys.com/products

[24] M. Kim, I. Larrabide, M. C. Villa-Uriol, and A. F. Frangi, "Hemodynamic alterations of a patient-specific intracranial aneurysm induced by virtual deployment of stents in various axial orientation," in *Proc. IEEE Int. Symp. Biomed. Imag.*, 2009, pp. 1215–1218.

[25] I. Larrabide, A. G. Radaelli, and A. F. Frangi, "Fast virtual stenting with deformable meshes: Application to intracranial aneurysms," in *Proc. Med. Image Comput. Comput.-Assisted Intervention,* Lect. Notes Comput. Sci., vol. 5247, 2008, pp. 790–797.

[26] J. Fox and R. Thomson, "Decision support and disease management: A logic engineering approach," *IEEE Trans. Inf. Technol. Biomed.*, vol. 2, no. 4, pp. 217–228, Dec. 1998.

[27] J. Fox, D. Glasspool, D. Grecu, S. Modgil, M. South, and V. Patkar, "Argumentation-based inference and decision-making – A medical perspective," *IEEE Intell. Syst.*, vol. 22, no. 6, pp. 34–41, Nov./Dec. 2007.

[28] *Arezzo Technical White Paper,* InferMed Limited, London, U.K. [Online]. Available: http://www.infermed.com/assets/infermed/docs/

[29] Fura. [Online]. Available: http://fura.sourceforge.net/

[30] S. Benkner, G. Berti, G. Engelbrecht, J. Fingberg, G. Kohring, S. E. Middleton, and R. Schmidt, "GEMSS: Grid-infrastructure for medical service provision," presented at the HealthGRID, Clermont-Ferrand, France, Jan. 2004.

[31] S. Benkner, I. Brandic, G. Engelbrecht, and R. Schmidt, "VGE – A service-oriented grid environment for on-demand supercomputing," presented at the 5th IEEE/ACM Int. Workshop Grid Comput. (Grid), Pittsburgh, PA, Nov. 2004.

[32] S. Benkner, G. Engelbrecht, M. Köhler, and A. Wöhrer, "Virtualizing scientific applications and data sources as grid services," in *Cyberinfrastructure Technologies and Applications*, J. Cao, Ed. New York: Nova Science Publishers, 2009, ISBN: 978-1-60692-063-3.

[33] A. Woehrer, P. Brezany, and A. M. Tjoa, "Novel mediator architectures for Grid information systems," *J. Future Gen. Comput. Syst. – Grid Comput.: Theory, Methods Appl.*, vol. 21, pp. 107–114, Jan. 2005.

[34] M. Antonioletti, M. Atkinson, R. Baxter, A. Borley, N. P. Chue Hong, B. Collins, N. Hardman, A. C. Hume, A. Knox, M. Jackson, A. Krause, S. Laws, J. Magowan, N. W. Paton, D. Pearson, T. Sugden, P. Watson, and M. Westhead, "The design and implementation of grid database services in OGSA-DAI," *Concurr. Comput.: Pract. Exp.*, vol. 17, no. 2–4, pp. 357–376, Feb. 2005.

[35] M. N. Alpdemir, A. Mukherjee, A. Gounaris, N. W. Paton, P. Watson, A. A. A. Fernandes, and D. J. Fitzgerald, "OGSA-DQP: A service for distributed querying on the grid," in *Proc. 9th Int. Conf. Extending Database Technol., LNCS*, vol. 2992, 2004, pp. 858–861.

[36] G. Wiederhold, "Mediators in the architecture of future information systems," *IEEE Comput.*, vol. 25, no. 3, pp. 38–49, Mar. 1992.

[37] K. Kumpf, A. Woehrer, S. Benkner, G. Engelbrecht, and J. Fingberg, "A semantic mediation architecture for a clinical data grid," in *Grid Computing for Bioinformatics and Computational Biology*. New York: Wiley, 2007.

[38] M. Boeker, H. Stenzhorn, K. Kumpf, P. Bijlenga, S. Schulz, and S. Hanser, "The @neurIST ontology of intracranial aneurysms: Providing terminological services for an integrated IT infrastructure," presented at the AMIA 2007 Annu. Symp., Chicago, IL.

[39] C. Qu, F. Zimmermann, K. Kumpf, R. Kamuzinzi, V. Ledent, and R. Herzog, "Semantics-enabled service discovery framework in the SIMDAT pharma grid," *IEEE Trans. Inf. Technol. Biomed.*, vol. 12, no. 2, pp. 182–190, Mar. 2008.

[40] B. V. Aduna and Sirma AI Ltd. (2006). *User Guide for Sesame.* [Online]. Available: http://www.openrdf.org/doc/sesame/users/index.html

[41] A. Kiryakov, D. Ognyanov, and D. Manov, "OWLIM – A pragmatic semantic repository for OWL," in *Proc. Int. Workshop Scalable Semantic Web Knowl. Base Syst. (SSWS),* WISE (LNCS series), LNCS 3807. New York: Springer-Verlag, Nov. 2005, pp. 182–192.

[42] T. Mason and R. Lawrence, "Dynamic database integration in a JDBC driver," in *Proc. 7th Int. Conf. Enterprise Inf. Syst.*, 2005, pp. 326–333.

[43] Protégé. [Online]. Available: http://protege.stanford.edu

[44] Web Ontology Language – Description Logic (OWL-DL). (2004). [Online]. Available: http://www.w3.org/TR/owl-features/

[45] RacerPro. [Online]. Available: http://www.racer-systems.com

[46] Pellet. [Online]. Available: http://pellet.owldl.com

[47] Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE). [Online]. Available: http://www.loa-cnr.it/DOLCE.html

[48] Foundational Model of Anatomy (FMA). [Online]. Available: http://sig.biostr.washington.edu/projects/fm

[49] O. Bodenreider, "The Unified Medical Language System (UMLS): Integrating biomedical terminology," *Nucl. Acids Res.*, vol. 32, pp. D267–D270, 2004, [Online]. Available: http://www.nlm.nih.gov/research/umls

[50] OASIS WS-Trust Specification. [Online]. Available: http://docs.oasis-open.org/

[51] OASIS. Security Assertion Markup Language. [Online]. Available: http://www.oasis-open.org/

Author's photographs and biographies not available at the time of publication.