IST – 027703

Integrated Project of the 6th Framework Programme

aneurIST

Integrated biomedical informatics for the management of cerebral aneurysms

www.aneurist.org

Information Society Technologies

White Paper

An introduction to the @neurIST System Architecture

April, 2009

The financial support of the European Commission is gratefully acknowledged. Material in this document reflects only the author's views and the Commission is not liable for any use that may be made of the information contained herein.

Authors

Christoph Friedrich Martin Hofmann-Apitius	Institute for Algorithms and Scientific Computing (SCAI), Fraunhofer Gesellschaft
Philippe Bijlenga Jimison Iavindrasana	Geneva University Hospitals
Antonio Arbona Guillem Cantallops Magdalena Escalas	Grid Systems S.A.
Guntram Berti Jochen Fingberg Nils Gruschka Peer Hasselmeyer Luigi Lo Iacono Guy Lonsdale Hariharan Rajasekaran	IT Research Division, NEC Laboratories Europe, NEC Europe Ltd.
Paul Summers	University of Oxford
Alejandro Frangi	Center for Computational Imaging & Simulation Technologies in Biomedicine, Universitat Pompeu Fabra
Steven Wood	Department of Medical Physics, Royal Hallamshire Hospital, Sheffield
Siegfried Benkner Gerhard Engelbrecht Martin Köhler Alexander Wöhrer	Institute of Scientific Computing, Faculty of Computer Science, University of Vienna

Acknowledgement: This work was generated in the framework of the @neurIST Integrated Project, which is co-financed by the European Commission through contract no. IST-027703

Foreword

This white paper provides a high-level introduction to the architecture of the distributed IT system developed for the @neurIST project. While the @neurIST project develops solutions for supporting clinical treatment and medical research related to a specific medical condition (cerebral aneurysms), the overall system design is appropriate for exploitation in a much wider healthcare setting. The general service-oriented approach and much of the ICT infrastructure behind the specific solutions – based on the architecture described here – can be applied to other healthcare areas. Thus, the target audience for this paper is the group of people wishing to develop a strategy for the deployment or development of integrated multi-site, multi-modal decision support and data analysis tools for healthcare. Details of the system architecture are provided in a publicly available project deliverable [D05] and the technologies developed for and exploited in specific @neurIST suites are described and documented in a number of papers and project deliverables that can be located via the project's web-site [pubs].

1 The @neurIST System – a template for distributed analysis and support for healthcare

The @neurIST project was set-up in response to the recognized necessity of federating data sources to allow rapid confirmation and discovery of clinical evidence. One of the key decisions taken when designing the project was that a future production realisation of the prototype system, to be developed in the lifetime of the project, should be deployable in a clinical setting within the constraints of the healthcare systems operating in Europe. On the one-hand, this means that we are dealing more with an inter-enterprise (or business-to-business) setting than with an eScience research environment and on the other hand that we have to deal with the trust, privacy and security aspects linked to processing of patient-specific data. These aspects have been reflected in the design of the system architecture: a service-oriented approach consistent with future incorporation of commercial, supported services; data privacy and access control being central components. The prototype infrastructure implementation based on the system architecture uses WebServices standards to ensure conformity with the target of integration with commercially operated systems and services. Where appropriate, *de facto* Grid standards with broad acceptance (such as OGSA-DAI) are also employed.

The above-mentioned “enterprise focus” does not mean that the medical research (eScience) users and use cases are excluded. On the contrary, the system extends the resources that the medical researcher can access (both geographically distributed data repositories and computational services) and provides a mechanism for knowledge generated to be fed into future clinical decision support systems and services. The different layers of the @neurIST architecture are shown in Figure 1.

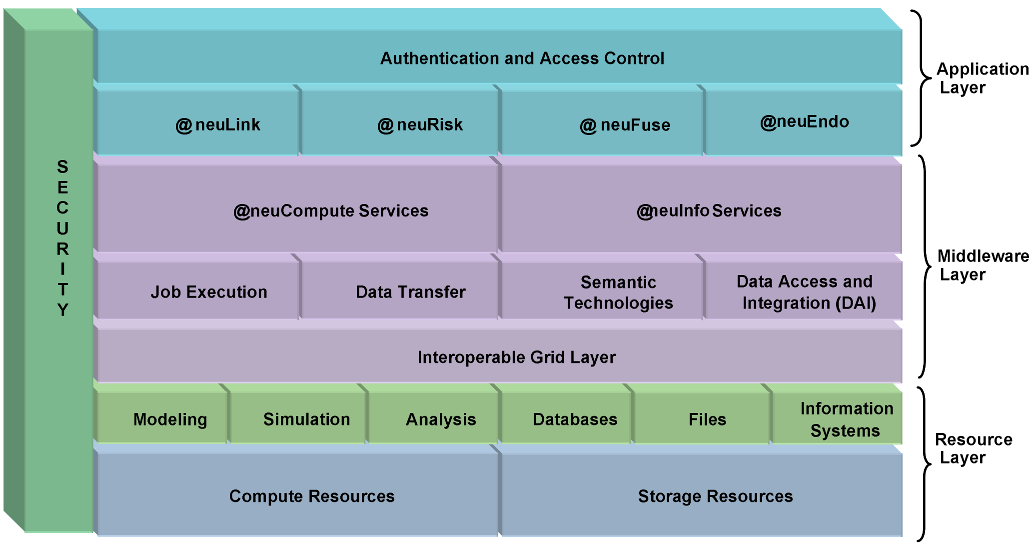

Figure 1:@neurIST layered view

- The Application Layer builds on top of the system infrastructure, to handle the users' work needs through the application suites.
  - @neuFuse handles complex analysis of the biomechanics of cerebral aneurysms for patient-specific geometries obtained from medical images, @neuEndo extends this to the analysis of endovascular treatment using stents;
  - @neuLink managing knowledge extraction via the integration of information from multiple sources and of different types – from genetics through bioinformatics analyses to biomechanical indicators;
  - @neuRisk delivering clinical decision support.

- The Middleware Layer provides the application layer with the tools and services required to access data and compute resources distributed both geographically and administratively. They hide the complexity of remote access from the user and are independent of the suites or aspects relating to specific diseases.
  - @neuInfo offers a generic data management and integration framework that supports the provision and deployment of data services.
  - @neuCompute enables service providers to virtualize High Performance Computing applications available on clusters or other parallel hardware as compute services that can be accessed on-demand by the application suites.

- Within the Resource Layer, the biomedical infostructure (BioIS) in @neurIST offers secure access to clinical databases to the geographically distributed front-end systems (suites). The data is secure since only anonymised data is contained in mirrored databases in the clinic's de-militarized zone (DMZ), or in the case of the very forward looking "on-the-fly" model, a virtual anonymised database, entries are extracted from the clincal system and anonymized at run-time.

- The realisation of the BioIS within @neurIST uses the clinical reference information model (CRIM), which is a description of all the clinically relevant patient information. The model covers aneurysm-related treatment data in great detail in order to support aneurysm research. Although the CRIM is specific to the @neurIST project, ongoing work related to the @neuRisk suite indicates that integration with suites handling the Virtual Medical Record being proposed by the HL7 standards body is feasible.

- Security: The security architecture of the system is designed to ensure that only authorised personnel gain access to the system and patient data while preserving the privacy of patients involved. The RelationshipManager provides the security infrastructure for distributed access control based on WebServices specifications including WS-Trust and SAML. Privacy preserving mechanisms such as pseudonymisation and de-personalization are used to remove personally identifiable information of the patients from the medical data made available to the users of @neurIST.

While refactoring of the application suites may, to some extent, be possible, it is foreseen that adaptation to other clinical contexts would focus on the development of specific applications with revision of the lower levels limited to accommodation of newly encountered resources, and specification of the CRIM for that context.

2 @neurIST System Highlights

2.1 Data Mediation

The @neurIST system deals with medical data spread across geographically distributed sites and hosted at different institutions each having their specific infrastructure to deal with such data. In this scenario is important to hide the details of distributed data sources and resolve the heterogeneities with respect to access language, data model and schema. The @neuInfo data services of the @neurIST middleware virtualize heterogeneous information sources and enable transparent access to and integration of relational databases, XML databases and flat files.

As preserving the autonomy of data providers and ensuring access to live data are key requirements, data integration is based on a mediator approach wherein local data sources are integrated bottom-up by mapping local data base schemata into a virtual global schema. Queries against the virtual global schema of a data mediation service are translated on-the-fly into local queries and result data from local sources is automatically integrated on its way back to the user. The provision of virtual global schemata and the corresponding mappings between global and local views is facilitated through semantic annotation of local data schemata according to the @neurIST ontology. Complex data integration scenarios may be optimized based on distributed query processing facilities.

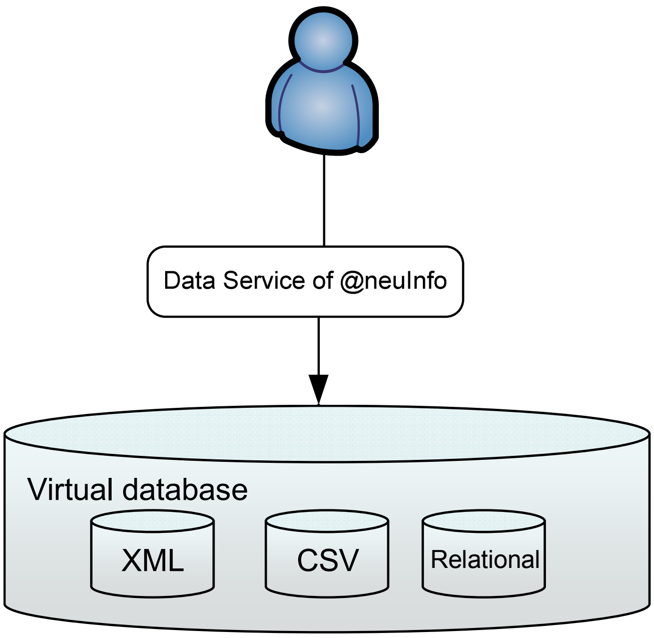

Figure 2: @neurIST Data Mediation Service

2.2 CRIM & BioIS

Clinical data related to aneurysms must be made available to the clinical researchers in @neurIST to help them in their research studies. The CRIM is a specification of the subset of the patient clinical data needed for aneurysm-specific research and clinical care. The BioIS makes this clinical data available in a pseudonymized (i.e. identifying attributes have been removed to protect patient privacy) and normalized form for research purposes. The data is pseudonymized to ensure that a patient can be tracked down in case a follow-up is required based on research results.

The BioIS uses two architecture models to deliver this research data. In the anonymized model (ANO), the clinical data is anonymized (to the degree of de-identification and pseudonymization) in bulk and hosted in a demilitarized zone (DMZ) of the clinical center. Regular updates are carried out to keep the anonymized data sources synchronised with the data in the clinical databases. In the on-the-fly (OTF) model, the data present in the clinical information systems are anonymized on-the-fly when they are requested. This eliminates the updating process required for the ANO model. The ANO model is a stable and pragmatic approach for setting up a patient data sharing system that respects user privacy while the OTF model is an advanced prototype for the use of depersonalization technologies on live clinical systems. The lessons learned in the OTF model will be used for data sharing schemes of the future. Figure 3 shows the two different BioIS architectures.

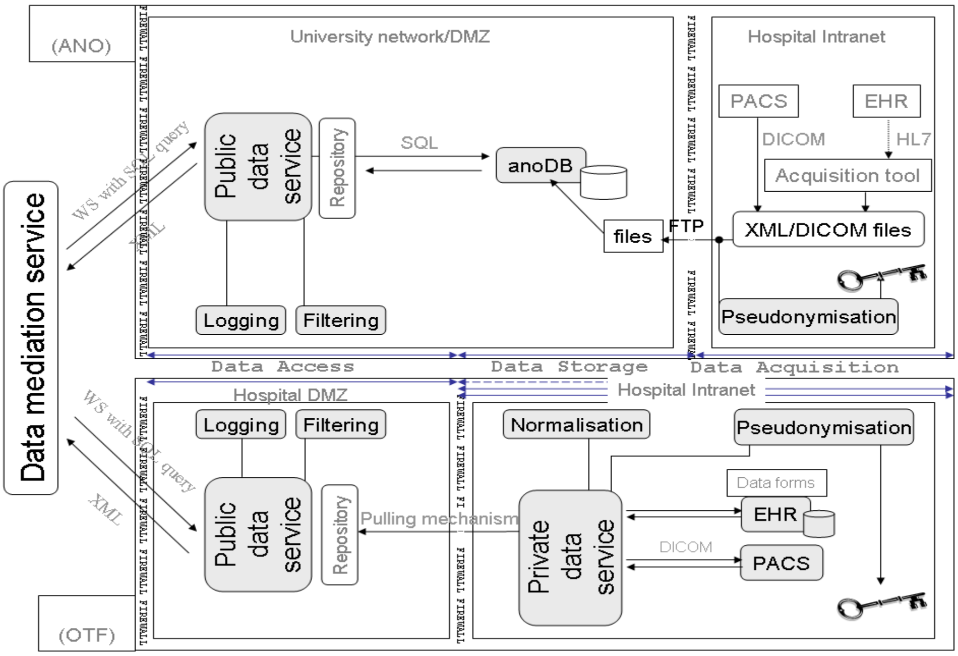

Figure 3: BioIS architectures

Were the target institutions looking to adopt the @neurIST architecture within a clinical context (for example, within a national health service), the OTF model could be adapted to forego the pseudonymization process. Such a strategy would allow all elements of a patient's clinical records to be accessed from within the network regardless of which specific institution holds the individual parts.

2.3 RelationshipManager

The @neurIST security architecture allows cross-Enterprise interaction and data exchange where it is possible for a user from one security domain to access resources hosted in another domain that operates a different security infrastructure.

To provide the required flexibility and manageability, the security architecture does not rely on any centralized component. Local and remote security management is separated removing the need for a system-wide harmonization of local identification and authentication policies and schemes. @neurIST follows a hybrid security model which is a combination of a local model and a distributed model. Within a security domain all the security is concentrated and placed under the responsibility of this domain whereas between different security domains, local credentials are mapped to inter-domain credentials that can be exchanged. The inter-domain credentials are issued by a Security Token Service (STS) located at each of the participating sites.

To allow the authorized entities within each institution to handle these access control functionalities easily, the RelationshipManager has been developed (Figure 4). It incorporates the existing identity infrastructure on the local site and combines it with the STS to issue or validate @neurIST specific credentials.

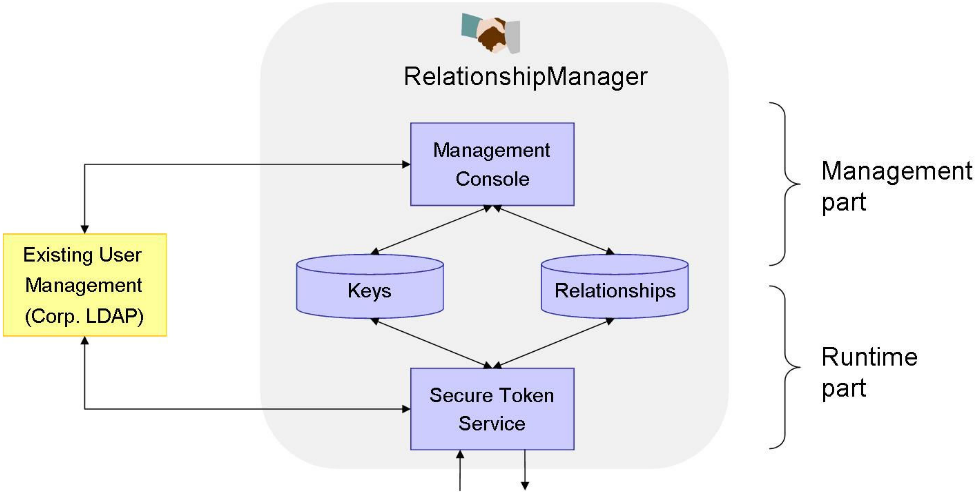

Figure 4: RelationshipManager Architecture

3 The @neurIST Workflows – combining technologies for specific use scenarios

The impact and effectiveness of the @neurIST architecture as realised within the project is demonstrated by five different workflows that capture the different aspects in which the @neurIST infrastructure is put to use in the treatment and study of aneurysms. Each of the short descriptions below will include an illustration of how the (mainly generic) @neurIST components are combined to realise a solution for the workflow.

Some key developments linked to the computational analysis services and the results produced by those services are components used across the workflows: automation of complex processing from patient's image data via the use of an abstract problem definition; the use of a standardised data model for the derived data; the integration of such derived data with other information for @neurIST's concept of a virtual patient metaphor.

3.1 Workflow 1: Integrative Risk Assessment Using Distributed Guidelines

This workflow deals with the clinical decision support feature of the @neurIST infrastructure. The @neuRisk application suite is used by clinicians to assess rupture risk on a per-patient basis and for evaluating the pros and cons of available treatment options, including that of no treatment. @neuRisk uses @neuLink results in the form of decision rules and associated evidence. It connects to @neuInfo data services to retrieve indexes and other patient-related data, such as clinical history and findings, family history, genomics profile, etc. The computational services offered via @neuCompute are used to compute risk profiles.

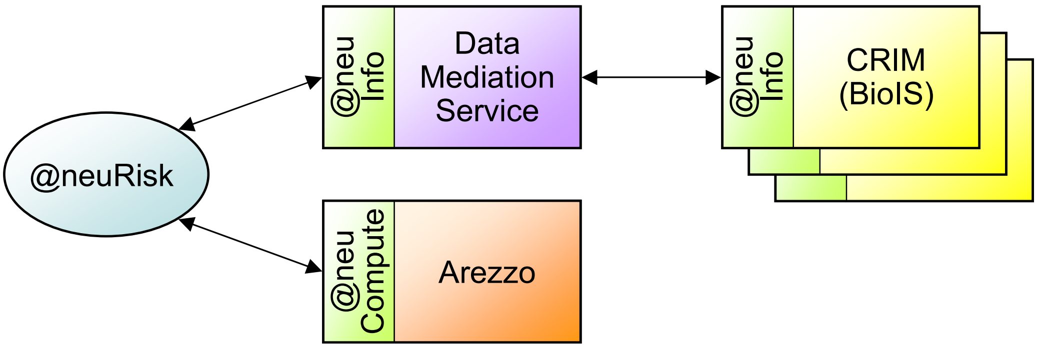

Figure 5:Realisation of Workflow 1 in the @neurIST System Prototype

3.2 Workflow 2: Knowledge Discovery from Unstructured and Structured Data

This workflow shows how @neuLink is used by researchers in the study of cerebral aneurysms. @neuLink uses datamining techniques to gather aneurysm related knowledge from patient data (CRIM), genetic data, derived data from analysing patient data and medical publications from repositories. Risk factors for individual patients can be obtained by comparing their profiles against a knowledge database derived from the data analysed by @neuLink.

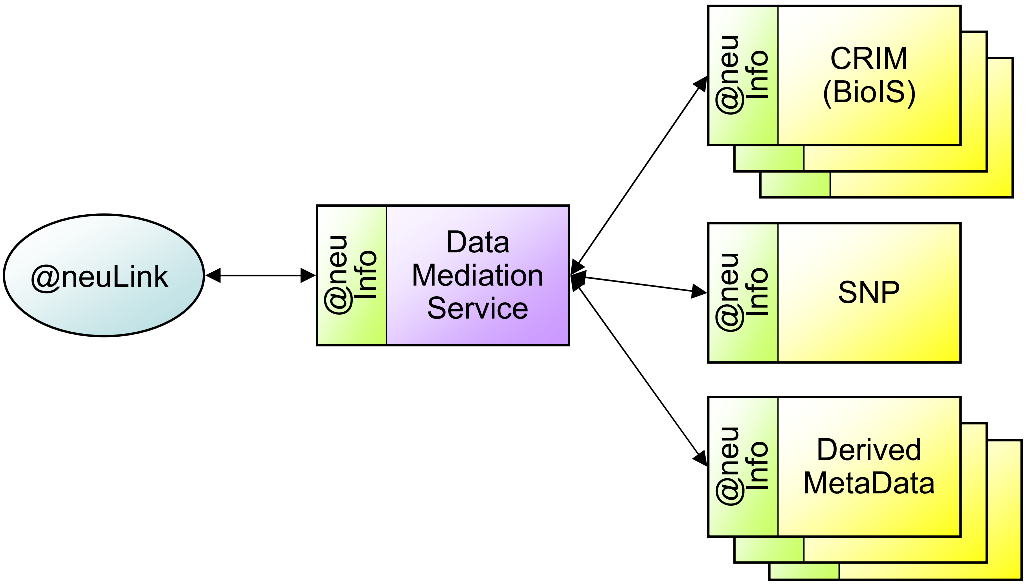

Figure 6:Realisation of Workflow 2 in the @neurIST System Prototype

3.3 Workflow 3: Case-specific Shape Analysis, Structural and Haemodynamic Simulation from Medical Image Data (Single Case)

This workflow's objectives are two-fold. One is to populate the knowledge databases of shape and mechanical indexes derived from patients in the study. These indexes are used by workflow 2 (knowledge discovery) to derive risk factors for @neuRisk. Second, in the clinical context, it serves to derive the shape and mechanical parameters needed as input to the *individual* risk assessment using @neuRisk and the risk factors covered bz the CRIM described in the first workflow. This workflow makes use of the @neuFuse application suite together with the services provided by @neuInfo and @neuCompute to perform and visualize complex shape, flow or wall stress analysis of blood vessels to determine the risk of their rupture.

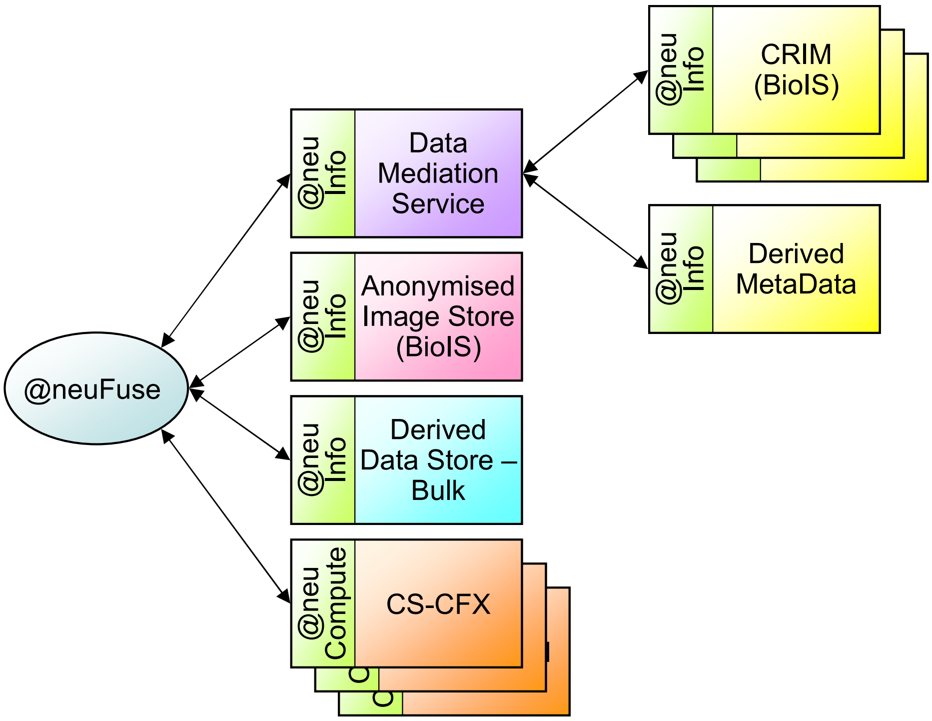

Figure 7:Realisation of Workflow 3 in the @neurIST System Prototype

3.4 Workflow 4: Virtual Stenting

One method of treating aneurysms involves the insertion of tubes called stents to prevent aneurysms from rupturing. This workflow serves to simulate mechanical impacts of stenting. There are two main variants of the workflow: First, to try out *in silico* several possible treatment options for a specific aneurysm using flow simulation *(clinical case)*. This is used to assess the impact of placing a particular stent to treat an aneurysm of an individual patient. The second variant is to test a specific stent design in detail against different real, typical aneurysm geometries *(commercial case)*. This is useful for commercial stent manufacturers to optimize their stent designs.

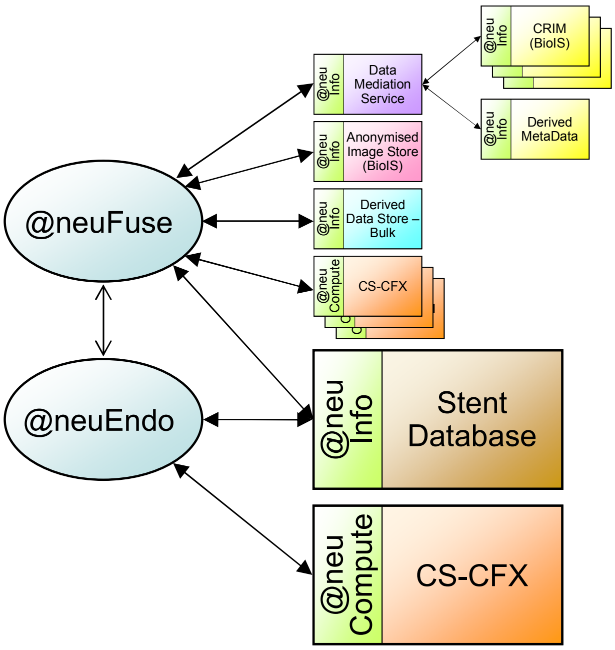

Figure 8:Realisation of Workflow 4 in the @neurIST System Prototype

3.5 Workflow 5: Analyses of shape and mechanical indexes of aneurysm geometries obtained from medical images (Multi Case)

This workflow can be seen as a generalization of workflow 3. Instead of working on a single aneurysm case as in workflow 3, here a number of related cases are studied and simulated in a “batch” mode. This is expected to be useful in populating a knowledge database with data derived on haemodynics in realistic cases. Although there are a number of other potential use cases for this functionality, this workflow is used in @neurIST for a virtual patient metaphor inspired scenario in which the simulation of many cases is automated. For each case, two blood flow simulations are to be performed: one with a personalized (VPM-derived) 1D model and one with standard parameters. Results of the simulations will be analysed by data mining tools to evaluate the influence of the VPM on simulation results.

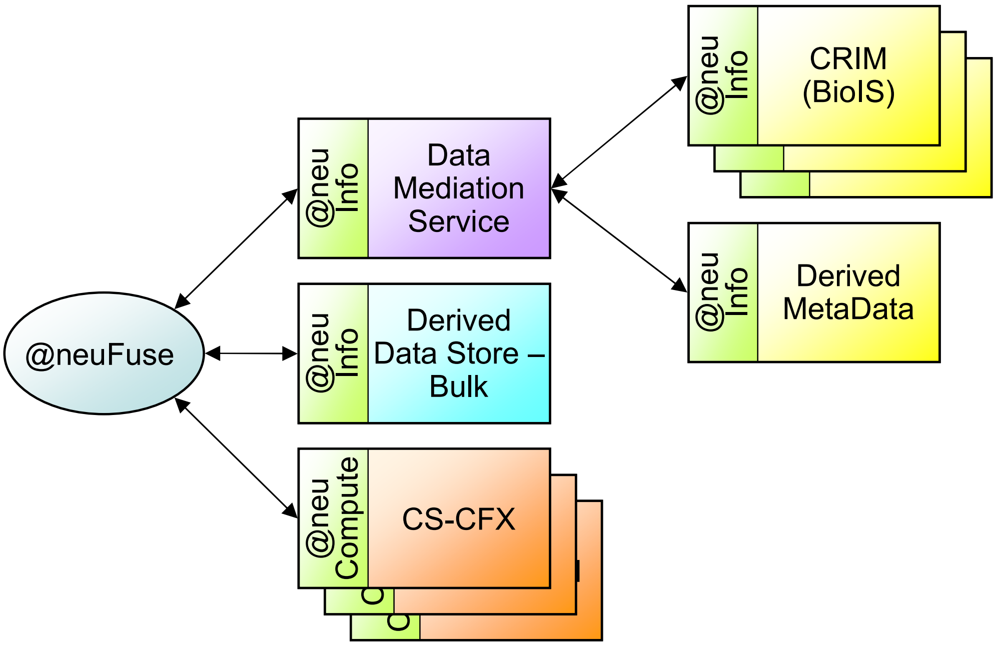

Figure 9:Realisation of Workflow 5 in the @neurIST System Prototype

4 Concluding Remarks

The @neurIST system architecture is a general approach for integrating data and compute services distributed between clinical and medical research institutions and at service providers. By design, the architecture is appropriate for and adaptable to a broad range of medical and clinical scenarios; the realisation for research and treatment for cerebral aneurysms created within the @neurIST project does not impose any limitations on the system architecture.

The service-oriented approach, incorporating appropriate security, privacy and trust management components, makes it an appropriate architecture for the development of future analysis and (decision) support systems to be deployed in a clinical context. It is important to emphasise that, as was seen in the introduction to the major system components provided by this document, the system architecture and realisations of it created in the @neurIST project are independent of the specific medical focus (analysis and treatment of cerebral aneurysms) being addressed within the @neurIST project. This means that the architecture is extendable in a natural way to other medical sectors and disease areas (or indeed to other distributed enterprise ICT system requirements).

Although the system design as such excludes management of data collection and corresponding database creation, the suites currently developed and deployed within the project will be augmented by database browser facilities building on the same distributed access infrastructure. Future operational systems would clearly benefit from integration with software suites and service systems to support data collection and management of both clinical trials and medical research experiments.

Such systems will exploit the current development of electronic health records (EHR) being progressively deployed across healthcare systems. While the current @neurIST system takes information extracted from electronic health records and federates it for the benefit of one research consortium, ultimately the vision of the @neurIST system would be global harmonization and integration of the activities of multiple research consortia or collaborating health services.

From the early stages, the duration of the project was clearly identified as being too short to allow for such global extension and the high-level clinical trials and medical research experiments management tools was similarly deemed out of scope. However, there are a number of EHR systems already commercially available and global clinical trials registries are being pushed forward by national institutions such as the National Institute for Health in the USA [NIHreg] and the World Health Organisation [ICTRP]. Thus, the building blocks are there to follow the first steps taken by @neurIST towards integrated, geographically distributed EHR and extended database systems to empower clinical and medical research through optimal use of resources and knowledge sharing.

5 References

[D05] @neurIST System Architecture, Public deliverable. Available at: http://www.cilab.upf.edu/UserFiles/File/PUBLIC_DELIVERABLES/d05_v2_public.pdf

[ICTRP] International clinical trials registry platform, http://www.who.int/ictrp/

[NIHreg] www.clinicaltrials.gov

[pubs] @neurIST Web-Page, Section "Publications" http://www.cilab.upf.edu/aneurist1/index.php?option=com_content&task=view&id=45&Itemid=64