Virtual Physiological Human network of excellence

Virtual Physiological Human Network of Excellence

Grant Agreement: 223920

Deliverable D3.6

VPH ToolKit Development

Work Package 3

PM23 (Apr-2010)

Version 1.1 30-Apr-10

SEVENTH FRAMEWORK PROGRAMME

European Commission
Information Society and Media

This page is intentionally blank

Document Information

IST Project Num	FP7 – 2007 - ICT - 223920	**Acronym**	VPH NoE
Full title	Virtual Physiological Human Network of Excellence		
Project URL	http://www.vph-noe.eu		
EU Project Officer	Joël Bacquet		

Deliverable	**Number**	3.6	**Title**	VPH ToolKit Development
Work package	**Number**	3	**Title**	VPH NoE ToolKit Development

Date of delivery	**Contractual**	PM23	**Actual**	PM23
Status	Version. 1.1		Final ☑	
Nature	Prototype ☐ Report ☑ Dissemination ☐ Other ☐			
Dissemination Level	Public ☐ Consortium ☑			

Authors (Partner)	UCL	Coveney, Diaz, Mendes, Zasada, Haidar		
	UOXF	Cooper, Fletcher, Gavaghan, Kohl, McCormack		
	CNRS	Thomas, Friboulet, Cervenansky		
	INRIA	Sermesant, Bleuzé		
	UPF	Larrabide, Martelli, Omedas		
	UOA	Hunter, Britten, Christie, Lloyd, Brooks		
	EBI	de Bono		
	USFD	Fenner, Lawford, Wood		
	IMIM	De Fabritiis, Sanz, Villa-Freixa, Giorgino		
Responsible Author	Keith McCormack		**Email**	K.M.McCormack@sheffield.ac.uk
	Partner	UOXF	**Phone**	+44 114 236 7522

Abstract (for dissemination)	This document updates the information on VPH NoE ToolKit strategy and implementation during Year 2 of the Network's operation
Keywords	VPH NoE Model Tool Data Simulation Library Standards Ontology

Version Log

Issue Date	Version	Author	Change
28-Mar-10	0.1	WP3	Skeleton
30-Mar-10	0.2, 0.3, 0.4	WP3	Some flesh added - draft to WP3 partners
17-Apr-10	0.5, 0.6, 0.7	WP3	Most sections populated
18-Apr-10	0.8, 0.9	WP3	Further population
19-Apr-10	1.0	WP3	Release to internal review
30-Apr-10	1.1	WP3	Final

This page is intentionally blank

Table of Contents

This page is intentionally blank

EXECUTIVE SUMMARY

This document describes the evolving approach to the provision of a central repository for the Tools, Models and Data that are to be made available to the VPH community.

The report begins by discussing the developments that have taken place during the second year of ToolKit development, examining progress in constructing the ToolKit Portal and equipping it with content. The extended activities of partners, in imaging, data, models, Grid, annotation and ethico-legal matters, are also reported.

The ways in which the approach to content provision is adapting to the identified needs of users is considered, taking into account the need for sustainability beyond the life of the NoE. To this end a process of standardisation is being introduced, intended to ensure high-quality ToolKit contributions and involving the development and publication of formalised guidelines for submissions, adherence to which will ensure that content can provide maximum utility to other users. Working Groups drawn from WP3 partners are being formed to further this approach.

In a further major strand of development, also involving the creation of Working Groups, the need for an integrated approach to the collection and curation of candidate ToolKit content, from the Exemplar projects, the VPH-I activities and from external sources is discussed. Here it is recognised that in many cases a process of adjustment may be need, working with content providers to ensure that the proposed content is aligned with the developing standards.

Licensing restrictions are a significant issue for ToolKit users and a specific strand of activity is addressing this issue. Training and Dissemination are equally important activities if the reach of ToolKit content is to be maximised.

Finally the sources of ToolKit content are reviewed in detail, with Exemplars, VPH-I projects and external sources considered separately.

Introduction

The VPH ToolKit concept is evolving, as NoE members begin to appreciate both the scale of the task and the issues that must be addressed for a sustainable future. The ToolKit is certainly not intended merely to be a collection of isolated tools used by VPH researchers and clinicians, nor can it be a single monolithic entity capable of fulfilling the needs of all users; the latter is impossible, and the former has little impact or utility.

In contrast, the ToolKit must evolve into a curated set of tools that are sufficiently usable, flexible, and interoperable (a key point brought up at the first EC review) that they can be configured and connected by researchers to provide a range of “shrink-wrapped” solutions for clinicians, instead of starting from scratch for each new use-case. This goal imposes requirements on ToolKit development (especially if the ToolKit is to have any lasting impact), necessitating that such activities should become self-sustaining and continue uninterrupted at the end of the VPH-NoE’s lifetime. It also requires involvement of the wider VPH community, since the VPH-NoE does not have the resources to achieve this vision by itself. Both communities must work together to promote development practices that can achieve mutually beneficial goals.

In this update to the ToolKit description, the emerging concepts of sustainability and the need for underlying content management are examined, and the evolving structure of activities and working methods to achieve these results are described. This document outlines early progress towards sustainability, centred on ToolKit development and contribution as a result of community interaction. The first section considers the work of WP3. Subsequent sections feature the Seed Exemplar and VPH-I projects, with the document completed by a critical analysis of ToolKit sustainability to date. Particular developments include the formation of Working Groups to address the need for documentation to support the ToolKit approach to standardisation, and the requirement to work with content providers to ensure optimum content quality.

Finally, just as the evolution of the ToolKit requires the full involvement of the VPH community to shape and populate the structure, so it must operate in an environment of secure and controlled access. A future priority is that WP3 must interact with the consortium that will be awarded the opportunity to construct the VPH infrastructure as a result of the sixth Call for proposals.

Help with this Document

This Deliverable is not structured as a linear document, rather it is an information resource reflecting developments of the VPH NoE ToolKit. The diagram below shows the structure of the document and may help the reader's navigation.

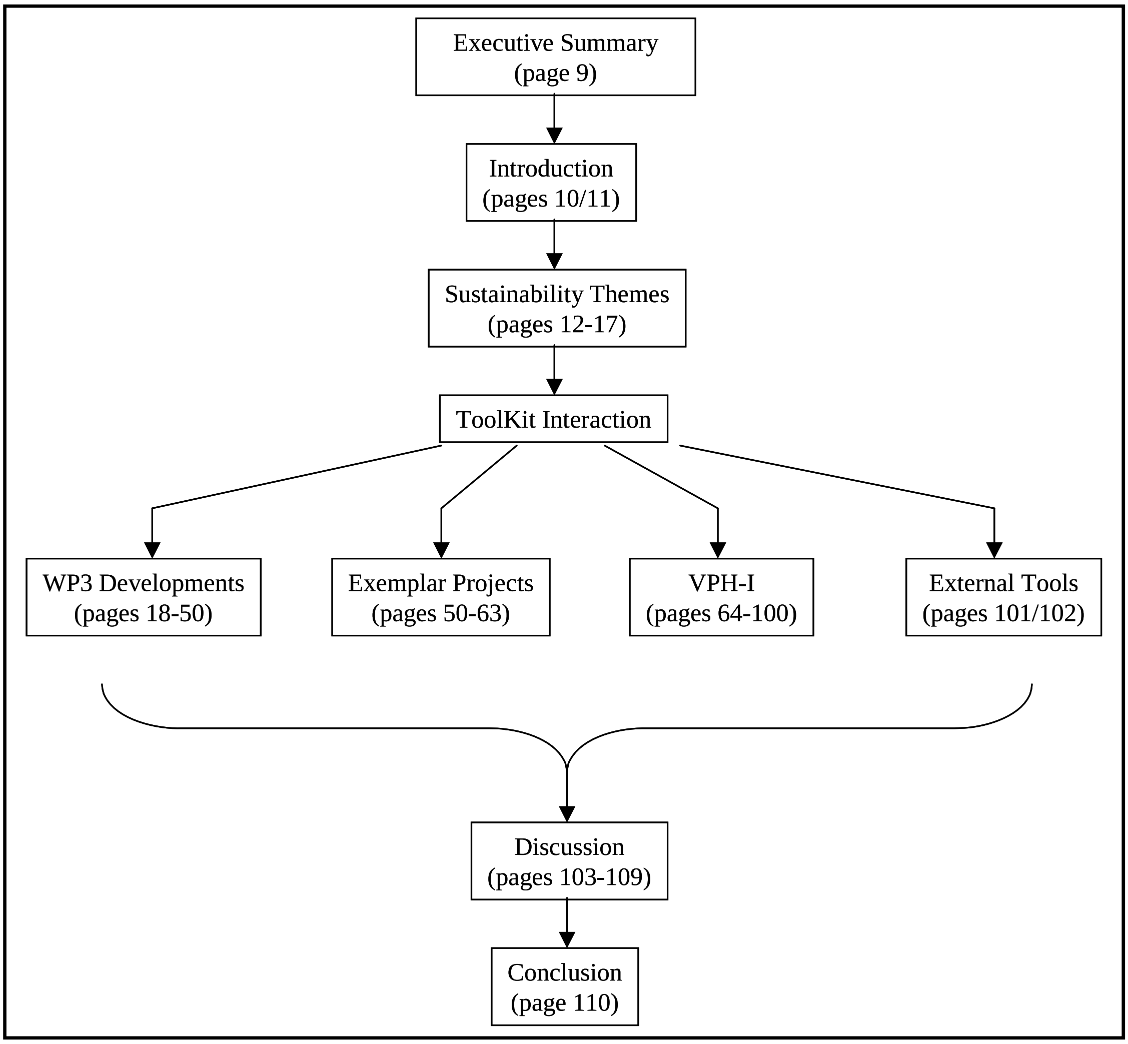


Figure 1. Guide to Document Structure

ToolKit Evolution

The first year of the NoE established the ToolKit concept and enabled it to be realised as a functioning repository, supported by a portal for interaction. Subsequent months have established the principle of long-term sustainability as central to the role of WP3 and identified the need for several strands of activity. These undertakings are implicit to the original Description of Work, but are reformulated below in the context of sustainability. Three principal strands have emerged, interoperability, engagement and standards.

- Interoperable content: High-quality content for the ToolKit needs to be collected. Wherever possible, this should be sought from existing developments external to WP3, in order both to maximise engagement and make best use of NoE resources. WP3 should not be reinventing any wheels, and so where the NoE is currently funding development that could be achieved from other sources, a redirection of effort should be considered. The focus should be on making existing tools interoperable with other ToolKit content.
- Engagement: The EPs (WP2) and the VPH-I are natural sources for ToolKit content, as well as being our primary users. Support is needed (and is being requested) from WP3 in order for these groups both to contribute to, and benefit from, the ToolKit. Engagement with these partners is a high priority, to assist both in identifying useful outputs from these projects, and increasing their interoperability. A major goal for WP3 is to promote ToolKit component interoperability, and obtain backing from relevant parties, willing to invest effort to achieve this. Interaction with WP4 can enhance this, using training workshops to develop ToolKit use further.
- Standards: The use and development of standards is of primary importance in achieving interoperability within the ToolKit, and again promoting this idea within the VPH-I, and further afield, is a key goal of WP3. This activity has many facets: identifying crucial standards that already exist, improving tool support for such standards, identifying gaps in provision and working with interested parties to develop new standards, and working with WP4 to provide training in the use of standards.

The foci of WP3 have therefore become standards development, ToolKit component integration and community engagement. Of course other work will continue, notably on ToolKit Portal development (Section 4.8) and training-based interaction with WP4 (Section 3.4). Each WP3 partner is committed to allocating effort to standards and interoperability through engagement with the community. Interaction with ToolKit content providers is expected to expose rich content from multiple sources, including resources that extend beyond the Exemplars and the VPH-I projects.

Sustainability Themes

Sustainability does not happen without careful consideration and planning. For software to be long-lasting, it must satisfy a range of criteria, most important amongst which is the need for a strong user and developer community to ensure its long-term support and relevance. Activities to strengthen community interaction have been undertaken at various levels. In particular, delegates from workpackages 2 and 3 of the VPH-NoE, and from almost all of the current VPH-I projects, met in Brussels in early 2010 for a detailed technical workshop to consolidate sustainability strategies for the VPH. Discussions drew out five key sustainability criteria for ToolKit content.

- Standards – these are crucial to enable tools to interoperate, and for researchers to be able to share their data and models.
- Ontologies – a sub-theme of standards, but sufficiently important to merit its own focus. Ontological annotation of models and data facilitates interconnection of ToolKit resources.
- Certification – ToolKit content needs to be demonstrably of a high standard, well documented, and properly validated.
- Interoperability & workflows – tools should not exist in isolation, but be usable in tandem to achieve clinical goals.
- User friendliness – essential to the uptake of tools by users.

A further significant issue for ToolKit sustainability concerns the business model for the survival of the infrastructure that will allow it to flourish. This issue cannot be divorced from developments taking place outside the NoE itself, with which the Network is engaging. Included within these considerations are:

- Existing facilities within VPH Partners
- Specific consideration of Biomed Town & PhysiomeSpace, both of which are forward thinking in respect of sustainability, since they offer the basic service free-of-charge, but operate on a pay per use basis for extra or advanced functionalities, and opportunities for third party advertising are also available. Vertical solutions derived from the core technology have their own specific exploitation/commercial plans.
- The activities to be funded under the FP7 Infrastructure Call, currently being assessed. Interaction with these activities will be a strong NoE focus.

Working Groups - Standards & Guidance

These sustainability 'themes' are central to ToolKit survival, but it was agreed that more explicit guidance was required if content providers were to realistically meet the requirements of the ToolKit. Accordingly the WP3 partners are working together to agree and document those features that are considered to be essential elements of ToolKit content. The areas for standardisation and the contributions from WP3 partners are tabulated below, and the guidelines will be easily obtainable from the ToolKit portal as soon as early drafts are available.

Guidance Docs	**UCL**	**UOXF**	**CNRS**	**INRIA**	**UPF**	**UOA**	**EBI**	**USFD**	**IMIM**
Tool characterisation	Yes	Yes	Yes	Yes	Yes	Yes			
Model characterisation		Yes	Yes		Yes	Yes			Yes
Data characterisation		Maybe	Yes	Yes	Yes	Yes		Yes	Yes
Ontological Annotation		Yes	Yes	Maybe	Yes	Yes	Yes	Probably	Yes
Provenance		Yes		Maybe		Yes		Yes	
Licensing		Yes		Maybe	Yes	Yes		Yes	
Ethico-legal issues			Some					Yes	
Interoperability	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Usability?	Yes	Yes	Yes	Yes	Yes	Yes			

Table 1: Essential elements of ToolKit content and WP3 partner contribution

Working Groups - Content Provision

A clear need has been identified for proactive engagement with the providers of candidate ToolKit content, and WP3 is now equipping itself to provide an appropriate service for the development and curation of such material, in accordance with the developing guidance being established by the Standards Working Groups.

Activity	UCL	UOXF	CNRS	INRIA	UPF	UOA	EBI	USFD	IMIM
Exemplars									
SeedEP1			Yes					Yes	
SeedEP2			Yes	Yes	Yes			Yes	
SeedEP3		Yes	Yes					Yes	
SeedEP4	Yes		Yes					Yes	Yes
SeedEP5	Yes		Yes	Yes	Yes			Yes	
EP6			Yes				Yes	Yes	
VPH-I Projects									
ARCH	Yes		Yes					Yes	
ARTreat								TBA	
ContraCancrum	Yes		Yes	Yes				Yes	
euHeart		Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
HAMAM			Yes	Yes	Yes			Yes	
IMPPACT	Yes		Yes					Yes	
NeoMARK			Yes		Yes			Yes	
PASSPORT			Yes	Yes	Yes			Yes	
preDiCT	Yes	Yes			Yes	Yes		Yes	Yes
PredectAD			Yes				Yes	Yes	
VPH2	Yes	Yes						Yes	
VPHOP	Yes		Yes					Yes	
C4P1: RICORDO		Yes		Yes	Yes	Yes	Yes	Yes	
C4P2: MSV			Yes		Yes	Yes		Yes	
C4P3: TUMOR		Yes						TBA	
C4P4: NMS Physiome								TBA	
C4P5: Sim-e-child		Yes						TBA	
External Sources									
Tools - Imaging	Yes		Yes	Yes	Yes	Yes		Yes	Yes
Tools - Translate		Yes		Yes				Yes	
Tools - Other	Yes	Yes	Yes			Yes		Yes	Yes
Models - Cell		Yes				Yes			
Models - Anat		Yes				Yes			
Models – M'scale		Yes	Yes			Yes			
Models – Other		Yes				Yes			
Data - Images				Yes		Yes		Yes	
Data - Physio			Yes			Yes	Yes	Yes	
Data - Other						Yes		Yes	

Table 2: List of ToolKit content providers and nominated WP3 partner contacts in support of content delivery

Supporting Content Delivery

The content provision table is not exhaustive and is a reflection of the fact that WP3 does not have the resources to cover all aspects . Nonetheless several additional areas would benefit from support and three elements are highlighted below:

- Licensing
- Training
- Dissemination

Supporting Content Delivery - Licensing

A common issue for developers of scientific software is dealing with intellectual property considerations, in particular software licences. The requirements of the host organisation(s) involved in developing the software need to be taken into consideration, as well as the licences of any software libraries used in the final product. These topics need to be carefully considered when deciding how to release software, and under what licence. They are also worth considering when beginning any software project, since decisions made earlier in development may have a significant impact on any later decision to release.

In an effort to assist VPH projects in understanding these issues, and also assist users of VPH tools in understanding their licenses, WP3 has provided some basic guidance on software licensing, along with links to relevant information from third parties, to the VPH ToolKit Portal (see http://toolkit.vph-noe.eu/component/content/article/9).

The VPH-NoE advocates the use of an Open Source licence whenever possible, and recommends selecting one from the list discussed on the portal. This is to keep the number of licences used by the VPH community manageable, and avoid the legal complexities that arise from building on software with unusual licence configurations. The licences listed are: AGPL, Apache v2, 3-clause BSD, 4-clause BSD, CeCILL v2, CeCILL-B v2, CeCILL-C v2, EUPL, GPL v2, GPL v3, LGPL v2.1, MIT, and MPL v1.1.

For every licence listed, the main features of each licence are set forth in a manner that is free from legal technicalities. Each licence is categorised according to 4 criteria:

- Whether or not it has been approved by the Open Source Initiative.
- Whether the licence is business friendly - that is, whether software released under the licence can be incorporated into a closed source software product that is then released under a commercial licence.

- Whether the licence is compatible with the GPL. This is important because the GPL is a viral copyleft licence. If GPL-licensed code is incorporated within software, it can only be released under the GPL. If other complementary components are released under incompatible licences, the GPL restrictions may mean that it becomes impossible to release the software. The GPL is only compatible with licences whose restrictions are subsets of those in the GPL. A table has been produced, indicating which licences are likely to be compatible with each other.
- The legal jurisdiction specified in the licence, if any. Some groups are reluctant to use GPL because it is unclear whether all its conditions are enforceable outside the USA. Also, at least one UK University is known to be averse to using MPL, since it specifies that it shall be governed by California law provisions, and they are reluctant to adopt the legal risks associated with this.

This information is designed to support the VPH ToolKit Portal since it includes a licence field for each entry. Where the licence (or one of the licences) used for a technology is one from our list, the entry's display will then link to our information page on that licence.

The optimum approach to licence management, and the possible issues affecting inter-use and reuse of ToolKit components, will continue to be examined as part of the ToolKit portal structure. Methods of easily identifying licence conflicts and resolutions will be considered, and an understanding of the steps being taken by the Open Source Initiative to resolve licensing conflicts will be studied.

Supporting Content Delivery - Training

Cooperation between NoE Workpackages 3 and 4 means that there is now a systematic approach to the process of ToolKit training, which will be further elaborated as ToolKit guidelines are written and further content is introduced. In essence ToolKit components will be used increasingly as the centre of training activities, with directed activities focusing on the various categories of tools and their application. The first such general introduction to tool interoperability will feature image processing, and the workstream approach to component selection and integration will be used to solve practical community problems in a sustained exercise to be held during the summer of 2010 at UPF in Barcelona.

In addition to training in the background and methods of tool use, the cooperation with WP4 will be explored more deeply, to determine whether training in software development methodologies and software testing and evaluation should be introduced.

Supporting Content Delivery - Dissemination

With a clear strategy for ToolKit content emerging, the WP3 team is now beginning discussions with Workpackage 5 to identify the optimum sustainable approach to ToolKit provision, hosting and dissemination. The clear target is to have a software repository system that is sufficiently valuable and attractive to the community that it will live on and continue to be supported without perpetual NoE involvement.

A key element in this process will be the engendering of a significant level of cooperation with the Integrated Project that will be selected to build the VPH Infrastructure as a result of the Call for Proposals that is closing as this document is being prepared. The likely introduction of a sophisticated, secure and highly-capable delivery mechanism allowing the VPH community to share and cooperate in the development and use of tools, models and data is a central part of the EC's VPH investment strategy, and it is clear that cooperative engagement of the NoE with this endeavour will have a clear prospect of continued success beyond that which the Network might achieve in isolation.

Technical Developments

Many WP3 partners are involved in development activities designed to provide ToolKit infrastructure, content or both. In this section the developments during the last period are examined. Many categories of work are described, roughly corresponding to the WP3 tasks listed in the DoW. These include the development of tools for VPH research, addressing the availability and utility of VPH models and data, workflow systems for orchestrating disparate systems and technologies to achieve VPH goals. It also includes the use of high performance computing, the over-arching themes of standards and ethico-legal guidelines, and finally discussion of the VPH ToolKit Portal, along with model and data repositories. In each section, developments are described with particular reference to the evolved goals of the ToolKit, and hence long-term sustainability. Plans to achieve this in the future are also outlined in each case.

Content brought into the ToolKit from VPH NoE Exemplar Projects, VPH-I contributors and sources external to the NoE is discussed in detail in later sections.

Generic Tools (NoE Task 3.4)

Virtual Research Environment (UOXF)

Rather than continue to concentrate on tool development, through a joint appointment with the preDiCT VPH project, we have sought to further direct efforts within preDiCT to be in line with NoE goals. These are hence mainly described in Section 6.9. The primary output is the Virtual Research Environment (VRE), the main component of which is the Chaste cardiac simulation software. Version 2.0 of Chaste was released in April 2010, and among other enhancements now includes support for ontological annotation of CellML models of single cardiac cells, improved documentation and usability for non-technical users, and features to enable closer integration with the VRE, in particular the use of Chaste within scientific workflows. Compatibility with other software has also been extended, both through support for result visualisation using cmgui or any VTK-compatible visualiser, and through support for more versions of the libraries Chaste uses.

Application Hosting Environment (UCL)[1]

The Application Hosting Environment (AHE) a lightweight grid interface that provides a simple set of services, allowing users to interact with grid resources without requiring specific knowledge of the details of each HPC resource they wish to use. Version 3 of AHE is scheduled to include support for urgent computing scenarios (essential for many medical applications) and the hosting of whole workflows as single applications. In addition, AHE 3.0 will make grid computing more transparent through an innovative resource brokering mechanism, that will mean the users need not concern themselves with where their application is running.

AHE 3.0 will also include a usable security mechanism, **Audited Credential Delegation**, which allows end-users to acquire credentials from their local site administrators to access grid resources. This means removing X509 digital certificate from end-users' experience. The mechanism has authorisation and auditing components that enable delegation of grid certificates between a group of users of AHE, locally auditing and controlling users' requests for accessing grid resources. A local audit trail will record information, such as which certificate and proxy were used in a user's request to perform a task on the grid and unauthorised attempts to access grid resources.

There are currently seven VPH-I and two NoE Seed EP projects making use of DEISA infrastructure. In addition, UCL is involved in negotiating the MAPPER project with the Commission's Infrastructure unit, which is expected to start on September 1, 2010. MAPPER will deliver distributed multiscale computing services to a large number of user communities, including the VPH. The link with the MAPPER project will provide all required infrastructure support, as well as easy access to European e-Infrastructures. Moreover, UCL also has substantial trans-Atlantic grid and HPC experience, and has been involved very actively in TeraGrid and with the US Department of Energy laboratories in work aligned with VPH. We intend to publish instructions on how to gain access to these infrastructures using AHE via the NoE ToolKit portal.

[1] Ali N. Haidar, Ali E. Abdallah, P. Y. A. Ryan, P. V. Coveney, B. Beckles, J. M. Brooke, and M .A. S. Jone. "Formal Modelling of a Usable Identity Management Solution for Virtual Organisations". Jeremy W. Bryans and John S. Fitzgerald (Eds.): Formal Aspects of Virtual Organisations 2009 (FAVO2009), Electronic Proceedings in Theoretical Computer Science (EPTCS) 16, 2010, pp. 41–50, doi:10.4204/EPTCS.16.4
B. Beckles, Ali N. Haidar, S. Zasada, P.V. Coveney, "Audited Credential Delegation: a Sensible Approach to Grid Authentication", Accepted in 5th Int'l IEEE e-Science conference, Security in e-Science workshop, December, 2009, Oxford, UK.
S.J. Zasada, P.V. Coveney, Virtualizing access to scientific applications with the Application Hosting Environment, Computer Physics Communications, Volume 180, Issue 12, 40 YEARS OF CPC: A celebratory issue focused on quality software for high performance, grid and novel computing architectures, December 2009, Pages 2513-2525, ISSN 0010-4655, DOI: 10.1016/j.cpc.2009.06.008.

GS Engine (UCL)

GSEngine is a workflow tool that provides the ability to create workflows through graphical or scripting interfaces, and to share workflows amongst a community of users through a workflow repository. We have worked with ViroLab developers to allow workflows to be constructed from applications hosted in AHE. We have also worked with ViroLab developers to deploy a test installation of GSEngine on NoE virtual machine hosting facilities. We will make this system available through a future ToolKit release, but are expecting to work with the ContraCancrum project to implement their workflows in order to fully test this system.

End user GUI software tools: Cmgui (UoA)

http://www.cmiss.org/cmgui

The main emphasis in cmgui development is the completion of a C API to allow control of all functionality by external applications using it as a library. Cmgui 2.7 (April 2010) incorporates a major revision to the API to follow a consistent style for which we can give improved promises of future compatibility, plus new or enlarged APIs for regions, fields, materials and time. A new main object 'context' supports staged start-up enabling non-graphical functions to be used without the overhead of creating a user interface. In many cases significant refactoring has been carried out to implement the new API.

Two other noteworthy features in Cmgui 2.7 include: a new and more accurate isosurface algorithm with support for tetrahedra and triangle wedge elements; tetrahedral meshing support using the netgen library.

Over the past year, 4 projects have proceeded in parallel code branches and involve ongoing developments and integration into the main code-base in coming weeks or months:

1. A restructuring of scene graph and graphics data structures, on which the new graphics API is based.

2. General purpose fitting and optimisation, including a changeover to double precision mathematics mid-April.

3. Incorporation of the OpenCASCADE library to view and interact with CAD shapes as geometric fields.

4. Development of a new zinc browser plugin for embedding the cmgui visualisation engine in web pages. This is based on the npruntime interface supporting more browsers than the existing firefox-based xpcom zinc (Safari, Chrome, Opera, but not IE). The new zinc plugin will use only external API and no insecure legacy commands.

The other major current work is implementing input/output of FieldML version 0.2. Some refactoring of cmgui's finite element data structures is in progress to better handle FieldML's more general data model, plus new models exported from the OpenCMISS computation engine. The cmgui team has had much input into the design of FieldML, ensuring it is capable or extensible to support all current cmgui data structures and projected future needs. When complete, FieldML will become the main file format for field input and output within cmgui.

The FieldML, graphics restructuring and API, zinc-npruntime and fitting/optimisation developments are slated for mid-year completion, with only uncertainty on delivery of CAD functionality in standard builds due to its size.

See also section 4.1.2 under data fusion.

OpenCMISS (UoA)

http://www.opencmiss.org/

OpenCMISS is an open source interactive computer program for Continuum Mechanics, Image analysis, Signal processing and System Identification developed in a collaboration of a number of institutions, including the University of Auckland, Oxford University and Universitat Pompeu Fabra (UPF). OpenCMISS progress has been steady, and progress has been made on solving the following problem types: Laplace, Poisson, linear elasticity, finite elasticity, fluid dynamics (Stokes, Navier-Stokes, Darcy flow; for all of these, both steady state and dynamic problems are handled). In preparation for solid/fluid interaction, support is being added for the arbitrary Lagrange-Eulerian (ALE) formulation.

Progress has been made on creating the FORTRAN language API layer, which is now used by the OpenCMISS testing examples that are part of the daily automated testing of OpenCMISS. This allows end users to write FORTRAN programs that use OpenCMISS via its API. The next step is to make this library accessible from a wider range of languages, starting with the C programming language, and then other languages via the SWIG system. Significant progress for access from C has already been made.

Physiome Model Repository 2 (PMR2) (UoA)

http://www.cellml.org/tools/pmr

PMR2 version 0.1 replaced PMR1 in June 2009. Key features of PMR2:

- Facilitates model exchange directly between modellers, in possibly distant locations, without reliance on a central repository;

- A detailed change history record for each model;
- User access workflows to control privacy when required;
- Embedded workspaces to enable model reuse;
- The ability to include extra information, such as the experimental data on which the model is based, to enhance the model descriptions (metadata).

Version 0.2 of the PMR2 (Physiome model repository 2) software, was deployed in January 2010, and provides enhanced support for CellML 1.1, which will promote the library based approach to CellML modelling that is at the core of the CellML 1.1 design. PMR2 is also ideally suited for being used as the software for hosting the FieldML repository, and work on this has been underway, and will be deployed as version 0.3 of PMR2 in mid 2010.

Vascular Connectivity API (EMBL-EBI)

The Foundational Model of Anatomy (FMA) provides an excellent description connectivity, allowing information to be mapped between:

- named blood vessels (via the branch_of and continuous_with_distally relationships),
- named blood vessels and anatomical regions they supply or drain (via the arterial_supply and venous_drainage relationships).

However, two key shortcomings of the FMA representation of the human cardiovascular systems are that:

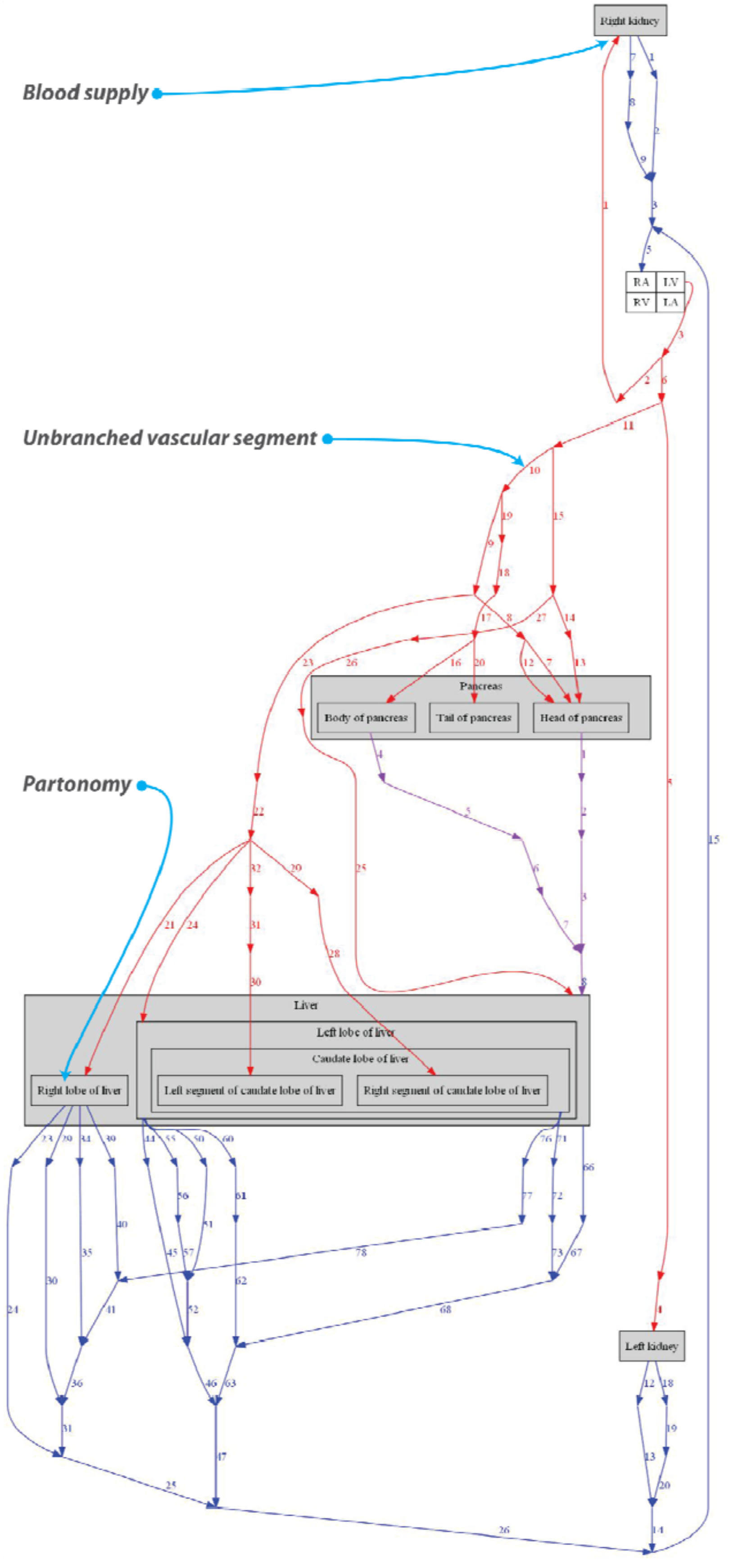
- While parent vessels are mapped to branches, the order of their branching is not represented, making it difficult to build a set of connected vascular segments that correspond to real anatomical structures;
- Over 90% of existing FMA vascular supply/drainage connections to anatomical regions relate only to the heart or brain.

Ongoing efforts by the EMBL-EBI partner have addressed the above two problems to create an extended vascular connectivity dataset that adds branch ordering and vascular connectivity to physiologically-relevant visceral organs in the thorax, abdomen and pelvis. This work has improved on the FMA data by:

- mapping anatomically correct vessel (unbranched) segments to named blood vessels;

- adding over 400 new blood vessel terms;
- doubling the number of arterial_supply connections to body regions;
- increasing the number of venous_drainage connections by a factor of four.

Furthermore, an API tool that carries out graph analysis and layout of the vascular connectivity network has been developed to support the visualization and VPH data overlay. An example of the use of this tool is show in the figure below.



An example of a connectivity graph of anatomically correct (numbered) cardiovascular segments, as they link to body regions.

This graph:

1. Represents cardiovascular networks as graphs, such that each edge in a graph has a stable identifier and represents a distinct unbranched segment of a named blood vessel;

2. Depicts vascular anatomy correctly: to this end, the vascular branching order was also added to the representation;

3. Maps vessels to the anatomical regions they service: for example, in the graph, terminal blood vessel segments are directly linked to the anatomical locations they supply or drain;

4. Captures the partonomy relationships of non-vascular structures of the anatomy: this allows the tracking of regional dependencies for vascular connections;

Segment colours are red: arterial, blue/purple: venous).

Figure 2. Cardiovascular Connectivity

Imaging Tools

Analysis from the RTAE[2] showed the great variety of medical images visualisation and processing software in use and under development within the VPH. It also has exposed the lack of high level interoperability between software as a fundamental issue.

Currently, basic interoperability ("level-0" interoperability) is achieved through the adoption of the common data format: DICOM. The DICOM format is the clinical standard for image data and also the most used by the VPH related research community. Hence the fundamental need for full DICOM compatibility of imaging software and at higher levels of DICOM management.

"Level-1" interoperability would enable shared libraries or an Application Programming Interface (API) to be developed within one development environment and operate within another environment. However, 'full' level-1 interoperability is extremely difficult[3] or even impossible to attain due to the wealth of medical image visualisation and processing software available, which are moreover often environment dependent[4]. To overcome these interoperability issues, the Imaging Tools subgroup has proposed two axes of development for VPH ToolKit: DCM API and GUIDE.

DCM API will be an open source and cross-platform mid-level DICOM management layer including a GUI and advanced features (images indexation, PACS/HIS access and GUIs that could be plugged directly into existing software).

GUIDE (GUidelines for Image Development Environments) will be an online tool, part of the VPH ToolKit portal, which aims at federating existing image visualisation and analysis tools and providing the VPH community with support for enabling sharing and open use.

DCM API (CNRS/UPF)

The development of DCM API is based on the results of a survey[5] which identified needed features. From this survey, the Imaging Tools subgroup has provided a software specification

[2] http://www.vph-noe.eu/vph-repository/doc_download/106-vph-noe-first-annual-report-1st-june-08-to-31st-may-09

[3] https://vphnoewiki.comlab.ox.ac.uk/DeliverablesSection?action=AttachFile&do=get&target=D5.6.pdf

[4] But some initiatives, as the CTK consortium -http://www.commontk.org-, want to try to success this challenge.

[5] http://www.vph-noe.eu/vph-repository/doc_download/141-dicom-survey

requirements[6] (SRS) document to expose a global overview of DCM API and the related constraints. A second document, the software design document[7] (SDD), describes the architecture of the API, its interactions, its mechanisms and its behaviour using UML diagrams (use-cases, state, data flows sequence, class). These standardised design documents were made publicly available, thus providing a framework for integration in software developments by any VPH community member.

The VPH ToolKit version of DCM API is in the process of being implemented by the Imaging Tools subgroup, the state of development being available on a subversion server[8] with a project tracking system[9]. Currently, a pure abstraction layer is implemented in order to allow basic interoperability.

The development plan for the DCM API is:

- ✓ The first implementation version for June 2010 (PM26). This release will be an alpha version with basic functions implemented.
- ✓ The second version (beta) is scheduled for December 2010 (PM32) with all functionalities and several GUI interfaces. This version will be exposed in the VPH ToolKit Portal. An integration phase in NOE Imaging Tools subgroup partner’s tools (Creatools, GIMIAS, MedInria) will be done.
- ✓ A first dissemination to “image related” Exemplar Projects is scheduled for June 2011 (PM38) which will provide the first feedbacks and corrections.
- ✓ A second dissemination beyond NOE will start after January 2012 (PM45) to VPH-I partners, “image” community (Kitware, ...).

To ensure long term sustainability for the DCM API, different thorough documents (with a firm standards adherence) have been and will be produced: SRS, SDD, verification and validation plan. These documents will provide a strong canvas for future developments. Moreover, it is to be noted that DCM API will be compliant with the recent DICOM standards new specifications[10].

The fact that DCM API will be part of the development framework of the Imaging Tools subgroup members will ensure that all features are correctly implemented.

6 http://www.vph-noe.eu/vph-repository/doc_download/140-dcm-api-specifications
7 http://www.vph-noe.eu/vph-repository/doc_download/165-dcm-api-design-document
8 svn+ssh://tux.creatis.insa-lyon.fr/svn/dcmapi
9 http://www.pivotaltracker.com/projects/62991
10 ftp://medical.nema.org/medical/dicom/supps/sup118_lb.pdf

GUIDE (CNRS/UPF)

GUIDE aims to federate existing imaging visualisation and analysis tools, providing support for their sharing and open use. The tools exposed by GUIDE could be software, libraries and/or scripts, online application and development environment in function of end users (software developers, researchers, clinicians or industrial users). However, providing only a list of software tools is not sufficient to help the user because medical images analysis is very problem dependant. This is the reason why GUIDE will also reference processing use cases and use them to drive the user in search for a tool. The structure of GUIDE will allow for the assessment of the exposed items in terms of documentation, licensing, reliability and sustainability.

The tools and use-cases provided by EPs and VPH-I will be the first exposed through GUIDE. A first interview to test GUIDE concepts and feed GUIDE has been performed with the seed EPs[11] providing a description of typical imaging workflows related to their project, i.e. the different steps involved to go from the loading of specific data to an end result. A description of the different tools used for these steps and the generated formats was also obtained. For VPH-I projects, a preliminary work is summarised by the table below. It shows that many VPH projects will have one entry or more in GUIDE.

VPH-I Project	*Future* tools developed/exposed	GUIDE Index	GUIDE use-case
ARCH	VmTk (Vascular Modelling Toolkit) Gnuid (3D CFD Solver) pvNS (Vascular network solver)	**YES**	**YES**
Contracancrum	Doctor Eye (annotation platform – GUI - for identification and delineation of tumours in medical images - DICOM Viewer	**YES**	**YES**
euHeart	plugins for Gimias MedINRIA 2.0 (image processing platform)	**YES**	**YES**
HAMMAM	MeVisLab Xnat (The Extensible Neuroimaging Archive Toolkit)	**YES**	**YES**
NEOMARK	Image feature extraction tools, based on MITK	**YES**	**YES**
IMPPACT	Segmentation / Visualisation 3D / Data management	**YES**	**YES**
PASSPORT	3D-IRCADb (3D Image Reconstruction for Comparison of Algorithms Database) VR-Render	**YES**	**YES**
VPHOP	Hyperrmodel technology software layer to connect models from different scale : OpenMAF 3	**YES**	**YES**

Table 3: GUIDE and the VPH-I

11 https://vphnoewiki.comlab.ox.ac.uk/DeliverablesSection?action=AttachFile&do=get&target=D5.5.pdf

To ensure long term use, several reference documents have been produced. The SRS document[12] has been designed to set the sphere of activity of GUIDE and to define the links between tools and use-cases. The SDD document[13] has been written with ERCIM collaboration to ensure compatibility with VPH portal functionalities and design. A preliminary analysis[14] has been performed in order to define the fields needed to assess the tools and use-cases that will be exposed in GUIDE.

The possibility to characterise some GUIDE items through an ontology[15] is in progress. In particular, previous work done toward a DICOM ontology and a medical image ontology could be used as a formalism to describe use cases using common vocabulary and thus help in formalising the classification of GUIDE use-cases with the integration in RICARDO project.

The development plan of GUIDE is as follows:

- ✓ A first online version for September 2010 (PM28) will expose imaging subgroup tools and use-cases.
- ✓ Dissemination through the VPH community with the exposure of EPs and VPH-I tools and use-cases for February 2011 (PM33)
- ✓ A large dissemination beyond NOE for June 2011 (PM37).

The sustainability of GUIDE will be ensured if a strong community is created around use-cases, associated discussions, comments or questions to increase their pertinence and their application domain definition. The maintainability of GUIDE with the help of ERCIM (to track tools updates, dead links, and to provide the GUIDE host server) is an important key for the generation of a pool of expertise and ensure GUIDE's growth beyond the scope of VPH.

12 http://www.vph-noe.eu/vph-repository/doc_download/142-guide-specifications
13 http://www.vph-noe.eu/vph-repository/doc_download/142-guide-design-documents
14 http://www.vph-noe.eu/vph-repository/doc_download/142-guide-fields-values
15 http://www.vph-noe.eu/vph-repository/doc_download/ontologies-for-GUIDE

Fusion Tools (NoE Task 3.8)

Horizontal Fusion (INRIA)

The objective of the horizontal fusion toolbox is twofold:

- Propose fusion algorithms to the community within an intuitive interface, and
- Integrate fusion algorithms from other groups into this application.

The new architecture of the toolbox is now in place and is used in the new release (05/2010). It is based on the new version of the MedINRIA software, which kernel will be open-source (http://www-sop.inria.fr/asclepios/software/MedINRIA/). A generic API for fusion algorithms is also being designed in order to ease the integration of external contributions. The plan is to use the future ontological descriptions of data and algorithms discussed within GUIDE and the imaging tools. In order to decide on the algorithm that we implement in the application and its parameters, extensive testing and parameter exploration has been done.

Reports on these tests and software releases can be found here:

http://www-sop.inria.fr/asclepios/projects/noe/

Other VPH software does include horizontal fusion capacities. The goal of this toolbox is to be able to interoperate with a wide range of tools.

For instance, cmgui (http://www.cmiss.org/cmgui) developed by the Auckland Bioengineering Institute (ABI). The Cardiac Atlas Project (www.cardiacatlas.org) includes a software called the *CAP-client*, which is based on cmgui software,. The CAP-client software allows time series of cardiac magnetic resonance images to be fused via the fitting together in space coordinates of different DICOM cine images, as well as the fitting of a cardiac left ventricular model to the DICOM cine. The Breast Modelling project at the ABI is also making use of cmgui as well as mechanical modelling via CMISS-cm to develop a software system for clinical use, where different breast diagnostic imaging modalities (e.g. mammography, ultrasound, X-Ray tomography) are fused. Mechanical modelling is required since the breasts are imaged under different mechanical loading situations for each of these modalities.

Another example is GIMIAS (http://www.gimias.org), developed by UPF. It has a plug-in architecture and fusion functionalities. UPF already collaborated with the cmgui team to interoperate both. While GIMIAS focuses on providing a framework for developing workflow tools, these workflows usually include some form of horizontal data fusion. Further discussion on cmgui developments is included in section 4.1 "Tools".

To conclude, the Toolkit content proposed here is a "Data Fusion toolbox" with a generic API and a plug-in mechanism so that VPH researchers will be able to perform fusion and integrate their own tools. The sustainability of this content will be eased by the fact that the generic platform will be open source. Moreover, it will be used and maintained by INRIA even after the end of the VPH NoE. Attracting a community of developers would ensure a lasting lifetime of this toolbox.

CMGUI and Data Fusion (UoA)

Cmgui is focused on visualisation and manipulation of "field data and models", which encompasses a wide range of functions and applications. One of the key applications to which cmgui is ideally suited is horizontal data fusion (different data modalities at same spatial and time scales) and vertical data fusion (different spatial and time scales). For example, the Cardiac Atlas Project (www.cardiacatlas.org, also work done by the Auckland Bioengineering Institute (ABI)) includes software called the *CAP-client*, which is based on cmgui. The CAP-client software allows timeseries cardiac magnetic resonance images (cardiac MRIs) to be fused via the fitting together in space coordinates of different DICOM cine images, as well as the fitting of a cardiac left ventrical model to the DICOM cine.

The Breast Modelling project at the ABI is also making use of cmgui as well as mechanical modelling via CMISS-cm to develop a software system for clinical use, where different breast diagnostic imaging modalities (e.g. mammography, ultrasound, X-Ray tomography) are fused. Mechanical modelling is required since the breasts are imaged under different mechanical loading situations for each of these modalities.

Two commercial entities in Auckland have also based key parts of their software on cmgui, with applications that also relate to data fusion. UPF have collaborated with the cmgui team to trial the use of the interim cmgui API from the GIMIAS software platform. While GIMIAS focuses on providing a framework for developing workflow tools, these workflows usually include some form of horizontal data fusion.

Further discussion on cmgui developments is included in section 4.1 "Tools".

Models (NoE Task 3.2)

A key need of VPH modellers is access to high quality models in standard forms that have been well curated to indicate the kind of scientific questions they are capable of addressing. The development of reference descriptions of models that incorporate computer-readable descriptions of the simulations performed, in order to allow researchers to reproduce the results easily, is also an area of considerable interest, as evidenced by the MIASE effort.

At present there are a multitude of electrophysiology action potential cell models (around 140 are in the CellML model repository as of March 2010[16]). Often the properties of these models are detailed in many different places: the original papers; consequent studies; or are not published anywhere. As a result a modeller or experimentalist wishing to use the model in a single cell or a tissue level simulation may not be aware of its limitations or suitability for their study. Consequently, a new thread of work is commencing at UOXF (supported by preDiCT, the VPH-NoE, and internal effort) to address several key goals of the ToolKit in the context of cardiac electrophysiological modelling. The aim is to develop a system to run a suite of tests on cell models encoded in CellML. The results of these will indicate clearly what experimental conditions the model is capable of reproducing. As well as providing a useful resource for modellers, this system will demonstrate the potential for the use of standards and ontologies in improving model curation. Ontological annotation will enable the relevant biological entities in each model to be identified, addressing the problem of differing modelling conventions resulting in a lack of exchangeability. The tests themselves will be described in a standard format, SED-ML (Simulation Experiment Description Markup Language, http://biomodels.net/sed-ml/), and this work will inform the development of that standard.

While the system will be targeted at cardiac models, similar functionality would certainly be useful in other fields as well. The use of standards in developing this system will ensure that it is relatively easy to extend to such applications. It will be designed to allow the addition of further tests, referencing any CellML (or indeed SBML) model that is appropriately annotated with terms from any ontology (chosen by the test author).

CellML core specification and metadata specification (UoA)

http://www.cellml.org/specifications/cellml_1.1

The CellML **core** specification version 1.1 is the latest version of the specification, and this has been available for a few years. The CellML **metadata** specifications include: graphing

16 http://models.cellml.org/electrophysiology - note that some papers introduce several model variants.

metadata, simulation metadata, citation metadata.

Status:
We continue to receive feedback from current users and potential users. This feedback has highlighted that some incremental enhancements are required, and the design team's analysis has shown that there is also scope for a fundamental revision, which could be based on "lambda calculus", and allow for a wider range of ways of expressing the essential structure of models.

The core specification design team has made progress on both of these fronts. Regarding incremental enhancements, a proposal for representation of "reset rules" for handling certain types of events has been put forward, and a prototype implementation in the CellML-API is in progress and near completion, after which point the representation will be finalised. The specification team takes the approach of proving specification changes work by means of prototype implementations. Another identified potential incremental enhancement is the representation and API support of delayed differential equations.

The graphing and simulation metadata specifications are supported by OpenCell (the flagship CellML software tool). The CellML developers have also contributed to the minimum information about a simulation experiment (MIASE) standard, as well as the simulation experiment description mark-up language (SED-ML). Related papers on these are in progress. A draft of the new citation metadata standard (version 1.1) has been written and is available publicly for comment.

Progress has also been made on standardisation of methods for biological annotation of CellML models.

CellML API development (UoA)

http://www.cellml.org/tools/api

The CellML API has been in a mature state for some years. The objective is to incrementally improve the API, enhancing usability, functionality and robustness.

Status:
The CellML API is under continued development and incremental improvement and enhancement. Also, an overview paper describing the CellML API has been accepted for publication. Progress has been made integrating the CellML API with OpenCMISS, with the goal of being able to use CellML models for a variety of purposes within FieldML type models and simulations: for example, electrical propagation using CellML models to describe cell

membrane electrophysiology; tissue mechanics using CellML models to describe material constitutive properties.

A prototype for handling reset rules has been developed.

Progress has been made on the handling of a broader range of differential algebraic equation (DAE) models by means of the Sundials IDA numerical integrator. For example, this is expected to lead to successful handling of the complete Guyton model (see discussion under SeedEP 1). Currently, subunits of the Guyton model are handled via an indirect DAE handling method, but this does not cope with the full model. IDA is a leading DAE numerical integrator, and can now be used via the CellML API, and this feature will be released in the next release of the API, version 1.8, planned for mid 2010.

Progress has also been made on supporting users of the CellML API. An automated tool for setting up the development environment under Windows is now available. Also, Java developers are now able to use the CellML API, though this has so far only been used on Linux.

Work has also been done on supporting biological annotation via the API. Initially, this has been made available in Java for the SAINT annotation tool (Newcastle University, UK).

End user GUI software tools: OpenCell (UoA)

www.opencell.org

Status:

OpenCell (previously named PCEnv, was also briefly termed "enCORe") is the flagship CellML software tool for development and exploration of CellML models. It is open source, and available for multiple software platforms (Windows, Mac, Linux). In collaboration with Oxford University, a key achievement has been implementation of the first version of the text language for editing CellML models that is based on that used by the COR software. This has been done by essentially merging these two projects at the design level, using the original PCEnv codebase. The COR text support will be available in the next version of OpenCell (i.e. has not been released yet). Further features from COR are likely to be added in future versions. Extensive work has been done on improving the software via bug fixes and core technical work, as well as improving the build system and automatic test system so as to improve software quality. Since OpenCell is based on the CellML-API, improvements to the API usually directly translate into improvements in OpenCell.

Work is underway to perform a major redevelopment of OpenCell based on either Qt/C++ or the Java JVM, since the current XulRunner platform has not gained broad acceptance as a platform for desktop software development. The new platform will ease the task of supporting a high quality software project on multiple operating system platforms, as well as allow for easier addition of key features based on third party libraries available for both Qt/C++ and Java JVM.

FieldML specification development (UoA)

http://www.fieldml.org/

The FieldML format is under active development. A pre-cursor to the FieldML format is widely in use, and is used by the ABI and collaborators to interchange field models between cmgui and CMISS. An early design of the XML FieldML format was developed in 2004 (FieldML version 0.1), but has not been adopted. During 2010, FieldML version 0.2 will be made available, as well as a β-version of the FieldML API. The initial usage will be for interchange of field models between cmgui and OpenCMISS.

FieldML 0.2 design work has aimed at defining an improved specification for an XML representation of FieldML. Cmgui and OpenCMISS already support a pre-cursor to FieldML (called exnode/exelem format).

Status:

Good progress is being made. A rigorous method for describing the interim design is being used which consists of representing the design using object oriented Java code, which compiles and is executable, providing some degree of continuous validation. The Java objects represent the main FieldML data structures. An automated mapping from the Java design to XML "mock-ups" has also been implemented, allowing preliminary XML representations to be demonstrated. There are a range of test-cases that are represented in this way, and the design work is progressing by working through further test-cases.

Cmgui's internal field representation is being reworked to converge with the FieldML design, and good progress has been made.

FieldML API development (UoA)

Goal: To deliver a standard API to enable software tools developers to easily incorporate support for FieldML.

Status:
Work on the β-version of the API is underway, with the goal being to be able to use it for FieldML input into OpenCMISS and cmgui during May 2010.

However, a precursor to the FieldML API is available in both cmgui and OpenCMISS. Cmgui's API has been available for some years now, but is still under active development.

The OpenCMISS system is being designed as primarily a library to be accessed via an API, although a standard graphical user interface (GUI) for OpenCMISS will also be provided.

BioSignalML (UoA)

BioSignalML aims to standardise the interchange of timeseries data, e.g. electrocardiogram (ECG), electroencephalogram (EEG). This is being developed by means of work for a PhD at the ABI by David Brooks.

The focus has been on developing a software library to implement concepts from the abstract model for BioSignalML, using Scala (Java VM compatible). Progress so far:

- Defined core API elements and services.
- Imported metadata from EDF signal files and separate annotation files into a Jena based RDF store.
- Developed a prototype web-server providing signals to a Javascript/HTML5 web-browser application for visualisation.
- Imported EDF signal data into a new, HDF5 based file format.

Next activities will be:

- Provision of signal data to applications (both via streaming and direct use of the API).
- Set up of a Semantic Web server to allow exploration of Jena store.
- Bring everything together as a store/repository that is usable as an initial release).
- Write and submit a journal paper covering this work to date, the aim is to have this ready around the end of April.

CellML Annotation (UOXF)

While primary effort in this area has been by other partners, UOXF has been involved in ongoing discussions with UoA and EBI on the best approach to annotating CellML models with ontological information. We have also contributed to work on MIASE, a reporting standard for simulation experiments. Work on both these fronts is expected to increase in the next period

DAEML Systems Model Markup Language (CNRS Evry)

DAEML is a language with similar concepts to CellML, but designed to be easier to use in the context of renal modelling. Notably, it allows easier use of XSLT to convert DAEML into other formats, such as input files for M2SL. For CellML, such a task must be accomplished through use of the API. Ideas from DAEML are now being fed into CellML. The user interface has received considerable effort and is now almost ready for use. Its development has brought some changes to the underlying DTD, and we will be testing the whole process from model markup in the DAEML GUI, implementation of the resulting XML and interactive running of simulations through the M2SL GUI (also nearly ready) between now and end of SAPHIR funding (mid July).

Assuming this goes well, we will submit specific suggestions to the CellML team for revisions in future CellML versions to make it amenable to this scenario as well (i.e., via XSLT converters).

Data (NoE Tasks 3.3, 3.7)

The WP3 Data Subgroup has a primary responsibility to explore solutions for data exchange and interoperability within the VPH/NoE. This encompasses far more than data format conversion and probes at issues that are at the heart of the Virtual Physiological Human. The pertinent questions relevant to widespread interaction with VPH data are deceptively simple, namely:

- VPH data – Where is it?
- VPH data – What kind of data is it?
- VPH data – Can I use it?
- VPH data – How can I contribute my data to it?

The first item is an exercise in cataloguing and searching, both of which depend on the second bullet point, which involves description of the data (ontologies, metadata etc.). The third issue touches on practicalities such as data formats (can the data be read by my application?) and authorisation (security, regulatory control, ethics, copyright, validation etc.), whilst the final question raises issues about upload/download, quality control, curation, storage, integration etc. The DoW Task 3.7 broadly alludes to these challenges, characterising them as follows…

- Formalisation of data formats, data access channels and curation policies
- Identification and cataloguing of data (curation)
- Integration of ontologies
- Support for exemplar (EP) demonstrators
- Exploration of clinical data issues (PACS, HIS, ethics, legal etc.)

These elements are designed to secure interoperability of ToolKit content and will continue to be sought, but they must now be assessed in the wider context of sustainability. The implications are significant and were instrumental in precipitating the technical WP3 meeting in February (Brussels), which considered numerous aspects of sustainability through integration, engagement and standards. The sustainability perspective has many facets, but only three categories relevant to VPH-data are presented below (engagement, standardisation, ethico-legal).

Engagement – Exemplar Projects

Infrastructure efforts by USFD have resulted in the provision of a 1TB SFTP resource for local data storage and exchange within the NoE. This use is secondary to the server's primary function as a data hosting facility for (eventual) exposure of USFD clinical data and therefore numerous limitations are associated with the use of this repository (see Appendix for details). The terms of use are articulated in a concise document that is distributed amongst those wishing to use the resource. Currently, the resource stores data for Exemplar Projects that have participated in the NoE and it is anticipated that this will continue as new EPs are funded.

The limitations of the current set-up highlight the need for an underlying, flexible and secure infrastructure for the exchange of data between participants of the VPH. Services like PhysiomeSpace* provide an example of wider data hosting, but many issues peculiar to the VPH remain to be solved before viable data hosting becomes a reality (e.g. a robust framework for the ethical and legal exposure/exchange of clinical data is essential). Avenues to implement a more clinically acceptable infrastructure are being explored within the NoE. The Sixth Call of the Framework Programme (FP7) has highlighted the importance of infrastructure, and the NoE will engage with those projects that are funded, to improve the infrastructure of the VPH.

Engagement - Clinical

Previous work within the NoE has established in principle that clinical data from the Biomedical Research Unit (BRU) at Sheffield, can be exposed for use by the biomedical research (VPH) community. This is subject to certain conditions and assurances. Examples include the need for an infrastructure that can demonstrate robust anonymisation, effective security, guaranteed data integrity etc.). Furthermore, there must be mechanisms that support data curation and safeguard data quality. The work at Sheffield has focused on several of these qualities in the past year, developing a path or workflow that will ultimately permit exposure of clinical data from the BRU. Currenty, it is envisaged that in response to a request for clinically-related data from a 3rd party within the VPH, the clinical centre will provide anonymised data, with the capacity to expose a breadth of data types typically available within the hospital environment. At Sheffield, this includes demographic data (name, address etc.), results from laboratory tests (eg. blood/gas analyses), imaging data etc., but typically these data sets are distributed across multiple, independent and incompatible databases within the hospital system. Hence, an external query involves an intermediate process in which relevant data is identified, transferred and amalgamated from the many data sources behind the hospital firewall.

* PhysiomeSpace - https://www.biomedtown.org/biomed_town/LHDL/users/repository/access/

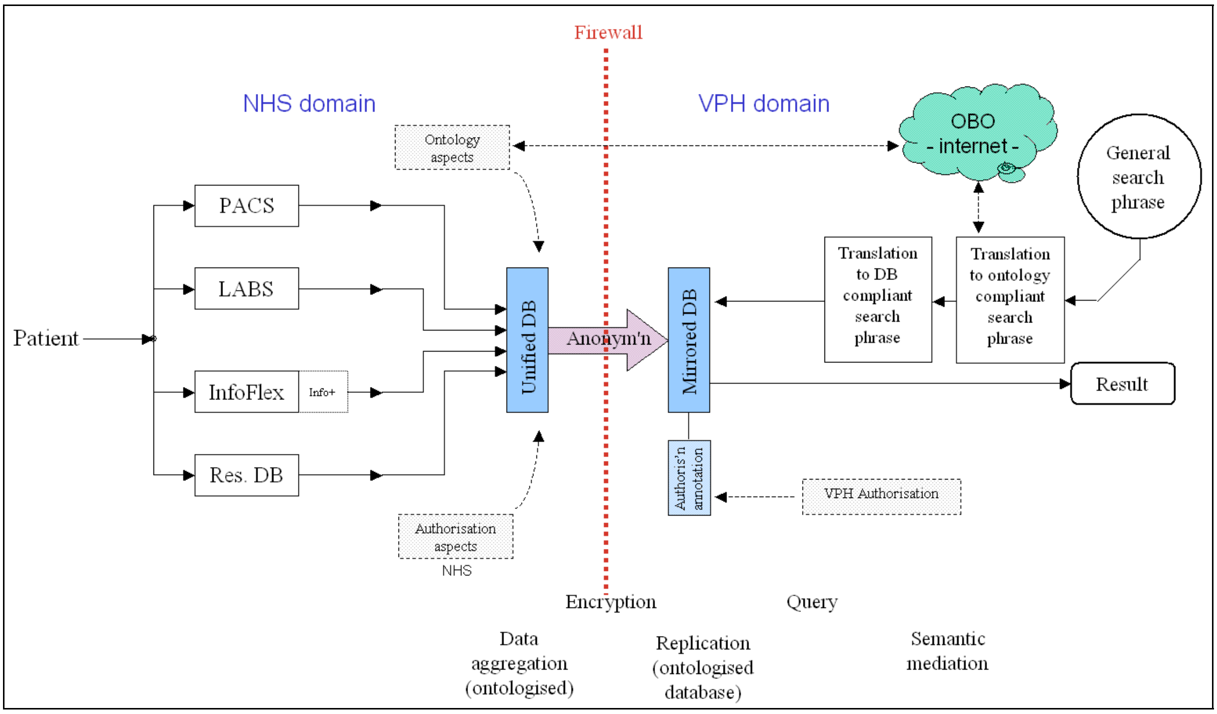


Figure 3: *Schematic of clinical data exposure illustrating the use of an intermediate database for amalgamation of PACS/LABS/HIS/Research data. Authorisation and ontology annotation also feature in the process of clinical data exposure.*

This may involve synthesis of a coherent data set in an intermediate data store or buffer, before being anonymously mirrored to an external VPH-compliant broker on the other side of the firewall. An effective anonymisation step is a prerequisite prior to data release, as is ontology annotation and authorisation.

Ontology annotation is central to the concept of data curation, and several interactions with the ontology community have permitted incremental, steady advances in this area at Sheffield. The ontology component is important because it standardises the nomenclature and meaning of terms held in the clinical databases, thus enabling effective search terms to be compiled in response to a user query. An added benefit of ontology annotation is that it provides an effective fall-back position for those centres unwilling to expose their clinical data. If the ontology alone is exposed in the absence of the clinical data, external queries can still establish which centres possess data of interest. These can then be approached independently to obtain the data (or not) by more traditional means (e.g. a telephone introduction, followed by a personal visit).

In summary, work to date at Sheffield has been in support of the data amalgamation process. Mechanisms have been developed to interface with and extract data from the HIS/EHR and separate mechanisms are now being developed to access the laboratory and PACS data for amalgamation. This constitutes the next year's clinical engagement work, and will include provision for anonymisation and ontology annotation.

Standards

The data subgroup has previously established links with standards bodies (e.g. ETSI) to demonstrate its commitment to the principle of data standardisation. This is important, because early in the project it became clear that the writing of data format conversion tools to support interoperability of every format encompassed by the VPH was going to be an impossibly large task.

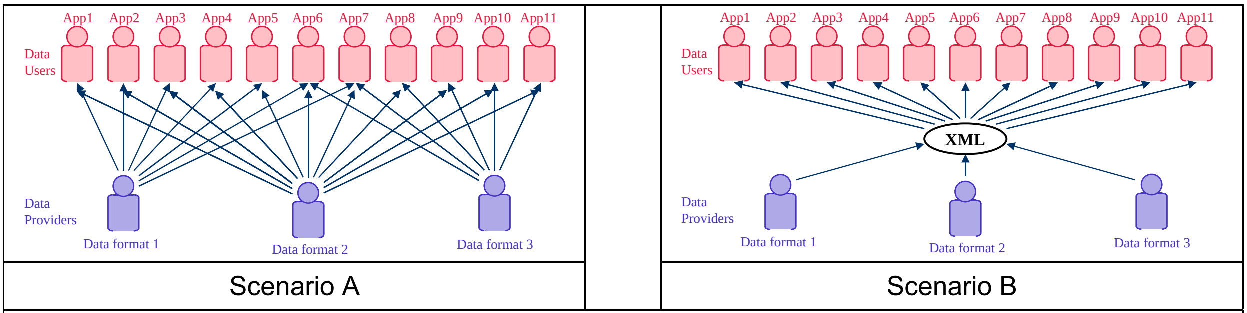


Figure 4: *The use of an intermediate standard reduces proliferation of data format conversion tools. In this example, the use of the intermediary standard reduces the conversion burden from 33 to 13.*

Standardisation is a means of limiting conversion proliferation by providing a common and accepted intermediate format through which data interoperability can be achieved. Indeed, suitable choice of an intermediate format can dramatically ease the data conversion problem and a machine/human readable format (e.g. XML and validated schema) is proposed.

Over the course of the NoE, several exercises to clarify the data format landscape have been undertaken. The simplified table of results below illustrates an example collection of data categories, with predominant formats present in each category. Interoperability can now be characterised by two types of conversion activity:

i. Intra-format conversion: between data formats within the same category

ii. Inter-format conversion: between different data format categories

The NoE strategy is to consider intra-format conversion first, identifying mechanisms by which conversion within each category is possible, and characterising the information that is lost in the conversion process. This information could then be documented for general VPH use. In parallel, this is an opportunity to consider underlying XML formats/schemas that can suitably represent the data in each category. Eventual implementation of such formats will provide a means by which both intra- and inter-category conversion can occur. The standards bodies become valuable at the point of formalising these intermediate formats (XML-based) for wider data exchange. A particular challenge is to be adaptable to the ever-changing formats landscape and astute in the face of competing standards (e.g. data specific representation in an optimised format vs incorporation within an increasingly diverse data 'host' such as DICOM). Standardisation within the VPH is a truly challenging exercise. The NoE does not have the resources to solve the data standards problem, but it can be instrumental in moving the process forwards.

Type of Data	Data Formats	Other Formats
Quantitative Data	Structured markup [xml] Delimited [csv] MS Excel [xls(x)] Matlab [mat] PostgresSQL [bin]	MS Access [mdb] OpenDocument Spreadsheet [ods] dBase [dbf] Dat Files [dat] ARCH Network{Graph, Mesh} [xml] Sepran [bin]
GeoSpatial Data	Stereolithography CAD [stl] Visualisation Toolkit [vtk] OpenInventor files [iv vrml wrl]	3D studio Max file [3DS] OBJ file [obj] PLY file [PLY] ESRI files Libmesh [xda, xdr] Nifti Gipl [vol]
Metadata & Ontologies	Resource Description Framework [rdf] Web Ontology Language [owl]	
Qualitative Data (includes source code & software)	C/C++ source [c cc cpp h hpp] Java source [java] Fortran source [f] Executable [exe, bin] Archives [zip tar gz] Structured markup [xml]	Matlab source [m] CellML [cellml] ASITIC file [tek] Patient Folder [mfo fxz]
Digital Image Data	DICOM file [dcm] JPEG file [jpg] TIFF file [tif] PNG [png] GIF [gif] RAW [raw]	BMP [bmp] Adobe PDF [pdf/a pdf] Software Specific Formats [img/ hdr nii rdc] MetaImage [mha] InrImage [ascii] MHD NII
Digital Image Data	DICOM file [dcm] JPEG file [jpg] TIFF file [tif] PNG [png] GIF [gif] RAW [raw]	BMP [bmp] Adobe PDF [pdf/a pdf] Software Specific Formats [img/ hdr nii rdc] MetaImage [mha] InrImage [ascii] MHD NII
Digital Audio Data	MPEG Audio Layer [mp2 mp3] Waveform Audio Format [wav]	Free Lossless Codec [flac] Ogg Vorbis [ogg oga]
Digital Video Data	AVI file [avi] MPEG file [mpeg mp4] Quicktime file [mov] Windows Movie [wmv wmf]	Shockwave Media [swf] Theora [ogv]
Documentation	Plain Text [txt] MS Word [doc(x)] MS Powerpoint [ppt(x)] MS Excel [xls(x)] LaTeX [tex] Adobe PDF (pdf) Hypertext Markup Language (html)	Rich text Format (rtf) Open Document Format (odt) XML with DTD (xml xhtml)

Table 4: A collection of data formats

Grid Developments (NoE Task 3.9)

Application Hosting Environment (UCL)

The Application Hosting Environment (AHE) a lightweight grid interface that provides a simple set of services, allowing users to interact with grid resources without requiring specific knowledge of the details of each HPC resource they wish to use. Version 3 of AHE is scheduled to include support for urgent computing scenarios (essential for many medical applications) and the hosting of whole workflows as single applications. In addition, AHE 3.0 will make grid computing more transparent through an innovative resource brokering mechanism, that will mean the users need not concern themselves with where their application is running.

AHE 3.0 will also include a usable security mechanism, **Audited Credential Delegation**, which allows end-users to acquire credentials from their local site administrators to access grid resources. This means removing X509 digital certificate from end-users' experience. The mechanism has authorisation and auditing components that enable delegation of grid certificates between a group of users of AHE, locally auditing and controlling users' requests for accessing grid resources. A local audit trail will record information, such as which certificate and proxy were used in a user's request to perform a task on the grid and unauthorised attempts to access grid resources.

There are currently seven VPH-I and two NoE Seed EP projects making use of DEISA infrastructure. In addition, UCL is involved in negotiating the MAPPER project with the Commission's Infrastructure unit, which is expected to start on September 1, 2010. MAPPER will deliver distributed multiscale computing services to a large number of user communities, including the VPH. The link with the MAPPER project will provide all required infrastructure support, as well as easy access to European e-Infrastructures. Moreover, UCL also has substantial trans-Atlantic grid and HPC experience, and has been involved very actively in TeraGrid and with the US Department of Energy laboratories in work aligned with VPH. We intend to publish instructions on how to gain access to these infrastructures using AHE via the NoE ToolKit portal.

Standards, Ontologies and Workflows (NoE Task 3.5)

It was clear from the RTAE, and has been re-emphasised repeatedly since, that VPH research is a very broad field. It covers the full range of human physiology, in both normal and pathological states, using a wide range of mathematical formalisms and simulation approaches. It is primarily for this reason that the ToolKit cannot consist of a single system addressing all needs, but rather must comprise a set of interoperable components based on common standards.

Standards

Standards play a central role in facilitating the exchange and interpretation of the outcomes of scientific research, and in particular of computational modelling – they are necessary in order for data and models to be integrated, leading to greater knowledge and understanding. A central theme of the VPH, and indeed of the International Physiome Project more widely, is that models and data should be accessible and re-usable. In order to achieve this goal, they need to be made available in software-independent formats which also incorporate the semantic information needed to understand and utilise the content.

A key focus of WP3 work is the promotion of this vision, encouraging the VPH-I projects to recognise the utility of standards, advocating the use of relevant common standards in each area (rather than groups developing their own formats), and facilitating the development of new model or data standards where none yet exist. Naturally, this is a long term process, requiring extensive community engagement, and iterative development to trial and refine emerging standards. It should be emphasised that WP3 is not pursuing formal standardisation processes such as those employed by ISO; rather a community-centric approach is envisaged, catalysed by the NoE, where different groups researching in the same or similar area work together to ensure their tools can communicate. This approach has been most successfully demonstrated in the SBML community.

Four areas in which standards work is needed are ontologies, data, modelling, and infrastructure interoperability. The latter three are discussed elsewhere, and ontologies are described below.

Ontologies

The importance of standardisation for interoperability is clear. However, this requires more than standardised formats (modelling, data etc.) and must extend to a standardised nomenclature that removes ambiguities in terminology and understanding across the VPH domain. This is particularly acute at the interface between domains (eg. where science meets the clinic). Consequently, an ontology strand is an important feature of NoE WP3 activity. Its role is to promote awareness of the need for ontological annotation, to encourage explicit allocation of ontology effort to VPH activities and, on a grander scale, unify competing ontological definitions that may be present within our domain.

The first of these strands has seen significant activity through links with the established ontology community (eg. OBO) and beyond, with promotion at seminars, and interactive sessions at workshops and conferences. These have deliberately covered a breadth of expertise, from sessions introducing the most basic concepts (dependents, continuants etc.) to erudite discussions on the definition of ‘disease’. This approach is intended to engage individuals across all levels of capability, and places the ignorant alongside experts for valuable VPH-style cross-fertilisation. The message promulgated by these forums is clear, namely:

- Ontologies can be considered an aspect of standardisation and are important to interoperability
- Coordinators and work-package leaders should consider allocating specific ontology effort to tasks within each project
- Bespoke ontology development should be avoided if possible. Rather, first seek to identify an established ontological term/identifier if it exists, and failing that, approach an appropriate ontology group that will accommodate your term(s) as necessary

Examples of ontology integration with VPH related activity include practical application, like ontology support for the Guyton model of renal hypertension. This has been a productive exercise that has highlighted the challenges of annotating physics-based physiological models, using unique ontology identifiers. Ontology integration closer to the clinical interface is underway at Sheffield (Section 4.3), demonstrating the practical hurdles that must be overcome if exposed clinical data is ever to be meaningfully interoperable with the tools of 3rd party researchers. Ontology efforts are seeing wider application through initiatives such as the recently funded (FP7 Call4) RICORDO project. Here, integration of models and ontologies (related to medical physiology) is a core aim that is intended to elicit interest from the pharma industry.

In all of the above examples, there are clear opportunities to contribute to the ToolKit. This could be in the form of documentation, clarifying best practice in ontological annotation for clinical application. Alternatively, contributions could highlight ontology developments at the interface between models and physiology, which might include resolution of competing ontological terms. There is also a need to bridge the many domains within OBO by a VPH infrastructure that provides for seamless ontological continuity. In some areas, new ontological concepts need to be accommodated, such as spatial/temporal localisation and associated hierarchies. Ultimately, ToolKit contributions are worthless without sustainability, and this depends on increasing integration with the open source ontology community (OBO). The interests of the commercial sector offer further incentive here, with benefits to be obtained by interoperability through effective ontology support. Finally, the ontology effort undertaken within the NoE provides expanding centres of experience and expertise that can further promote and help interested parties to obtain interoperability through standards/ontology integration. Such efforts will continue through the coming year and strengthen the ontology strand of the ToolKit.

Workflows

As with ontologies, workflows are gaining an increasingly important profile as the work of the NoE progresses. Workflows consist of modular, interconnected activities (often represented in schematic form) that offer a way of organising and characterising the collections of activities that contribute to a complex task. The process of constructing a workflow is useful in itself since it clarifies the sub-tasks that need to be undertaken as well as their temporal sequence and interdependencies. It also provides a convenient framework on which to hang other concepts, illuminating the nature of additional activities that are necessary for effective completion of the main task.

By way of example, the diagram presents a synthetic and idealised representation, indicating the modular interconnection of sub-tasks describing a clinical, therapeutic patient interaction. (It also illustrates a capacity to

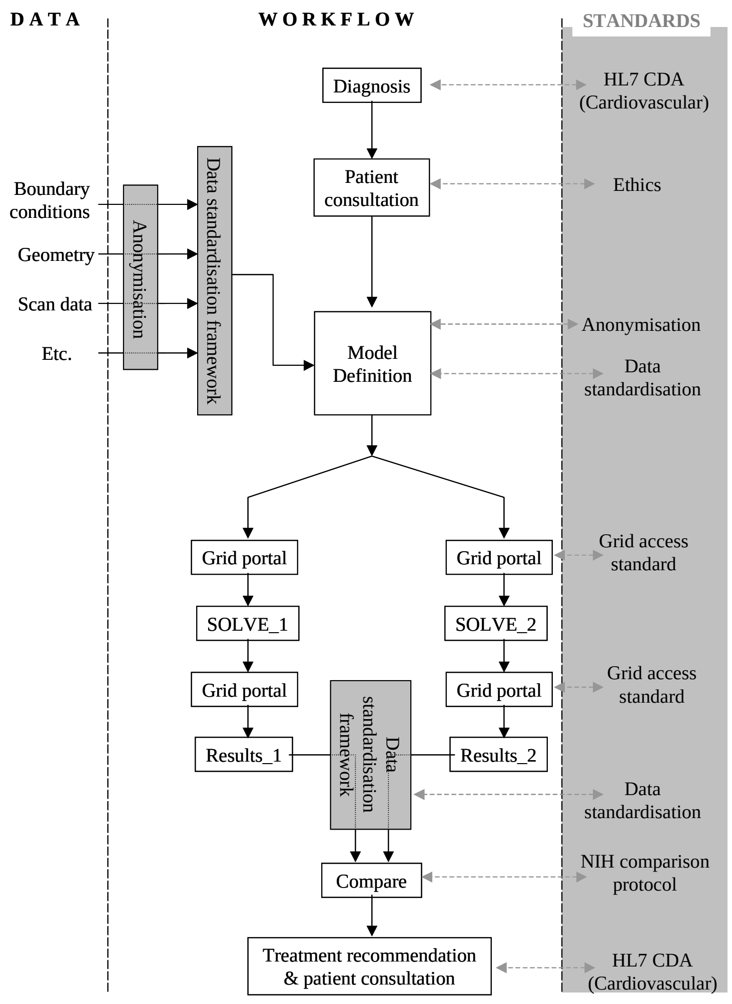


Figure 5. *An idealised patient-specific workflow illustrated in the context of data, workflows and standards*

highlight associated issues, such as data standards). The figure implies that patient referral has lead to numerous diagnostic tests which have resulted in a positive diagnosis related to cardiovascular disease.

The patient status at this point is most succinctly summarised by the clinical report, ideally formulated in accordance with a ratified standard supporting ontological annotation for that diagnosis (eg. HL7[*] Clinical Document Architecture (CDA)). This not only classifies the pathology, but is a mechanism for identifying relevant ontological associations and available data types (blood pressure etc.) that might be able to contribute to the patient specific simulation. Subsequent medical consultation with the patient leads him/her to consent to optimisation of treatment outcome through simulation. The model definition file for this patient-specific treatment scenario relies on data from many sources (including data from previous/other simulations) all of which have to be anonymised according to standardised anonymisation procedures, and the data translated to a standardised accessible data format via some form of standardisation framework.

Reduction to a standardised data type enables data to be readily imported into the model definition files so that the solver can be configured for computational analysis. The magnitude of the problem may well require Grid resources for acceptable solve times and Grid access is supported through appropriate, standardised Grid access protocols, which includes facilities for reservation, failover etc. and ultimately leads to production of a solver results file. For the purposes of engendering clinical confidence in the solution, it can be argued that multiple simulations of the same scenario should be run and their solutions compared. This requires that the data is passed again through the data standardisation framework to provide the results in a standardised format, amenable to visualisation by appropriate tools. The comparison of multidimensional data sets (ie. the independently computed flow solutions) will permit an assessment of the confidence in the solution provided by the simulations. The outcome will influence the proposed management of the patient, which will involve further consultation and incorporation of the important details into the evolving clinical report (CDA).

NoE Engagement Strategy

In pursuit of increasing levels of engagement, the soliciting of workflows from VPH projects provides a coherent thread by which to address topics such as standards and ontologies. End-user workflows within each project must be identified, enabling clarification of

[*] Dolin RH, Alschuler L, Beebe C, BironPV, Boyer SL, Essin D, Kimber E, Lincoln T, Mattison JE. The HL7 Clinical Document Architecture. J Am Med Inform Assoc. 2001 8:552–569.

requirements for ontology support and data standardisation. This will result in a growing repository of workflows, which, following further analysis, can be rationalised and provide improved classification of tools submitted to the ToolKit portal. Activities/modules common to multiple projects can be selected, with the intention of harmonising output and reducing unnecessary replication of effort, whereas under-represented activities and bottlenecks can benefit from redirected effort that will advance the causes of the VPH.

Ethico-Legal Developments

The environment for healthcare research involving the re-use of medical data is enjoying a welcome period of re-evaluation, with it now becoming routine for clinicians to obtain approval for the recording and retention of clinical data for possible future research purposes - subject to appropriate consent from patients. In some cases the consent process has even been permitted to be implicit in the attendance at clinics where appropriate notifying information is prominently displayed. Yet the VPH NoE's ambition, soon to be shared by the consortium that will be awarded the contract to construct the VPH Infrastructure, is to identify and cater for the widest possible range of ethical and legal requirements, from the strictest to the least onerous, and to respect each in accordance with the highest possible standards of care. Indeed so important is this area to the NoE that it has been accorded its own Guideline Topic in the catalogue of curation activities that are to receive coordinated attention from the standards working groups described earlier in this document.

The intention is for the Ethico-legal working group to consult widely with the VPH-I and Exemplar projects to identify the nature of the changing landscape, and it is additionally hoped to elicit the support of an ethico-legal team (preliminary discussions have already been held with the University of Sheffield's Department of Law). The goal is to draw up practical, useful guidelines for those wishing to share information (in either direction - donor or recipient) and to identify an adequately comprehensive set of parameters with which to tag data items so as to ensure the capability of completely faithful recognition and monitoring of data transfer according to the ethical and legal constraints in place.

It is appropriate to consider the nature of the cooperation that will be required with the Call 6 Infrastructure contract-holders. It is anticipated that the activities to be included in the Description of Work for the successful Integrated Project will have a significant element of activity in the area of traffic provenance, such that the IP Infrastructure systems will be designed from the start to cater for exactly the comprehensive range of metadata that the NoE is planning to investigate. It is hoped that this activity will form an early part of the IP's activities, so that the NoE may benefit from involvement with those funded activities, and can additionally contribute its own ideas into the IP's development processes.

The planned Guideline documents will be released regularly as improvement are introduced, and it is hoped that the first batch of releases will take place before the end of 2010.

Portal & Repository (NoE Tasks 3.6 & 3.7)

The main front-end for the ToolKit is the VPH ToolKit Portal (http://toolkit.vph-noe.eu/), an interactive website providing information and access to the ToolKit components. As has been described in earlier reports, this website lists a database of tools, methods, and services for the VPH community, to which members of the community are invited to submit new entries, and provide feedback on existing entries.

Registration on the website was opened to anyone at the time of the first annual review of the NoE, and we have seen a steady increase in the number of entries since, at a rate of one per month on average. As also discussed in Section 3.2, work is ongoing to source relevant high quality content from a variety of sources (see also Sections 5, 6, and 7). Thus far, efforts have focused on encouraging partner institutions to contribute in order to build up a core community, and hence achieve critical mass. Wider advertising has occurred at selected scientific conferences, including the VPH-I day in September 2009, the fourth International Conference on Computational Bioengineering (also in September 2009), and the UK e-Science All Hands Meeting in December 2009, from which a publication on the VPH ToolKit has arisen.

New features have also been added to the website since September (eg. the addition of guidance material on software licensing - see Section 3.3). New fields are also now available for entries in the database, as WP3 works towards improving the curation of entries. Extensive work on this is pending, dependent on the consolidation of working groups within WP3, and wider community consultation. Initial work has been carried out nevertheless:

- Each entry may now have a list of associated VPH projects that support the technology. This indicates to potential users, the level of support they can expect.
- Each tool or service can have a list of supported input and output formats, selected from a list of key VPH formats (see also the discussion of standards elsewhere in this report). This provides an indication of which tools could potentially be used in concert, due to the use of compatible formats for transferring data and models between them.
- Registered users of the site may also add themselves to the list of users of a technology.
- Many categories now list one or two key contacts within the VPH-NoE who have particular expertise in that area, so that users know who to contact for more detailed advice if required.

The aim for the portal, and the ToolKit as a whole, is to provide more than a simple collection of independent tools by offering methods and services, that create an interoperable suite of VPH technologies. A focus for the next stage of development will be to provide more detailed indications of which tools can successfully be integrated. This will begin through specifying the standards supported, so that it can be determined which tools will interact. The addition of license compatibility information allows developments to determine which tools could be combined in a new piece of software. Beyond this, we will be adding concrete examples of the use of multiple technologies in concert to tackle specific scientific problems. These exemplar workflows will be drawn primarily from EPs and the VPH-I, but also from related projects occurring in WP3 partner institutions.

CellML

The CellML model repository (http://models.cellml.org) currently contains over 450 CellML models, and these are freely available for users to download, simulate, modify, and upload.

From the onset of the project, several objectives for the CellML model repository were identified:

- Promote the sharing of models;
- Be publicly accessible;
- Provide a user-friendly interface to access, query, and store models;
- Provide a consistent workflow where models can be submitted, reviewed, validated, and published;
- Facilitate feedback on models;
- Provide a mechanism for advertising new models or changes to existing models.

There have been three different architectures of the CellML model repository software system, but only the current architecture (PMR2), which was released in July 2009, satisfies all these design objectives.

PMR2 Software

The CellML model repository is powered by the Physiome Model Repository (PMR2) software. This has been described in section 4.1. Also, as mentioned there, PMR2 will be used for a soon to be deployed (Mid 2010) FieldML repository. An alpha version of PMR2 for FieldML-precursor models will be available publicly in April 2010. The initial contents of this FieldML repository will be exnode/exelem files (see section 4.1 discussion on FieldML).

These will be converted to FieldML files once FieldML 0.2 is available (i.e. during the latter half of 2010).

Model Curation

The CellML curators supervised several student projects over the Southern hemisphere summer and the proportion of curated models has been considerably increased. Certain categories of models have been well curated, including cardiac electrophysiology, immunology, endocrine models, and cell cycle models. Further, there has been an effort to create an OpenCell session file, with an embedded SVG diagram, for each curated model, making the models more accessible to users who are unfamiliar with modelling.

In a move to replace the current curation star system with a set of more meaningful curation flags we have looked towards using MIRIAM as the minimum set of data required. In turn this has highlighted our need to annotate the CellML models with biological information, and also associate each model with a specific distribution term (two of the MIRIAM requirements). The former has led us to propose a model annotation scheme (based on MIRIAM uris) and collaborate with researchers at Newcastle University developing an annotation tool called SAINT. For the latter we have identified the Creative Commons Attribution license as an appropriate license for the model repository content.

BluePrint for Ontology-based Data and Model Integration

Mechanistic models in the VPH domain describe physiology knowledge in a formal and quantifiable manner. In particular, such models depict physiological processes in terms of the influence physiological variables have on each other. A significant proportion of variables in physiology models represent either some direct biological measurement or an inference calculated from these measurements. In most cases, variables take the ontological form of a quality (e.g. pressure, concentration, rate of some biological process, etc.) linked to an anatomical location.

Equations in mechanistic models formalise the precise effect that the value of one variable has on another. In such models, however, provision is rarely made to explicitly depict the anatomical connectivity that conveys the influence variables have on one another. Therefore, the integration of quality and location knowledge pertaining to VPH data (measurement) and models (inference) onto a communal ontological connectivity framework is an important step in providing VPH resource interoperability on a large scale.

The objective of the Use Cases below is to assist in the development of such a framework, in collaboration with the RICORDO project. The annotation pipeline emerging from this work will support the consistent representation of knowledge associated with VPH data and models. An illustration of this ontology-based annotation work that has been carried out on a clinical (cardiovascular) database schema, carried out by the EMBL-EBI and Sheffield partners, is also outlined in Appendix B.

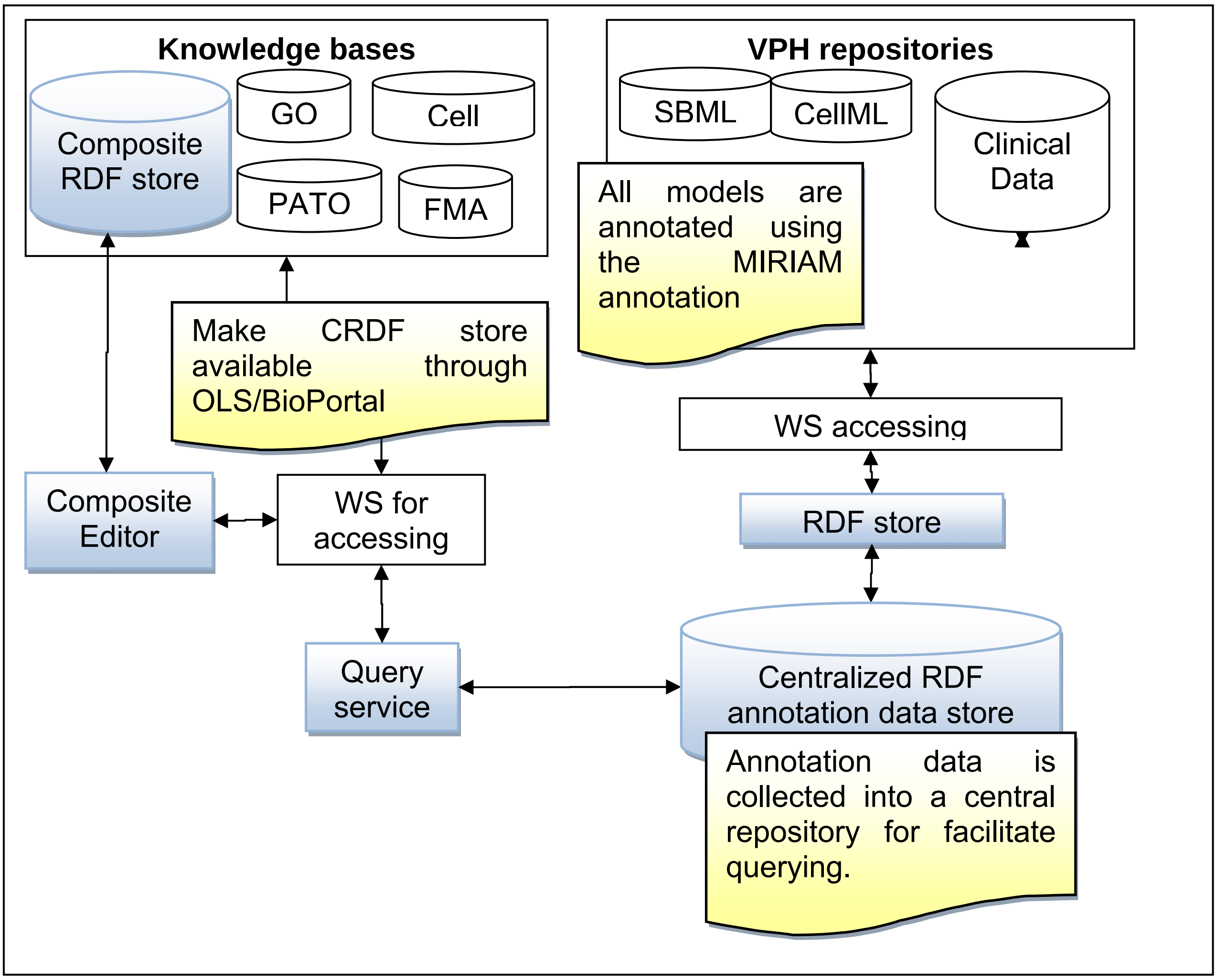


Figure 6. VPH Ontology-Based Infrastructure

The above figure illustrates the design blueprint for a VPH ontology-based infrastructure to integrate physiology models with clinical data. This plan is the result of the collaboration between the EMBL-EBI and Sheffield partners on clinical data, and with the Auckland partner on CellML model annotation. The key steps involved in the implementation of this plan are listed over the following Use Cases:

Use case 1: Creating composites

- A simple web based editor will be developed for creating composites. These composites will be stored in a centralized RDF store.
- The composites will be exposed to the community via the Ontology Lookup Service (OLS) and BioPortal. Using existing services will also enable us to introduce our knowledge base to the community easily. Users can continue to use their annotations schemes to annotate with composite terms.
- Composite editor will use the OLS/BioPortal to query knowledge bases. Using OLS/BioPortal will also help us to search for relevant composites when users want to create composites of composites.

Use case 2: Storing annotated data

- *The aim is to have an up-to-date centralised RDF store of this annotated data. This will enable us to carry out speedy querying of the data. Storing the data in RDF will also support complex querying.*
- *Existing MIRIAM specification needs to be extended to support composites. These MIRIAM* annotations will be stored in the RDF store. This will allow us to query both composites annotations and simple annotations.
- Using MIRIAM will enable us to annotate other types of data such as models coded in programming languages (*e.g.* MATLAB), clinical database schema etc. It may be necessary to introduce an XML schema for representing clinical database schemas which can then be annotated using MIRIAM.
- A package 'RDF store manager' will be developed for accessing web services supported by the external repository providers to update the central RDF store.

Use case 3: Querying annotated data

- A query service application will be developed for querying the RDF data store. The query service application will interact with OLS/BioPortal to access the knowledge bases for executing similarity searches.
- A set of algorithms/RDF queries needs to be developed for searching the RDF annotated data store and sorting queried data.
- External repository providers must register their web services in the query service application. This will enable 'RDF store manager' to locate the external repositories that need to be queried for annotations.

- An interface will be developed to display queried data. This will help users to interact with the queried data.

Use case 4: Annotating models

- Existing annotation editors such as SAINT supports OLS and MIRIAM. This will allow users to use existing tools to annotate SBML and CellML models with composites.
- Currently, there are no methods or tools for annotating MATLAB models or clinical database schemas. It may be useful to explore possibilities for extending SAINT for annotation support.
- The solution described in this document requires users to use two different tools; composite editor for creating composites and annotation editor for annotation. It may be useful to integrate these two tools together.

Content Source 1: Exemplar Projects

The authors of the VPH NoE wished to ensure that a thorough and comprehensive appreciation of the complexities of tool provision was obtained, the better to enable the NoE teams to interact with external tool providers. It therefore arranged for internal development of tools to take place, ensuring that content would emerge, and that the complexities of enabling content to be generalised, accessible, well-documented and highly usable were experienced first-hand.

The NoE Exemplars have produced – and continue to produce – a rich variety of material that is being prepared for release via the ToolKit portal, and this section describes the activities that are taking place.

SeedEP1 A multi-organ Core Model of arterial pressure and body fluids homeostasis (CNRS)

Long-term regulation of arterial blood pressure (BP), is determined by the balance between fluid and salt intake and excretion, characterised by the relation between BP and urinary output and manifested in the renal function curve (RFC). This serves as the basis for understanding normal BP regulation and the treatment of hypertension and was developed by Arthur Guyton who based it on quantitative control-theory, firmly grounded in experiment.

For the VPH a core modelling environment (a CME – the ANR-funded SAPHIR project) inspired by the Guyton models is being built, implemented as a set of replaceable open source modules (heart, kidney, lung, muscle mass, and nervous and hormonal regulatory systems). This seed EP involves:

- Integration or link-up of the SAPHIR quantitative parameter database (QKDB) with a more generic VPH DB (to be implemented under WP3),
- Contribution of the Multiformalism Multilevels Simulation Library® (M2SL, developed at LTSI INSERM, Rennes), to the VPH ToolKit collection of numerical solver packages,
- Specification of inter-module I/O connection protocols,
- Markup language versions of the various modules (in DAEML and also, in most cases, in CellML)
- Cross-discipline ontology integration (and development where necessary).

In short, a prototype modelling environment, with an example "core model".

Progress and Outputs

Topic	Status
Model markup	All modules marked up in CellML. Submodules tested. Several versions of the Guyton model are fully operational: (Fortran, Simulink, some in Berkeley Madonna, the main modularised version in M2SL). The working markup language within the project is DAEML, a CellML offshoot. Two GUIs are in final development (one an aid for DAEML markup, the other for running M2SL simulations. Markup and model simulation are now separated.
Data	• Solute concentrations, flow rates, gas partial pressures, hematocrit etc • Compartment volumes; • Anatomical dimensions, counts of multiple structures (nephrons); • Electrical potentials • Biophysical/chemical data such as constants, viscosities etc. • Signals (time-series) of many kinds (ECG, respiratory, BP…) Lack of standardisation means that SeedEP1 is therefore very keen to participate in the data standards development of the VPH ToolKit.

Topic	Status
Ontologies	• Anatomical labels in the QKDB database of kidney parameters submitted to the FMA (EBI's VPH reference ontology initiative) • An OBO ontology of variables and parameters in the Guyton and Ikeda models contributed to EP6 (EBI Guyton-reference ontology EP) • Text-mining/analysis tool to aid entry of physiological data into QxDB-type databases, transferable to other domains.
Demonstrator	An interactive version of the multi-organ systems model of blood pressure regulation (the Guyton model) will be provided, running under M2SL. The GUI is currently in the final phases of development.
Specific contributions and interactions	SAPHIR modelling environment will be explicitly adapted to the protocols of the VPH ToolKit: • Integration/link of the SAPHIR quantitative parameter database (QKDB[17]) with a more generic DB called QxDB now available. • Contribution of the Multiformalism Multilevels Simulation Library® (M2SL, developed at LTSI INSERM, Rennes), to the VPH ToolKit collection of numerical solver packages.

ToolKit Ultimate Content

Item	Name	Description	General	ToolKit
Tools	M2SL	Multiformalism Multilevels Simulation Library® solver	Yes	June 2010
	QxDB	Generic package to build a parameter database, based on QKDB (Apache/PHP/MySQL)	Generic	June 2010
	DAEML	Model markup language, with XSLT converters for M2SL, Berkeley Madonna…	Generic	June 2010
	GUI for DAEML	GUI for model entry and conversion	Generic	June 2010
Models	Guyton72	Open source working implementations of Guyton 1972 model of blood pressure regulation, in Fortran, C++ (for M2SL), and Simulink	The original Guyton model, as described in the landmark 1972 paper	May 2010
	Guyton92	Open source working implementations in Fortran, C++ (for M2SL). NB: a CellML version of the individual modules is available at cellml.org	The standard shared version of the Guyton model	May 2010
	G92 with GUI	User-friendly interface to run G92 (or other models) in M2SL	The GUI is generic in the sense that it allows one to run any model (with M2SL solver)	June 2010
	IKEDA	Open source implementations (Berkeley Madonna, C++ for M2SL) of IKEDA 1979 multi-organ model of acid-base regulation		
Data	QKDB	Parameter database for kidney modeling	Applies to Kidney	Yes

17 http://physiome.ibisc.fr/qkdb/

SeedEP2

Integrated multi-level modelling of the musculoskeletal system (ULB)

Seed EP2 will define the modelling needs of the musculoskeletal field and will communicate these to the WP3 ToolKit development team. It will then develop an ontology to allow further integration of the musculoskeletal models with other sub-systems (e.g., the nervous system and both cardiac and vascular systems). Available ICT tools will be considered for integration into the VPH ToolKit to answer common modelling needs. Solving sophisticated problems requires not only multi-level integration, but also the development of multi-organ and multi-tissue modelling techniques based on robust optimisation algorithms and advanced visualisation tools. Such an approach has been the focus of several EC-funded projects (VAKHUM, Multimod, LHDL), whose main result is the availability of a shared ICT technology, MAF2, that aims to perform the required integrative research.

Progress and Outputs

<table>
<tr><th>Topic</th><th>Status</th></tr>
<tr><td>Model markup</td><td>CellML describes models using systems of ODEs and nonlinear algebraic equations. EP2’s work, which includes the description of spatial transformations and deformation of geometrical objects uses PDEs. UoA’s FieldML, designed for encoding spatial fields, is not yet sufficiently advanced, but being based on XML (like the native format of the MAF library, used by the Seed EP2, communication between MAF and FieldML should be facilitated. Discussions to encode and handle transformation between objects for various types of joint motions are underway, with the objective of allowing input/output exchange between models created within a MAF environment and FieldML models.</td></tr>
<tr><td>Data: Morphology</td><td>
<table>
<tr><th><u>Topic</u></th><th><u>Format</u></th><th><u>Comments</u></th></tr>
<tr><td>Medical imaging</td><td>DICOM 3.0 and 4.0</td><td>From Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) scanner</td></tr>
<tr><td>3D bone models</td><td>SURF, STL, VRML, VTK</td><td>Basis of ULB database</td></tr>
<tr><td>Musculoskeletal model</td><td>XML (native format), ASCII</td><td>Models including 3D bone models + geometrical dissection information of muscle fibre + spatial transformation</td></tr>
<tr><td>Muscle weight & volume</td><td>xls (Excel file)</td><td>Muscle morphological parameters</td></tr>
<tr><td>Muscle tissue biomechanical data</td><td>XML (native msf format), Excel</td><td>Data obtained from dissection (origin, insertion, tendon length, muscle length, pennation angle)</td></tr>
<tr><td>Virtual spatial location of skeletal landmarks</td><td>XML, ASCII</td><td>Required for registration towards subject-specific dimensions</td></tr>
<tr><td>Manual location of skeletal landmarks</td><td>XML, ASCII</td><td>Required for registration of subject specific morphology</td></tr>
</table>
</td></tr>
</table>

<table>
<tr><th>Topic</th><th>Status</th></tr>
<tr><td>Data: Kinematics</td><td>
<table>
<tr><th>Topic</th><th>Format</th><th>Comments</th></tr>
<tr><td>3D marker trajectories measured with a motion analysis system</td><td>C3D, ASCII, CSV</td><td>C3D is the de facto standard in motion analysis</td></tr>
<tr><td>Anatomical landmark trajectories</td><td>C3D, CSV</td><td>Trajectory of ALs during a particular task</td></tr>
<tr><td>Joint angle data</td><td>ASCII, CSV</td><td>Measured joint angle data, processed from AL trajectories, of subjects performing tasks. No proper standards exist.</td></tr>
<tr><td>Reconstructed motion model</td><td>XML</td><td>Models including 3D bone models + AL locations + motions + local anatomical frames</td></tr>
</table>
</td></tr>
<tr><td>Data: Kinetics</td><td>
<table>
<tr><th>Topic</th><th>Format</th><th>Comments</th></tr>
<tr><td>Ground reaction force</td><td>C3D</td><td>From gait analysis</td></tr>
<tr><td>Electromyography</td><td>C3D, MSF</td><td>From gait analysis</td></tr>
<tr><td>Foot pressure</td><td>RsScan</td><td>From RsScan foot pressure plate</td></tr>
</table>
</td></tr>
<tr><td>Ontologies</td><td>An improved functional mapping of the FMA ontology to address key physiological connections between anatomical systems (cardio-vascular, nervous) and the musculoskeletal system is underway. This ontology will allow the description of anatomical and physiological links between the systems. To develop this, new classes of objects must be developed since they are not available from the FMA ontology (for example, joint motion, muscle excitation, nerve potential of action, etc.). A first draft of the ontology including all new classes of physiological parameters required for musculoskeletal modelling has been developed in Protégé and is now under revision by NoE Partner EBI.</td></tr>
<tr><td>Demonstrator</td><td>Muscle spasticity is a highly handicapping pathology, and current therapeutic schemes focus on regular rehabilitation and training of the patients' musculoskeletal systems, with no facility for patient-specific customisation. EP2 is creating a VPH tool (lhpFusioBox) to allow such customisation.
EP2 is also prospecting ontology development, to build bridges between the musculoskeletal system physiology and the cardio-vascular system.</td></tr>
<tr><td>Specific contributions and interactions</td><td>SeedEP2 has a clinical focus on the development of a shared technology to produce infrastructure for improved musculoskeletal modelling. The main goal is the definition of the required data processing and data visualisation tools. As it extends previous work using the Multimodal Application Framework (MAF) that already integrates many routines required for MS modelling it is likely to be an important contribution to the VPH ToolKit.

Looking further ahead, this EP is also an important scientific driver for ensuring adequate consideration of the requirements for linking anatomical systems to other physiological modes. The MS system is linked particularly to the nervous and cardiovascular systems, and work is beginning, to determine how models of these systems might interact. Enabling such research will be a key aim for the VPH ToolKit.</td></tr>
</table>

ToolKit Ultimate Content

Item	Name	Description	General	ToolKit
Tools	lhpFusionBox	This software allows fusion of highly heterogeneous data such as palpation data on subject, motion data, in-vitro data, electromyography, foot pressure data, medical imaging. All data can be visualized independently, or after fusion, in 3D interactive windows through various viewers. Data can be processed and represented according to International recommendations (for example, motion data). Other processed data (such as muscle moment arms) can be determined, displayed and stored.	Some of the tools are general (for example, the data fusion tools, medical imaging viewers,etc) and should of interest to many anatomical systems. Other tools are more specific to the musculoskeletal community (e.g., muscle moment arms, etc).	The software could be distributed as such (after licencing checking). Some parts of the code (C++) could be distributed as well (after licencing checking).
Models	-	3D bone models (in various formats and resolution), finite element models	General for the musculoskeletal community	Yes, if required. Available directly (from PhysiomeSpace)
Data	-	20 full body CT-scans Anatomical landmark locations Gait analysis data. Full simulation of registered gait analysis to medical imaging.		Yes, if required. Some of them are available directly (from PhysiomeSpace). Others will be brought within one year.

The Vertical and Horizontal Atherome (WHAM) (KI)

SeedEP3

Our limited mechanistic understanding of atherosclerosis - the identity of the disease-related genes, polymorphisms, proteins, and their interactions, within and between cells and organs – is a limiting factor in developing drugs. The 'WHAM' program seeks to achieve both vertical and horizontal modelling integration. Available vertical data ranges from molecular information (gene-expression, protein-protein interactions data) to angiograms (imaging) reflecting the degree of disease from patients undergoing bypass surgery. Horizontal molecular information is obtained from tissue biopsies (liver, muscle, fat, affected and unaffected aorta). Similar data-types are available from the aorta from a mouse model prone for atherosclerosis. Molecular data from macrophages, a key cell-type involved in atherosclerosis, are also available. To analyse these data from different organs and model systems of different complexity, a blend of pattern detection techniques (statistics, machine learning), network identification algorithms, and mathematical modelling is required.

VPH tools enabling 'simple' things like storing different types of data (patient descriptions, experimental protocols, expression, SNPs etc) are urgently needed. At the other end of the spectrum, the issue of integrating the molecular information obtained across different model systems and imaging is most likely a central methodological problem for a VPH ToolKit addressing clinical needs. This includes visualisation but also how molecule X affects the tissue and 3D properties etc, interaction between flow (blood) and expression of various molecules.

Progress and Outputs

Topic	Status
Model markup	The SBML code for the WHAM model and a brief description on how Copasi was used to generate the SBML code have gone to the WP3 team. The markup was not trivial due some complications with the functions for this model, which is quite different from the usual metabolic models for which SBML was designed.
Data	An example of a gene expression dataset and sample output from the simulation model will be uploaded to the WP3 ftp server
Ontologies	It is not straightforward to assign/map an ontology to the SeedEP3 model. For example, an edge in the model (transforming process/bioentity A to B) should point to one/several terms in an ontology such as GO. The problem is how to do this in a systematic way in practice; this step appears nontrivial since it depends a lot on model understanding, and would thus be difficult to automate. The problem is being submitted to the SBML team
Demonstrator	A demonstrator of the visualisation of the dynamics of the atherosclerosis process based on the simulation model is under development.

Topic	Status
Specific contributions and interactions	**Exploring parameter and model uncertainty using optimisation algorithms for vertical systems – from genes, cells to composite tissues.** There are at least four hard conceptual and computational core problems: 1. Integration of data and models between different vertical (from gene to phenotype) 2. Integration of data and models between different horizontal levels (between entities such as organs, cell and molecules). 3. For any given model capturing a biological process at a given level, we are faced with the challenge of model uncertainty – the problem of deciding which elements (nodes, processes) and interactions (edges) that should be included in a model. 4. Finally, for a given model structure there is a challenge (iv) to identify one or several sets of parameters which are consistent with available data. We have developed a workflow approach that should be relevant to all VPH research projects aiming for causal understanding except those projects not mechanistically integrating across levels (i.e. statistical models) or exceptional cases where biological knowledge can uniquely determine parameters and model structure. The workflow uses a time-dependent computational model capturing the plaque growth in an organ (the arterial wall), involving several cell-types and biological processes down to the level of receptors, genes and proteins. • Review the state-of-art of biology of plaque growth • Formulate a time-dependent mathematical model of the plaque formation where the interactions between elements are generically parameterised • Search the parameter space using a particle swarm optimisation (PSO) algorithm. Here we use physiological constraints and data from biological experiments to define seven evaluation functions (including lesion size) which therefore produce a fitness landscape corresponding to the “feasible sets of parameters” • Analyse (statistics and machine learning algorithms) and identify different qualitative behaviours within the feasible sets of solutions from the simulations • Formulate robust core predictions on the tissue (plaque level), cell, and gene level which are based on groups of solutions • Perform appropriate experimental validation. From the ordinary differential equation model, containing 8 state-variables and 52 parameters, we computed more than 10000 different sets of parameters and approximately 10% were considered feasible parameter sets. Surprisingly, we find that the behaviour of a high-dimensional (parameters) vertical model of a complex disease reduces to only three qualitatively different dynamical fingerprints. The underlying reason is that that the existence of several vertical layers (tissue, cell-to-cell interactions, genes) in essence strongly constrains feasible input-output conditions for each vertical model level. In this specific sense, the integrative (1) vertical VPH challenge simplifies the (4) parameter challenge. This finding makes the intrinsic parameter complexity in the system caused by our incomplete knowledge tractable for research. The model can now be used for predicting how to effectively reduce the size of the plaque at different points in time.

ToolKit Ultimate Content

Item	Name	Description	General?	ToolKit
Models	WHAM	SBML model of WHAM	Atherosclerosis	Complete
		Ontological model annotation		Underway at EBI
		Gene Ontology process predictor		Uploading soon
		Demonstrator		Uploading soon
Data		Gene expression data for validation		Uploading soon

SeedEP4 — Multi-scale simulation and prediction of the drug safety problems related with hERG (IMIM)

An important field of application for the VPH concept is drug development.

To obtain useful *in-silico* predictions of the efficacy and safety of drugs, we require models sensitive to the differential molecular characteristics of the drugs, coupled with models simulating the biological system. The hERG-related cardiac adverse effects of drugs are a paradigmatic example of this approach. The vast majority of drugs associated with pathological prolongations of the QT segment of the electrocardiogram are known to interact with the hERG potassium channel. Simulation at the molecular scale using atomistic approaches will enable quantitative predictions of the effects of drugs on the electrophysiological parameters of hERG, with the aim of predicting the change in the QT segment generated by the drugs under study.

This approach is based on tools developed in several projects that are focused on the multiscale processes modelling and their computational implementation (PS3Grid and EC-STREP QosCosGrid) as well as on the translational research aspects of such a multilevel problem (EC-STREP BioBridge). This seed EP will aim to integrate existing software tools dealing with the several levels of complexity of the QT elongation. The expected outcome will be standardisation of formats for easy integration of simulation across scales.

Progress and Outputs

Topic	Status
Model markup	The CellML model of the activation potential used in the demonstrator, will be available from the demonstrator page starting from mid November.
Data	To be specified
Ontologies	SeedEP4 has agreed with EBI to send a detailed text description of the computational experiments performed in order to try to encode it in the MIASE markup (Minimum Information About a Simulation Experiment[18]). This markup seems still immature to serve for the operations required in computational biology

18 http://www.ebi.ac.uk/compneur-srv/miase/index.html

Topic	Status
Demonstrator	The ultimate aim of this EP is to produce accurate predictions of the cardiotoxic properties of new compounds, in particular, those related with the QT elongation. We are developing models that describe the phenomena involved at molecular and electrophysiological level, integrating them for producing a single prediction which aims to be more accurate that the results provided by the state-of-the-art methods. A web-based prototype server has been built (http://lamia.imim.es/HERG_server.php) able to reproduce the action potential in an automatic way. The user introduces the structure in a standard format and a prediction of the compound properties at molecular level is produced, the output of which is used as an input to an electrophysiology model. Each of the models is rather sophisticated, but the complexity of the process is hidden. Although only a prototype this illustrates the two main key goals: • The system can be used to obtain safer compounds • Models built to describe phenomena at different levels (receptors, cells and tissues) can be integrated to provide a better depiction of the reality than those provided by any single model.
GRID/HPC	SeedEP4 has used HPC resources to equilibrate the initial molecular structure of hERG, and has made extensive use of a distributed grid infrastructure via the GPUGRID.net project for molecular dynamics simulations. The UCL AHE and GSEngine workflow tools are appropriate for this project
Specific contributions and interactions	• The team has developed a variety of tools in molecular-level VPH modelling, particularly distributed molecular dynamics and molecular systems biology simulations. • The BioBridge portal also has much to contribute to the VPH ToolKit by providing an environment for the integration of clinical to -omics data, making the inclusion of molecular data an easier possibility for other modellers. • The project hopes to benefit from markup language expertise: o The elegant description of multi-cellular models which is currently handled by an adaptation of SBML files but that would benefit from the interaction with CellML o The use of general markup languages for force field definition, already being explored in the IMIM partner in the form of FFML, where links to FieldML will also be considered in the future o Metadata standards for simulation results.

ToolKit Ultimate Content

Item	Name	Description	General	ToolKit
Tools	Pentacle	Advance software to perform 3D-QSAR studies.	Applied in medicinal and computational chemistry research.	It is in the Toolkit.
	ACEMD	Accelerated molecular dynamics code	Computational biophysicists	Basic version available from www.acellera.com
	GPUGRID.net	Distributed network for molecular simulations	Computational biophysicists	Resource available to research groups
	RBoinc	Virtual access to distributed computing projects	Distributed scientific computing	Embedded in BOINC open source middleware
Models	Endo.cellml Mcell.cellml	CellML models of the action potential in a single guinea pig cell employing the Hodgkin-Huxley formalism.	Applied in computational electrophysiology research.	Available in the demonstrator webpage.
Data	hERG-dofetilied all-atom docking trajectory	An all-atom MD trajectory in which the dofetilide drug docks into the hERG channel	Computational biophysicists, drug discovery	Uploaded to VPH-NOE's SFTP server

SeedEP5 Modelling and visualising brain function and pathophysiology (ERCIM, Digital Patient Working Group)

SeedEP5 models brain function based on clinical data in order to understand the causality of brain diseases such as epilepsy, dementia, schizophrenia, and alcoholism.

- At the first functional level, linear and nonlinear synchronisation methods are applied to study neuronal dynamics (increasingly recognised as important in cortical and sub-cortical activity integration). Synchronous oscillations of certain types of such assemblies in different frequency bands relate to different perceptual, motor or cognitive states and may be indicative of a wider range of cognitive functions or brain pathologies.
- At a second level, source estimation models and graphical theoretical measures are applied to describe and understand the functional characteristics of brain networks.

The project also investigates brain tumours (especially glioblastoma) and normal brain tissue behaviour at the cellular and higher levels of biocomplexity. Such models will be individualised, therefore requiring pertinent image analysis, data processing and visualisation techniques in order to extract the necessary information, which will be the input to the cancer simulator. In particular, image analysis tasks (such as image registration fusion, segmentation, etc) will be applied at different scales (e.g. tissue 3DMRI images, microarray data, etc).

This seed EP will make use of generic VPH ToolKit components ‘fine-tuned’ to meet the brain's specific needs:

- Web-accessible repositories for data, annotations, patient information etc.;
- Model solutions to the inverse or forward brain source localisation problems;
- Patient-specific customisation of models;
- Data fitting
- GUIs tailored to the visualisation of causal and functional relations between different brain lobes.

Progress and Outputs

Topic	Status
Model markup	Because we do not operate at the level of systems biology, we do not have a brain model and we do not have a model of a biological process, CellML and SBML are probably unsuitable. However our interest in topic behaviour in the area of a dipole or an electrode has led us to investigate GraphML (http://graphml.graphdrawing.org/) as a mechanism for describing our brain network graphs using an ML.
Data	EEG data is held in ".edf" files, a broadly used format. Other data includes: **'.set' files**, which is the format accepted by the DIPFIT toolkit of EEGLAB in MATLAB. They contain the EEG measurements: - at different time point - at different channels - for different individuals and - for several repetitions of the experiment **The MRI image** (Colin 27 brain is used according to MNI http://imaging.mrc-cbu.cam.ac.uk/imaging/MniTalairach) with the retrieved dipoles located in it **A '.txt' file** with the exact locations of the dipoles (coordinates)
Ontologies	The FMA is sufficient for anatomy. For the *connections* between separate nodes of the brain network, it is difficult to create an ontological structure for our graph data. Expert advice has been to combine the ontology and the GraphML Markup Language information.
Demonstrator	Although the non-applicability of some tools (SBML, DICOM…) prevents complete representation, a demonstrator using certain VPH ToolKit features will be ready during Q2 2010.
GRID/HPC	We have been using DEISA, and more specifically the HECTOR supercomputer, with satisfactory results – faster execution. The UCL AHE and GSEngine workflow tools are appropriate for this project
Specific contributions and interactions	SeedEP5 is concerned with the detection of the sources of a specific brain activity in the brain interior, recorded by electroencephalography (EEG). Outputs include: • Dipoles (position, direction, sense and magnitude) are identified such that the reported electrode activity can be equivalently provoked by these dipoles' activity. • The findings should help in better describing and understanding the functional characteristics of normal brain networks, by fusing these results to complementary MRI brain scans. • Visualisation of results, based on MRI scans can be provided • The source localisation algorithms may be fused with other the Image Fusion ToolKit visualisation components

ToolKit Ultimate Content

Item	Name	Description	General	ToolKit
Tools	BrainNetVis	BrainNetVis is an application, written in Java, that displays and analyzes synchronization networks from brain signals. The program implements a number of network indices and visualization techniques.	This tool is not for general use. It is expected to be used by neuroscientists, doctors and bioinformaticians.	It will be uploaded very soon.
Models	Brain networks	The networks are modelled as graphs and we use algorithms and techniques from graph theory.	It can be used easily by experts.	It will be uploaded very soon.

EP6 Establishing ontology-based methods for the VPH ToolKit to improve interoperability between data and models: the Guyton case study

This Exemplar Project will investigate an interoperability framework for physiology models and gene expression data. In particular, this EP sets out to achieve the following goals:

- The design and implementation of representational classes for the annotation of anatomy and physics in physiology model parameters and gene expression data;
- The demonstration and verification of such an approach via the annotation of
    - The parameters of a classic model that represents blood pressure regulation (namely, the Guyton model), implemented in the SAPHIR project (from the VPH NoE seedEP1) and marked up in the CellML repository
    - Related human gene expression datasets in ArrayExpress that correspond to the anatomical locations these parameters address;
- The contribution of the outcomes of this work to the VPH ToolKit effort, as a reference example of the role of a communal anatomical and physics framework for resource interoperability.

Strategic co-operation between the OBO and VPH networks to co-ordinate ontology-based efforts has been agreed. Curational effort has been identified for the annotation of

- anatomical, cellular, molecular and physics concepts of model parameters and variables, as well as
- mapping to the gene expression datasets related to those locations.

There is no doubt that these curatorial aspects are considerable, ranging from

- consistency checking of those ontology terms utilised for annotation (e.g. crucial discussions are ongoing with Onard Mejino and Stephen Sammut regarding consistency issues pertaining to the FMA),
- establishing protocols for term requests and amendments,
- assessing the optimal way to share the annotations between the different incarnations of the Guyton model implementations, as well as ArrayExpress.

For the latter step, it will be useful to explore the solutions afforded by the SemSim/SemGen work, as well as the BioModels and CellML/openCell repositories, amongst others. The Guyton exemplar project will therefore need to

- Focus on both curational and representational issues that link together both biological models and data
- Compare different technologies that are of relevance to encoding and sharing such annotation.

ToolKit Ultimate Content

Item	Name	Description	General	ToolKit
Tools				
Models				
Data	Annotated model	Ontology map to physiology variables	Development of general annotation method	Early 2011

Content Source 2: VPH-I Projects

Engagement with the VPH-I projects has begun, with all twelve original VPH-I research activities (IPs and STREPS) having provided information on their activities, timescales, personnel and intended outputs. Work is also underway to obtain similar information from the five projects recently funded under Call 4 from the VPH funding stream.

In tandem with this engagement, the NoE Content Working Groups are now formalising the specifications and processes required for project outputs to contribute to the ToolKit, thus allowing the content to be deployed by other users as easily as possible. It is acknowledged that in some cases outputs might be sufficiently specialised that introduction to the ToolKit may require excessive effort that may not be justified by the likely user interest. The following pages outline the contributions from the VPH-I community, highlighting sustainability aspects important to the VPH.

ARCH

Introduction

ARCH is concerned with developing patient-specific computer models, based on images of the blood vessels, for improvement of vascular access in kidney patients on haemodialysis therapy. This is a response to issues relating to vascular access in patients suffering from kidney failure on haemodialysis. The principal challenge is vascular access (VA), typically accommodated through the introduction of a 'shunt'. VA dysfunction is the major cause of morbidity and hospitalisation in haemodialysis patients and is related to the haemodynamics of shunt placement. ARCH is developing patient-specific computer models for optimisation of shunt placement, based on images created from the blood vessels in the arm. Modelling tools are being designed to simulate the long-term adaptations of the vasculature and predict the function of the shunt for improvement of surgical planning and arteriovenous fistula management. Experimental validation is an integral part of the programme.

Outputs

ARCH is developing image processing tools that incorporate vessel segmentation and network graph editing. Vascular network models are being developed (0D/1D) and these are being extended to accommodate vascular adaptation. These modelling tools will be fused together as a unified computer application, which will be provided to clinical end users for improved management of haemodialysis patients.

Community

The ARCH community has its roots in the ARCH consortium which comprises 8 partners with academic, clinical and industrial backgrounds: Academisch Ziekenhuis Maastricht (Netherlands), ESAOTE Europe B.V. (Netherlands), Istituto di Ricerche Farmacologiche "Mario Negri" (Italy), Philips Electronics Nederland B.V. (Netherlands), Philips Medical System Nederland B.V. (Netherlands), Universiteit Gent (Belgium), University of Sheffield (UK), Univerzitetni Klinikni Center Ljubljana (Slovenia). The use and development of open-source software based on existing widely adopted libraries (VTK, ITK, Slicer, vmtk) secures a growing community that could be further strengthened by links with the VPH/NoE. Open source communities offer a route for sustainability (vmtk, Slicer, ITK, VTK) and some of these - National Alliance for Medical Image Computing (www.na-mic.org) – are supported by substantial numbers. Key aspects of their success relate to the availability of non-restrictive licenses (business friendly, e.g. BSD), openness, inclusiveness and solid documentation.

Sustainability and the ToolKit

The use of vmtk provides opportunities for sustainability through community development. For ARCH, sustainability includes a wide range of concepts such as community-building, standards establishment and adherence, ontological annotation, interoperability and workflow suitability, 'generalisability', ethical and legal progress, infrastructure-building, aspects of reproducibility, documentation-thoroughness and maintainability. ARCH is contributing in the following ways:

- The image analysis concepts have already contributed to the Vascular Modelling Toolkit (www.vmtk.org)
- A 3D CFD solver will be released as open source within the same initiative
- A modular, object-oriented 0D/1D vascular network solver framework will be released as open source at the beginning of the final year of the project.

In this context, a link with the VPH/NoE toolkit would benefit all partners and encourage a wider adoption of the tools. Data representation for vascular networks, meshes and simulation outputs has been formulated in XML to promote interoperability. The ARCH FTP client is a feature of infrastructure support and is available for wider use beyond the immediate confines of the project (it has already been adopted by euHeart). However, ARCH has no specific plans or commitments for maintenance of the tools after the project is completed, other than open-source community driven efforts.

The NoE is seen as a route to attract critical mass, perhaps through providing guidelines and a platform for exchange to maximise integration. This could help to unify diverse tools and documentation (follow NITRC example, http://www.nitrc.org/), supported through source code repositories.

ARCH	**Description**	**Comments**
Project Outputs	VA shunt management and simulation. A unified application supporting vascular network, adaptation, modelling and supporting tools.	
Content relevant to the ToolKit	Image processing and vmtk exploitation gnuid (3D CFD solver) pyNS (vascular network solver) archFTP client Standards – Data as XML	Sustainability through open source community involvement
Benefits of NoE involvement	Critical mass Data exchange infrastructure Source code repositories The UCL AHE and GSEngine workflow tools are appropriate for this project	

ARTreat

Introduction

The focus of ARTreat is the cardiovascular system, particularly atherosclerosis and its progression, leading to life-threatening events. A patient-specific computational model will be developed to improve management of the condition and associated quality of life. The end-user applications will provide environments for clinical decision support and training. The former will include tools to assist clinical cardiologists with personalised treatment selection complemented by real-time, on- the-fly advice for invasive interventions, such as stent positioning. More effective treatment is a recipe for reduced therapeutic cost, likely to be achieved through personalised treatment support. The same patient-specific model is also suitable for adaptation as a real-case simulator training tool. This is intended to support and augment the skills-development of clinical cardiologists in training

Outputs

ARTreat will rely on a patient simulation that works across three levels: (i) description of the 3D geometry of the arterial tree, (ii) modelling of blood flow and blood particle dynamics and (iii) accommodation of biological processes that lead to the creation and progression of atherosclerotic plaques. This will be coupled with advanced clinical support tools for plaque characterisation, and the discovery of new knowledge - associations among heterogeneous data - that can improve the predictive power of the patient-model. The medical expert is also in a position to contribute to the accumulated knowledge which can further inform the models through an adaptive patient-specific computational tool.

Community

ARTreat encompasses 18 partners, with strong representation from industry: Universita Degli Studi di Parma, Consiglio Nazionale Delle Ricerche, D.D Synergy Hellas Anonymi Emporiki Etaireia Parochis Ypiresion Pliroforikis, World Match Limited, Agfa Healthcare N.V., Sorin Biomedica Cardio S.R.L, Euro (PMS) Ltd, Intercon Sp. Z O.O., Universite Paris Descartes, Advanced Simulation and Design GmbH, The University of Cambridge, Foundation for Research and Technology Hellas, Tecmic-Tecnologias de Microelectronica SA, Lanza & Thompson Information Technology SRL, SAS Software Ltd, Technologica EOOD, Computer Sharing Bucuresti SRL, Univerzitet U Kragujevcu. Wider community interaction will come through training and commercial interest in the products.

Sustainability and the ToolKit

A suite of tools will be generated by this project, many of which have relevance to the NoE ToolKit. Decision support tools include automated 3D artery reconstruction from IVUS and angiography, with 3D arterial tree reconstruction (based on MRI/CT) supporting stented and unstented arterial models for the study of haemodynamics. Tools are also available for plaque characterisation, incorporating aspects of blood molecule dynamics and factors relating to wall adhesion. Further clarification comes from experimental data and development of a data mining application. In combination, these tools will contribute to a treatment support application and an interventional decision support application. Reconfiguration of the tools for training purposes is a cost effective method to further exploit the technology, whilst encouraging engagement with the community at large.

Members of the ARTreat consortium include organisations that are leaders in the areas of (i) medical IT and (ii) cardiovascular and arterial surgery. These companies will exploit the ARTreat applications to add new solutions to their product range, while academic and IT company partners will benefit from experience of the new generation of patient-specific healthcare tools. Links with the NoE offer wider exposure and uptake of these tools. The commercial context may ultimately benefit from developments in standards and awareness of regulatory affairs that the NoE could provide.

ARTreat	**Description**	**Comments**
Project Outputs	Patient-specific cardiovascular simulation for atherosclerosis and predication of associated life-threatening events. Application environments for decision support and training.	
Content relevant to the ToolKit	A suite of tools… Arterial geometry reconstruction Stented and unstented arterial haemodynamics Molecular dynamics Plaque characterisation Data mining application Training environment	Sustainability through community training and commercialisation opportunities
Benefits of NoE involvement	Wider exposure to user community (clinical and technical) Standards and regulatory affairs awareness	

ContraCancrum

Introduction

The focus of ContraCancrum is the development of an advanced multiscale simulation platform that can describe tumour growth and response to treatment. Its application is intended to optimise disease treatment procedures in a patient-specific manner. Real clinical needs are informing its development and it is intended that clinical translation will be in the context of clinical trials/tests. Two dedicated studies, one on lung cancer and one on gliomas, will access the ContraCancrum oncosimulator and evaluate its possible use in clinical practice. Validation of the integrated simulation system will reference clinical and other medical data which will be used to adapt and optimise the simulation appropriately. The ultimate goal is wider uptake in treatment practices in cancer therapy.

Outputs

With this simulator the clinician will be able to perform *in-silico* experiments corresponding to different candidate therapeutic scenarios for any cancer patient in order to facilitate and better substantiate his or her treatment decisions. Several tools will be available: 3D Image analysis, meshing and modelling. The timetable for developments of key components includes:

- Initial version of the models (non imageable tumours, imageable tumours and normal tissues) available at the project mid-point
- (Semi-) automatic tumour and normal tissue segmentation (lung and brain); optimised version of the TBINM (tissue level models + biomechanics + imaging + molecular models) integrated multi-level tumour simulator; fine tuning and optimisation of molecular models.

Community

The project involves numerous partners, spanning academia (University of Bedfordshire, University of Berne, University of Edinburgh, University College London), industry (FORTH) and the clinic (Saarland University Hospital). The cancer community is a large and well established force in medicine and ContraCancrum is a significant player in the research/translational area, with links extending internationally (eg. Harvard, USA). Creation of tools using resources like VTK ensures support through a natural community of technical enthusiasts.

Sustainability and the ToolKit

Suitable tools for contribution to the VPH/NoE ToolKit include automatic, smooth, voxel-based meshing implemented with the VTK libraries. Additionally, DoctorEye is a complete DICOM image analysis, segmentation/annotation, 3D visualisation, open platform for cancer image analysis and modelling. Complementary to this is DoctorsCompare, a cancer imaging data viewer that enables comparison of DICOM files and their annotations and displays the differences between them. A successor to ContraCancrum is already envisaged, continuing in a new project called TUMOR. This will focus on developing cancer model repositories and their seamless workflow-based interoperability with the already established CViT at Harvard. This Centre for the Development of a Virtual Tumour (https://www.cvit.org/), is a community of investigators with interests in the biomedical, computational and mathematical aspects of cancer research. In league with the NoE, it is the hope of such initiatives that more cancer modelling related activities can be brought together to generate a new community. Common goals/tools/services are invaluable in helping to drive developments and collaboration for future benefit.

Contra Cancrum	**Description**	**Comments**
Project Outputs	Oncosimulation for tumour growth and treatment response. A suite of tools for image analysis, meshing and modelling.	
Content relevant to the ToolKit	Voxel tools with and vtk exploitation DoctorEye DoctorsCompare Standards – DICOM	Sustainability through TUMOR project. Links with CviT.
Benefits of NoE involvement	Critical mass; wider community exposure Standards for tools and services The UCL AHE and GSEngine workflow tools are appropriate for this project	

euHeart

Introduction

euHeart is concerned with personalised and integrated cardiac care. Its focus is patient-specific cardiovascular modelling and simulation for in silico disease understanding and management, and for medical device evaluation and optimisation. The challenge of the euHeart project is to directly address Cardiovascular Disease by combining novel ICT technologies with integrative multi-scale computational models of the heart in clinical environments to improve diagnosis, treatment planning and interventions.

Outputs

Three key activities characterise the work, namely

- Integrated Virtual Heart – The curation and database recording of validated/verified model components
- Personalisation – Application of multi-scale and multi-physics modelling techniques to human data
- Integration – Achieving productive collaboration across disciplines on a common system (the heart)

This involves (i) development of euHeartDB (a database populated by 11 consortium models; anatomical content and bio-physical benchmarks; aortic flow model 'benchmarks'; FDA CFD model and solutions), (ii) production of CellML/FieldML models, (iii) development of a Biophysical Personalisation Library including model personalisation 'benchmarks' (i.e. estimation of physical and physiological parameters like the contractility of the muscle, the conductivity of the tissue, etc.…) and (iv) a user-friendly interface implemented through GIMIAS (including incorporation of problem specific GIMIAS plug-ins and integration of cmGUI).

Community

The consortium contains a mix of academic leadership, clinical sites, and industrial partners, in total combining eighteen industrial, clinical and academic partners, whose collective goal is the development of individualised, computer-based, human heart models. This represents a community of over 100 people, many of whom are active within the VPH community, who together form a productive base from which to promote the community ethos of the virtual physiological human. The VPH-community effort is fully supported by euHeart's desire to share best practice in model construction and development, and euHeart is a strong advocate of the NoE's stance on markup languages for the sharing of models and tools.

Sustainability and the ToolKit

The main outcome of euHeart will be an open source framework for the description and representation of normal and pathological multi-scale and multi-physics cardiovascular models, using international encoding standards (CellML, FieldML). In addition, a library of innovative tools for the execution of the biophysical simulations, the personalisation of the models and the automated analysis of multi-modal images are developed (OpenCMISS, SOFA, OPENFEM, LIFEV). Visualisation will also incorporate open-source resources (GIMIAS, CMGUI).

Aspects of sustainability are reflected by a commitment to development and application of standards. The CellML models for the lumped-parameter modelling of the circulatory system will be curated and maintained in the CellML repository. They can be applied by researchers in the community to provide improved boundary conditions for future CFD studies. Sustainability is also supported by euHeartDB which requires population of the data base from content by the community, with the intention that it would be inclusive of other projects. Furthermore, the 'benchmark' problems based on the FDA's CFD challenge and on carefully described and validated personalised aorta models, will be available from euHeartDB, in standardised ML formats, thereby providing a lasting legacy for the VPH community. Finally, adoption of GIMIAS helps by building a community around common objectives, which includes the standardisation of workflow specifications. Ultimately, sustainability comes through the use of tools by the community. euHeart is collecting evidence of clinical benefit and the quantification of potential impact through clinical validation (e.g. w.r.t. modelling power, prediction).

euHeart	**Description**	**Comments**
Project Outputs	Person-specific virtual heart simulation environment. A unified interface for heart simulation, supported by database, biophysical personalisation library.	
Content relevant to the ToolKit	Open source framework for multiscale/multiphysics modelling Tools library based around… OpenCMISS, SOFA, OPENFEM, LIFEV Visualisation based around… GIMIAS, cmGUI Database - euHeartDB Standards – CellML, FieldML	Sustainability through interaction with open source communities.
Benefits of NoE involvement	Exposure of tools to wider community Promotion of standards	

HAMAM

Introduction

HAMAM is an acronym for a breast cancer diagnosis project (Highly Accurate Breast Cancer Diagnosis through Integration of Biological Knowledge, Novel Imaging Modalities, and Modelling). It is a response to the fact that early detection and accurate diagnosis of breast cancer are still unresolved challenges. Currently, a variety of imaging modalities exist to identify and characterise morphology and function of suspicious breast tissue, but an effective, multi-modal solution for breast imaging is still missing. The purpose of HAMAM is to support these modalities with tools that can augment diagnosis.

Outputs

HAMAM requires the building of tools to integrate datasets / modalities into a single interface. This includes provision of pre-processing/standardisation tools that will allow for optimal comparison of disparate data, the building of spatial correlation information datasets to allow for new similarity and multimodal tissue models. Adaptability needs to be incorporated to allow for the integration of other sources of knowledge such as tumour models, genetic data, genotype, phenotype and standardised imaging. These elements in combination will be used to support detection and diagnosis of breast cancer. Data sets will become available in 2010, with the first prototype of workstation planned for early 2010. Segmentation, registration, pattern recognition, data mining methods will follow.

Community

The 9 partners include The European Institute for Biomedical Imaging Research (AT), University College London (UK), MEVIS Research (DE), MEVIS Medical Solutions (DE), the Swiss Federal Institute of Technology (CH), Raboud University Medical Centre (NL), University of Dundee (UK), Charite Medical University Berlin (DE), Boca Raton Community Hospital (USA). The HAMAM community extends to clinical users and incorporates a clinical advisory board. Wider community involvement may result from commercialisation efforts envisaged within the programme.

Sustainability and the ToolKit

The open XNAT toolkit supports the middleware layer of the HAMAM infrastructure, and interoperability is sought through XML structuring of database and patient data.

HAMA can also offer the following as potential contributions to the NoE ToolKit.

- Data sets – Multimodality breast imaging cases with annotation. For demonstration, training, and evaluation.

- Solutions to multimodality - image registration, computer aided diagnosis, segmentation, modelling.

The datasets will be openly available and the multimodality solutions will be published and partly further developed and integrated into *Workstation,* which is a development prototype that is headed for commercialisation. It is intended that some of the solutions may develop their own life as open source projects. Added value, sought by connection with the NoE, would include tools for high volume data storage and annotation, tools for public data exchange and batch/cloud computing facilities.

HAMAM	**Description**	**Comments**
Project Outputs	Multimodal breast cancer diagnostics. Single interface encompassing a suite of tools for multimodal tissue models, image processing and data analysis.	
Content relevant to the ToolKit	Open XNAT toolkit. Multimodal processing tools. Multimodal data sets Standards – XML structuring of database and patient data.	Sustainability through commercialisation of *Workstation*
Benefits of NoE involvement	Tools for high volume data storage and annotation. Infrastructure for data exchange and batch/cloud computing facilities.	

IMPPACT

Introduction

IMPPACT describes a solution for Image-based Multi-scale Physiological Planning for Ablation-based Cancer Treatment. It recognises that radiologists cannot accurately predict results of the Radio-Frequency Ablation procedure, and currently no training simulator exists for preparing radiologists to conduct and optimise RFA treatments. Any such tool must include understanding of processes during tissue heating and the resulting death of the cells. IMPPACT has approached this challenge from two perspectives, namely (i) multi-scale data analysis and validation and (ii) computational modelling and simulation of the RFA process. These tools will enable prediction of treatment results on a patient-specific basis, thus bringing down the risk of local recurrences and eliminating the need for repeated treatments. Ultimately it will enable treatment by RFA to be as effective as resection.

Outputs

The major objectives of this initiative are three-fold:

- Patient-specific computational modelling and simulation of the RFA process.
- Interventional planning system and augmented reality training simulator for interactive planning and visualisation of RFA treatment.
- Comprehensive model validation in histological experiments and clinical practice.

This will be provided through an Interventional Planning System (IPS) and a Physiological-Organ-Simulation-Toolkit (POST).

Community

The work involves a consortium that consists of the Institute for Applied Information Technology (Germany), Graz University of Technology (Austria), Medical University of Leipzig (Germany), Medical University of Graz (Austria), Helsinki University of Technology (Finland), University of Oxford (UK), NUMA Engineering Services Ltd (Ireland). Wider community interaction comes through clinical feedback and opportunities afforded through the training simulator. Other interaction comes through Biomed Town, web-based exposure etc.(eg. Medical Visualization @ ICG @ TU-Graz).

Sustainability and the ToolKit

The simulation environment will incorporate a modified bio-heat equation for appropriate modelling of cell heating/death which has relevance to the ToolKit. Microscopic findings will be interpreted as macroscopic equations and complement the modelling to provide patient specific planning. Results will be validated at multiple levels, involving the use of images compared against human physiology, and visual comparison of simulation and treatment results in a range of studies throughout treatment. Hence, tools are being developed for image-based reconstruction of liver/tumour/necrosis structures from multiple scans and histology images. The whole edifice will employ a user-centred software design approach, intended to guarantee suitability of the solution for clinical practice. In combination, the Interventional Planning System (IPS) and the Physiological-Organ-Simulation-Toolkit (POST) rely on interactive simulation of the RFA process, with tools to support fast computation, real time visualisation and manipulation.

IMPPACT provides patient-specific modelling of tissue properties that is a significant step forward in comparison with current interventional planning systems, but sustainability will only come with the support of an enthusiastic community. The NoE may be able to promote wider interest and help with critical mass. Further opportunities are afforded by the augmented reality training simulator, supporting the surgical community through training with the IPS.

IMPPACT	**Description**	**Comments**
Project Outputs	Modelling/simulation of radio-frequency ablation. Tools for treatment planning and visualisation with validation through histology/clinical practice.	
Content relevant to the ToolKit	Interventional Planning System (IPS) Physiological Organ Simulation Toolkit (POST) Real time visualisation Thermal modelling	Sustainability through community and clinical involvement
Benefits of NoE involvement	The UCL AHE and GSEngine workflow tools are appropriate for this project	

NeoMARK

Introduction

NeoMARK is an ICT initiative for prediction of cancer reoccurrence. NeoMark is interested in integrating heterogeneous clinical, laboratory, molecular and imaging data in a multiscale and multilevel modelling environment for prediction of neoplastic reoccurrences, with a particular interest in oral cancer. This approach will be applied to the monitoring of disease after remission, for early identification of local or metastatic reoccurrences. The project will be validated in two primary Clinical Centres in Spain and in Italy. In this phase the early exploitation of NeoMark will also be assessed through the use of a RT-PCR platform as a means of developing highly individual diagnostic tests to be used both at the time of first diagnosis, as well as for reoccurrence identification.

Outputs

Technical developments are oriented towards the following:

- Creation of a database of oral cancer patients (clinical data, genomic data and imaging data at multiple points in time)
- Development of data analysis tools for assessing cancer reoccurrence risk, risk factors and to identify relevant features
- Development of tools for storing and managing patient data (Database, Web Interface, Handling of sensitive data)
- Development of an image feature extraction tool (Imaging features of tumours and lymph nodes from CT and MR images)
- Development of a real time PCR chip to eventually replace the time consuming and expensive gene expression analysis

The database of heterogeneous data of oral cancer patients supported by tools for heterogeneous data entry should be available in June 2010, whereas risk factors, gene expressions and trained Reoccurrence Prediction algorithms are due at the end of the year. A real-time PCR chip is also under development.

Community

Project partners comprise Universitaria di Parma (Italy), Fraunhofer Gesellschaft Zur Foerderung der Angewandten Forschung E.V. (Germany), STMicroelectronics S.R.L. (Italy), Fundacion MD Anderson International España (Spain), Link Consulting – Technologias de Informaçao S.A. (Portugal), Universidad Politécnica de Madrid (Spain), MultiMed S.R.L. (Italy), Panepistimio Ioanninon (Greece), PLANET AE (Greece). Wider collaboration involves links with a range of hospitals, demonstration of tools and the provision of research data to the oral cancer community.

Sustainability and the ToolKit

With respect to the VPH the salient outputs of NeoMARK relate to a trained algorithm for oral cancer reoccurrence prediction supported by information relating to oral cancer risk factors. This will include a database for heterogeneous data for oral cancer patients and relevant gene expressions for the real time PCR chip. The latter offers replacement of comprehensive gene expression analysis by the PCR chip and can be used in other applications with different gene expressions. A software toolset is also being created for entering and storing heterogeneous data (clinical data entry, image feature extraction) and will be available for incorporation into the VPH/NoE ToolKit. These are designed to support oral cancer reoccurrence prediction within the clinical routine and are available to be adapted for use as tools in different applications. For instance parts of the Image Feature Extraction Tool could be reused to assess the response of lymphoma patients to therapy.

It is hoped that the exposure through the NoE will promote beneficial links with existing ontology developments and encourage wider uptake of the NeoMark approach, with other communities recognising that it can be applied to similar diseases, if suitably adapted data is provided.

NeoMark	**Description**	**Comments**
Project Outputs	Prediction of oral cancer reoccurrence based on data analysis (database), image processing and real-time PCR.	
Content relevant to the ToolKit	Trained algorithm for cancer reoccurrence prediction. Oral cancer database Software for data entry and analysis Image feature extraction Real-time PCR chip	Sustainability through adoption of tools in the wider community.
Benefits of NoE involvement	Promote wider uptake (technology translation to other areas – lymphoma) Ontology developments	

PASSPORT

Introduction

PAtient Specific Simulation and PreOperative Realistic Training for liver surgery: a project that is designed to support liver surgery in the context of acute liver dysfunction. Eligibility for liver surgery is based on the minimum safety liver volume remaining after resection (standardised FLR), but this minimum value varies over time and from one patient to another according to biological and mechanical properties of the liver. PASSPORT provides a patient-specific modelling environment that combines anatomical, mechanical, appearance and biological preoperative modelled information in a unified model of the patient. This complete "Virtual liver" is being developed in an Open Source Framework allowing vertical integration of biomedical data, from macroscopic to microscopic patient information. Dynamic liver modelling will provide a patient-specific minimum safety standardised FLR, which when coupled with the preoperative planning simulator, will inform strategies for optimised patient outcome.

Outputs

The underpinning open source framework for surgical simulation development (SOFA - www.sofa-framework.org) was up-and-running by 2008. Developments within 2009 have included an Open Database (containing a 3D model of the patient for algorithm comparison + the VR-Render viewer) and FW4SPL (an open source framework for computer assisted surgical software development). Virtual planning and anatomy software are also planned for release in the next 18-24 months. As a whole, the whole software suite will provide an environment to support surgical management of patients suffering from (imminent) liver failure.

Community

The community associated with PASSPORT has it origins in the 9 partners of the consortium. This comprises Eidgenoessische Technische Hochschule Zuerich (Switzerland), Technische Universitaet Muenchen (Germany), Imperial College of Science, technology and Medicine (United Kingdom), Institut National de Recherche en Informatique et en Automatique (France), Universitaet Leipzig (Germany), University College London (United Kingdom), Université Louis Pasteur (France), KARL STORZ GmbH & CO. KG (Germany), Institut National de la Santé et de la Recherche Médicale (France). Close links with WebSurg provide a substantial, ready-made community that is supported by access to the VR-Render WLE (software for 3D-visualisation of patients and DICOM images). Research and development interest extends to SOFA/vtk/itk, providing links with a technical community that can influence developments. These communities will come together for the first time in two events, scheduled for 2010. (i) The First LIVIM (Living Imaging) congress: www.livim2010.eu (September in Mulhouse France) and (ii) the First International Congress on simulation and serious game (October in PARIS France).

Sustainability and the ToolKit

The strength of the Websurg and open source communities is an important factor in PASSPORT sustainability (the Second International Congress on simulation and serious game - Strasbourg France - is already being planned for 2011). Furthermore, the software tools are ideally suited for inclusion in the NoE ToolKit. *VR-Render* is a 3D image viewer that is now available as freeware. *VR-Planning* provides for 3D surgical planning, and there is some debate as to whether it should be provided free-of-charge. *VR-Anat* is due to be released in 2011, and this will be a commercial product. In all cases, software has been developed from a new open source framework (http://code.google.com/p/fw4spl/), with the *VR-software* and *Fw4SPL* supported by IRCA.

Complementary to this is the SOFA framework (www.sofa-framework.org) which is based on a strong network and financially supported by INRIA. This is also open source, and enables the creation of complex and evolving simulations by combining new algorithms with algorithms already included in SOFA. It is primarily targeted at real-time simulation, with an emphasis on medical applications. Participation in numerous workshops has led to successful links with other VPH projects (e.g. Health-e-Child).

Open databases are also a feature of PASSPORT (IRCADb) with the first characterising 20 segmented livers, the second cataloguing respiratory anatomy/function and the third characterising the parathyroid (see www.ircad.fr/softwares/3Dircadb). IRCAD also supports MEDIC@ as a free, on-line **Med**ical **I**mage **C**omputerised @nalysis service.

It is hoped that links with the NoE can further our sustainability goals. This includes the need for standardisation of models (3D mesh, mechanical) and effective ontological annotation. It would desirable for Fw4SPL and SOFA to work and link with ITK and VTK and it would be invaluable to expose the 3D anonymous patient database. However, this requires that ethical/legal hurdles are overcome, with access supported by a secure, VPH-wide data sharing infrastructure. Finally, quality guarantees are required, based on reproducibility, documentation-thoroughness, maintainability etc.

PASSPORT	Description	Comments
Project Outputs	Framework for surgical simulation (liver) development. Database supported by tools for visualisation/planning.	
Content relevant to the ToolKit	VR-Render image viewer VR-Planning VR-Anat Open source FW4SPL Open source SOFA Open databases Standards - DICOM	Sustainability through community involvement - WebSurg, SOFA, vtk/itk, conference events etc. VR-Anat commercialisation
Benefits of NoE involvement	Model standardisation (eg. meshing) Ontology annotation Possibility of itk/vtk and SOFA/Fw4SPL cross-fertilisation? Exposure of databases to wider community	

PreDICT

Introduction

The goal of preDICT is to provide an efficient computational framework for the study of possible adverse cardiac side effects in the early stages of drug development. Many drugs fail to reach market because of side effects on the heart. preDiCT aims to create an open environment comprising validated computational models, tools and numerical methods that will enable simulations of drug actions on the electrophysiology of the human heart.

- Development of models describing the interaction of drug molecules with ion channels (WP3)
- Models and tools for understanding differences between species (WP4)
- Whole-heart simulator based on Chaste software and associated tools (e.g. PyCml) (WP5)
- Data warehouse supporting experimental, literature and simulation data with associated ontologies and GUI for data entry (WP6)
- Models of the heart incorporating realistic structure (WP7)
- Determination of sensitive new biomarkers for drug toxicity (WP7)

Outputs

Anticipated outputs for the coming year revolve around a Virtual Research Environment (VRE) encapsulating the following:

- Software for analysing species differences at the cell level
- Validated ion-channel models
- Myocardial tissue models for rabbit and human
- Highly scalable Chaste heart simulator for HPC (see also section 4.1)
- Data warehouse
- Protocols and algorithms to simulate drug-induced alterations in electrophysiological behaviour at the whole-organ level and in the ECG

Community

The work involves a consortium that consists of the University of Oxford, GlaxoSmithKline, Novartis, Roche, Universidad politecnica de valencia, Universitas scientiarum szegediensis, Aureus Pharma, Centre for Advanced Studies Research and Development (Sardinia), and Fujitsu. Community-based activities revolve around the developing exploitation plan, supported by links with the NoE and include:

- Consultation with pharmaceutical companies to determine VRE requirements
- Pharma companies represented on Scientific Advisory Board
- Pharma Workshop (October 2009)
- Building other relationships

Sustainability and the ToolKit

Models developed within preDICT support best-practice formats (CellML, etc.) and best-practice ontological annotation. The simulation code is open-source (LGPL) and architecture-independent with modularity to allow for specialised hardware. The resulting Virtual Research Environment supports reproducible *in silico* experiments (workflow, data capture, etc.) and is business-friendly open-source and fully documented (particularly with the needs of the end-users in pharmaceutical companies in mind). Note that long-term sustainability will depend heavily on adoption by pharmaceutical companies and others. Factors that can influence this include decisions like centrally hosted versus deployed VRE. This raises issues such as security of commercially sensitive data, end-user access through firewalls, and complexity of deployment in industry. Legacy data is also an issue. This might be an opportunity for the NoE to act as a hosting portal for data and administration.

PreDICT	**Description**	**Comments**
Project Outputs	Virtual Research Environment (VRE) for simulation of drug action on electrophysiology of the heart.	
Content relevant to the ToolKit	Open source cardiac simulation – CHASTE Validated ion channel models Myocardial tissue models (meshing) Standards – ontological annotation, CellML, FieldML	Sustainability linked to commercial interest. Hence acute awareness of the needs of pharma. Release 2.0 in April 2010; further will follow. Models will be placed in the CellML repository. Rabbit mesh available from Chaste website.
Benefits of NoE involvement	Aspects of security Legacy data Infrastructure for data and administration	

PredictAD

Introduction

The last two letters (AD) of PredictAD refer to Alzheimer's Disease. Alzheimer's disease accounts for 60-70% of all dementia cases. It is without cure, and reliable early diagnostic techniques are lacking. This debilitating disease causes long and oppressive suffering to patients and their relatives with considerable costs to society. About 25 million people suffered from dementia in 2000, and a 4-fold increase in this number is expected by 2050. The objective of PredictAD is (i) to find the best combination of biomarkers for AD diagnostics from heterogeneous data (imaging, electrophysiology, molecular level, clinical tests, demographics) and (ii) to develop clinically useful tools for integrating efficient biomarker results. In the face of increasing AD prevalence, this is an appropriate response to an era of new drugs and prevention strategies. Now is the time to exploit the vast potential of information hiding in heterogeneous patient databases to solve the AD diagnostics problem.

Outputs

Early diagnosis and progress monitoring of AD is a central element of disease management in the context of emerging drugs and prevention strategies that will become available. There is a strong indication that different biomarkers provide a reliable and early indication of AD prior to its major clinical signs. However, optimal early diagnosis requires information from a combination of different biomarkers to be used in a clinically useful way. Consequently, comprehensive biomarker discovery techniques and rigorous statistical models are being developed using the large databases of the consortium. The accuracy and usability of models and tools will be clinically evaluated, complemented by a cost-effectiveness study of heterogeneous data in AD diagnostic procedures.

Community

The PredictAD community has its roots in the consortium members: VTT Technical Research Centre of Finland (Finland), GE Healthcare Ltd (UK), Nexstim Oy (Finland), University of Kuopio (Finland), Imperial College London (UK), University of Uppsala (Sweden), University of Milan (Italy), Rigshospitalet (Denmark). Wider interaction involves collaboration with ADNI (Alzheimer's Disease Neuroimaging Initiative), with an arrangement to use data in exchange for analysis results. Negotiations for analysing data from other cohorts is underway.

Sustainability and the ToolKit

The goals of PredictAD are to generate new knowledge of biomarkers characteristic to Alzheimer's disease and compile information about the cost-effectiveness of various patient measurement procedures. Central to this is the production of a software tool for early detection of Alzheimer's disease and for progress monitoring. A prototype software solution is due to be ready in late 2010, followed by a final software solution in May 2011. Particular issues relating to this effort are standards, ontological annotations, workflow suitability, all of which is complicated by challenges associated with database and data file format(s).

There is an opportunity to contribute software to the VPH NoE kit thus enabling the analysis of potentially relevant and heterogeneous data from other groups. A key aspect of relevance to the VPH NoE is integration with clinical practices, and these again raise challenges associated with standards, ontological annotation, workflow suitability, ethical & legal. It would be useful if some project effort could be reserved for standardisation and integration work, linking with relevant authorities and other interested parties to create applicable standards and interfaces (e.g. standardisation of the database and data format used in the project software tool). NoE links can also hopefully assist with solutions that contribute to community building, infrastructure, etc.

PredictAD	**Description**	**Comments**
Project Outputs	Database combined with biomarker discovery and statistical models for Alzheimer's Disease (AD) diagnostics.	
Content relevant to the ToolKit	Heterogeneous data and software tool for early detection of AD.	Sustainability through community links (ADNI)
Benefits of NoE involvement	Database and data file format standardisation. Recommendations for integration with clinical practice. Community building and support.	

VPH2

Introduction

VPH2 refers to the Virtual Pathological Heart of the Virtual Physiological Human. Heart failure accounts for almost a quarter of all admissions to hospital for cardiovascular events and is associated with high mortality (median survival ~ 18months). The project will develop a platform to assist with management of ischemic heart disease, helping the cardiologist and the cardiac surgeon to characterise the severity and extent of disease in patients with Left Ventricular Dysfunction (LVD), with or without Functional Mitral Regurgitation (FMR). Specific computational methods will support clinical decision-making and the planning of optimal treatment, all of which will be ultimately validated against patient data. Improved patient outcome and reduced healthcare costs are anticipated.

Outputs

The computational framework will bring together three elements:

(i) The integration of clinical, biological and genetic data retrieved from medical records and laboratory results. This aspect is designed to support surgical and medical decision-making.

(ii) Advanced simulation and modelling tools based on MRI dataset analysis. Intended to promote better selection and simulation of specific surgical procedures for cardiac surgeons.

(iii) A new framework and multimodal approach, useful in other relevant clinical scenarios.

Community

The consortium comprises a mixture of private companies, universities/research labs and

public administrations/healthcare institutions. *Private companies:* NoemaLife (data/system integration), SCS (imaging, visualisation, integration), SORIN (system evaluation; exploitation and dissemination), Q&R (data mining), INTERCON (system integration), PATMOS (data mining, modelling). *Universities/Research Labs:* WWU (genotyping), Bedfordshire (visualisation), EPFL (modelling), CTI (data mining, modelling), POLIMI (modelling), IFC/CNR (MRI data analysis; system testing). *Public Administrations/Healthcare Institutions:* RegLom (user requirements, system assessment and validation). Sustainability comes through community interest and support and for this purpose a VPH2 website is active. VPH2 also has a presence in BioMedTown to increase awareness of the project in the VPH community and attract wider interest. A commitment to using MAF for visualisation also helps to secure longer-term sustainability.

Sustainability and the ToolKit

The project will deliver several elements of relevance to the NoE ToolKit. The data mining tool, will relate clinical, biochemical, genetic and image-based markers to ischemic heart progression. This will accompany tools for cardiac function assessment from MRI data with additional software for the prediction of cardiac surgery outcome. The VPH2 framework will underpin all aspects, including genomics, modelling, data mining and integrated treatment planning tools, all operating under a user-friendly interface.

More specifically, the functional assessment tool will permit biomechanical assessment of both left ventricle performance and that of the mitral valve from MRI data. The functional predictive tool will enable the scoring of different therapeutic strategies for the left ventricle (restoration, revascularisation, resynchronisation, regurgitation correction). This will enable assessment of the effect of annuloplasty on mitral valve performance. Finally visualisation support will enable the user to visualise results alongside patient data (integration and visualisation are based on the Multimod Application Framework (MAF)).

With respect to sustainability, possible exploitable results include:

- A diagnostic/prognostic tool for cardiologists and cardiac surgeons to score patients and select different therapeutic strategies for HF
- A simulation tool for (new) medical device testing
- A cardiac data visualisation tool
- A stand-alone left ventricle and mitral valve assessment tool
- A stand-alone left ventricle and mitral valve predictive tool
- A data mining module

Links with the NoE could offer numerous advantages, including the opportunity to explore the potential of other modelling tools (perhaps less accurate but time saving) or integration of VPH2 tools with other models for more diverse application (coronary bed stenting procedures, peripheral circulation, models of heart - peripheral organ interaction in disease)

VPH2	**Description**	**Comments**
Project Outputs	Virtual Pathological Heart simulation for the VPH. Tools to assist with management of ischemic heart disease (LVD, FMR) supported by a unified framework/interface.	
Content relevant to the ToolKit	A suite of tools that include… Data mining Functional assessment of left ventricle Functional assessment of mitral valve performance Predictive tool for scoring of therapeutic strategies Visualisation tool	Sustainability through community awareness, (BioMedTown, MAF). Potential (commercial) exploitation of tools.
Benefits of NoE involvement	Availability of alternative modelling tools Integration of VPH2 tools with other tools for more diverse applications	

VPHOP

Introduction

VPHOP (Osteoporotic Virtual Physiological Human) is a multiscale modelling exercise, providing tools to fight osteoporotic fractures. The fatalities from OP fractures equal those of breast cancer, and it is clear that the technology in current clinical practice is inadequate - accuracy in predicting fractures can be as low as 60%. VPHOP is developing a technology suite for Predictive and Personalised (P2) medicine for osteoporosis. The predictive aspect considers multiscale models representing the skeletal mechanobiology from the whole body down to the molecular constituents, simulating the skeletal loading in various conditions and predicting if the bones will fracture in each of them. The personalised component will configure the multiscale environment for personalised judgements using specific patient information. This includes the flexibility to accept a spectrum of information, so that the more information that is available, the more personalised the model becomes.

Outputs

The outputs of VPHOP are many, but those of particular relevance to the NoE and the VPH ToolKit are provision of community building services (Biomed Town), provision of digital library services (Physiomespace), MAF (the multimod application framework v3, and front-end applications) and a software layer ('Hypermodel technology') for the connection of models across different scales.

Community

VPHOP contributes a sizeable community by virtue of a large consortium of 20 partners, covering clinical, industrial and academic interests from across Europe. However, the project enthusiastically contributes to a much larger external community interaction by virtue of Biomed Town and other initiatives such as Physiomespace, MAF etc. Biomed Town (www.biomedtown.org) is a free community of peers who have a professional or educational interest in biomedical research and technology. It is a maturing resource with a community of approximately 2000 citizens. The imminent release of a new “town product” will allow the easy deployment of different towns for specialised communities and needs eg. biomedworld.org.

Sustainability and the ToolKit

Sustainability aspects are central to VPHOP and begin with Biomed Town (www.biomedtown.org), which embraces open source concepts, and development is based around open source frameworks such as Zope (application server), Plone (Content Management System) and Python (programming language). PhysiomeSpace (and other emerging services) will be intimately connected with BiomedTown and increase visibility and access. PhysiomeSpace is a sustainability initiative that offers a biomedical data repository and sharing service, with tools to support annotation of user data, sharing of data with other users and searching for new data. The application client supports interoperability, enabling users to import a wide range of biomedical data, visualise and annotate it and enable upload/download to a PC. Data search is possible through the use of ontologies and metadata curation. It is anticipated that the NMSPhysiome international cooperation project with SIMBIOS will extend PhysiomeSpace annotation to be compatible with the NIH semantic annotation. MAF is another sustainability initiative that relies on an open source framework for rapid development of multimodal visualisation applications (www.openmaf.org). It uses C++ based code and integrates many specialised libraries (VTK, wxWidget, ITK, Crypto++, XercesC, Curl and others). There is provision for different kinds of data inside a single data space (support for 4D hierarchical data) with multiple visualisations of data. More recent developments will exploit a modular architecture, extensible through plug-ins with modules that can be wrapped into scripting languages (Python, Java...) through a well defined public API. It is intended that modules will have a Facade and offer multithreading support. The communities behind these initiatives are enthusiastic and growing and offer much in the way of sustainability. Finally, a Hypermodel approach to modelling has been developed. This is a platform in which the composition of generic existing predictive models can be crafted into a new model. Each model represents the process at a particular scale or a specific sub-system and the hypermodel would represent the whole biological process across scales and sub-systems. VPHOP's experience of tool building to date makes this a potentially attractive and viable platform for a wide range of modelling tasks across the VPH.

VPHOP	**Description**	**Comments**
Project Outputs	Multiscale modelling of skeletal mechanobiology with facility for patient-specific simulation.	
Content relevant to the ToolKit	Community support – BioMedTown, PhysiomeSpace Open source MAF Hypermodel platform Standards – ontology annotation	Sustainability through community interaction (BioMedTown, PhysiomeSpace) and wider links (NMSPhysiome/SIMBIOS)
Benefits of NoE involvement	Exposure to wider community Standards promotion	

Action-Grid

This is an International Cooperative Action on Grid Computing and Biomedical Informatics between the European Union, Latin America, the Western Balkans and North Africa. The main objective is to collect the relevant achievements in the fields of Grid Computing and BMI that can be reused and transferred to Latin America, the Western Balkans and North Africa. By extending the methods proposed in ACTION-Grid there is potential for wider benefit to other regions and countries. Such knowledge reuse will be based on previous achievements of the consortium and, of course, from other people and groups. The agenda for ACTION-Grid developments includes:

- Surveying Grid-based and BMI initiatives in Europe, Latin America, the Western Balkans and North Africa. These results will be combined with data from a resource-ome, an inventory of Grid/Nano/BMI methods and services-, developed by the consortium.
- Fostering training and mobility in Grid, BMI and nanoinformatics, based on previous EC-based projects
- Developing a White Paper, in collaboration with a panel of recognised experts. This document will be delivered to the EC to establish a future agenda covering the Grid/Nano/Bio/Medical Informatics areas and develop new plans in Latin America, the Western Balkans and North Africa.
- Disseminating ACTION-Grid, by means of: (a) An international symposium on Grid and BMI. This conference will be carried out in Europe, with two satellite conferences (b) Scientific publications, (c) Dissemination strategies, such as a Website, Newsletters, Press releases, etc.

RADICAL

RADICAL is a Coordination Action that brings a coherent, in depth and scientific approach to analysis and recommendations for security and privacy enhancement in the Virtual Physiological Human. It will take into consideration technology advancements, business and societal needs, and consider ethics and other challenges that deserve to be addressed for societal benefit. RADICAL objectives are:

- To benchmark existing security and privacy technologies. This includes a special focus on Privacy Enhancing Technologies, which assist in designing information and communication systems and services in a way that minimises the collection and use of personal data and facilitate compliance with data protection regulations.
- To identify required technology developments and implementation challenges in order to clarify the gap between the present and future desired status of security/privacy.
- To identify societal needs and challenges that should be addressed in order to protect health patient records and regulate their usage. This will encompass implications of health data usage, with a special focus on genetic data.
- To capitalise on existing knowledge acquired by EC funded projects under the 6th Framework Programme, using their experiences and expertise to provide a policy paper/roadmap for the future agenda in medical and genetic data.
- To develop a Good Practice Guide, presenting the best practices that should be adopted by the many stakeholders of the VPH.

Content Source 3: VPH-I Projects (Call4)

ICT Call 4 enabled five projects with an international research agenda to be funded. These projects are very new and have only recently begun. Thus, the process of NoE engagement is not well established and arrangements for ToolKit interaction have not yet been specified. Details will be added over the coming months.

Call 4 Project #1: RICORDO

RICORDO (www.ricordo.eu co-ordinated by EMBL-EBI) is a project that began in February 2010, and has goal that are aligned with aims of the VPH. This includes a focus on interoperability framework that links physiology-related data and model resources. Ontologies are a natural feature of this work, complementing standards in the areas of mark-up languages and minimal information for representation. RICORDO will further develop, implement and demonstrate the effectiveness of such standards for ordinary differential equation (ODE) models of physiology and radiological models of anatomy. Furthermore, it aims to integrate models and ontologies related to medical physiology and human anatomy in collaboration with the OBO Foundry. The RICORDO infrastructure is already attracting collaborative links with modellers in the pharma domain (through the EBI's involvement in the Innovative Medicines Initiative project), and with groups that focus on the development of clinical terminologies. US Partner: University of Washington.

RICORDO	Description	Comments	ToolKit
Project Outputs	Composite Ontology term grammar		2012
Content relevant to the ToolKit	Webservice-based infrastructure for exchange and query of ontology-based annotation		2012
Benefits of NoE involvement	VPH-wide metadata standards for models and data		2012

Call 4 Project #2: MSV

In recent years, various terms – the Virtual Physiological Human (VPH), Integrative Biology, Physiome Research – have been used to describe the trend in biomedical research towards the consideration of systemic processes. These phenomena are commonly observed in living organisms but cannot be explained within a single sub-system, but reflect, rather, systemic outcomes that result from the interaction of multiple sub-systems.

Traditionally, when confronted by the complexity exhibited in biomedical problems, researchers have been forced to focus purely on individual sub-systems; the most common boundary separating these has been spatiotemporal scale. The current interest in VPH is demanding greater concentration on the study and simulation of biology systems at multiple scales, and multiscale data collection and multiscale modelling have recently become synonymous with integrative research. The many VPH projects, that will start to demand multiscale visualisation in the coming years, suggests that this area should receive urgent attention.

The Multiscale Spatiotemporal Visualisation (MSV) project aims, by international cooperation between the European @neurIST and VPHOP integrated projects, the USA National Alliance for Medical Imaging Computing (NA-MIC), and the New Zealand-based IUPS Physiome initiative: i) to define an interactive visualisation paradigm for biomedical multiscale data, ii) to validate it on the large collections produced by the VPH projects, and iii) to develop a concrete implementation as an open-source extension to the Visualisation Took Kit (VTK), ready to be incorporated by virtually any biomedical modelling software project.

International Partners: NA-MIC and IUPS Physiome.

MSV	**Name**	**Description**	**General**	**ToolKit**
Tools	MSV Toolkit	A library that solves visualisation parading for multiscale spatiotemporal biomedical data.	Yes	Expected to be ready in Dec 2012
Models				
Data		Collection of exemplary problems that the multiscale visualisation paradigm should solve effectively. Will also be made concrete in shared data collections that materialise each exemplary problem into a precise data space, and into a specific user scope.	Yes	Is not clear about restrictions to share the data.

Call 4 Project #3: TUMOR

TUMOR aims to develop a European clinically oriented semantic-layered cancer digital model repository from existing EU projects (ContraCancrum and ACGT). It will be interoperable with the US grid-enabled semantic-layered digital model repository platform at CViT.org (Centre for the Development of a Virtual Tumour, Massachusetts General Hospital (MGH), Boston, USA) which is NIH/NCI-caGRID compatible. The development of the project will be based upon specific clinical scenarios that will be implemented within an integrated EU-US workflow environment prototype for predictive, *in silico* Oncology-guided clinical studies. This will be deployed towards the end of the project. As an end result, a specific, clinically relevant workflow involving both EU and CViT models will be demonstrated, which will clearly highlight the need for and added value of interoperability. The project has only just commenced, with the kick-off meeting scheduled for May 2010. Detailed work plans and system designs will be developed during the first few months of the project.

EU projects: ContraCancrum, ACGT, VPH-NoE.

US partner: Massachusetts General Hospital.

Call 4 Project #4: NMS Physiome

The NeuroMusculoSkeletal Physiome project brings together two of the largest research efforts on personalised, predictive, and integrative musculoskeletal medicine with a major US partner.

The project involves a three year cooperation between some partners of VPHOP and some partners of SIMBIOS, the USA NIH Centre for physics-based Simulation of Biological Structures established at Stanford University (http://simbios.stanford.edu/). The project aims to create synergies between the work of the two consortia with respect to research communities (biomedtown.org and simbios.org), resource repositories (physiomespace.com and simtk.org) and to the modelling software tools (MAF and OpenSIM/FEBio).

Call 4 Project #5: Sim-e-Child

Sim-e-Child aims to develop a grid-enabled platform for large-scale simulations in paediatric cardiology. EU project: Health-e-Child; US partner: the American College of Cardiology, Johns Hopkins University.

There is a high demand for patient specific cardiovascular disease therapeutics. Paediatric cardiology, in particular, faces difficult challenges due to the evolving nature of a child's heart and vascular system. Comprehensive and accurate computer models reconstructed from patient specific data and simulated physical constraints are needed to help determine better and more reliably risk stratification to improve and personalise therapies, and ultimately to decrease morbidity and increase survival of patients.

The Sim-e-Child project proposes to develop a grid-enabled platform for large scale simulations in paediatric cardiology, providing a collaborative environment for constructing and validating multi-scale and personalised models of a growing heart and vessels. The project will establish an international cooperation, by linking the EC funded Health-e-Child project with leading institutions such as the American College of Cardiology, Johns Hopkins University, Technical University of Munich, and Siemens Corporate Research. Sim-e-Child is an extension of the Health-e-Child platform that:

- Interconnects the Health-e-Child database with new data from two prospective US multicentre studies;
- Enhances and expands the Health-e-Child heart model with existing models of the aorta, aortic valve and mitral valve, and with computational fluid dynamics;
- Integrates the Health-e-Child Gateway and Case Reasoner with versatile tools for simulation workflow composition (iKDDTM) and sharing of scientific experiments (SciPort).

The objective of the Sim-e-Child is to strengthen the impact of the Health-e-Child project by creating an international simulation and validation environment for paediatric cardiology, supported by integrated data repositories. The project will advance the state-of-the-art by providing comprehensive and patient specific models for the dynamic and longitudinal interactions occurring in the left heart, with a focus on the congenital aortic arch disease and repair.

VPH-I ToolKit Summary

The table below summarises the current state of knowledge of VPH-I projects.

Project	Content relevant to the NoE ToolKit	Comments
ARCH	Image processing and vmtk exploitation gnuid (3D CFD solver) pyNS (vascular network solver) ArchFTP client Standards - data as XML	Sustainability through open source community involvement
ARTreat	A suite of tools… Arterial geometry reconstruction and stented /unstented haemodynamics Molecular dynamics; plaque characterisation Data mining application Training environment	Sustainability through community training and commercialisation opportunities
Contra-Cancrum	Voxel tools and vtk exploitation DoctorEye DoctorsCompare Standards – DICOM	Sustainability through TUMOR project. Links with CviT.
euHeart	Open source framework for multiscale/multiphysics Tools library based around… OpenCMISS, SOFA, OPENFEM, LIFEV Visualisation based around… GIMIAS, cmGUI Database - euHeartDB Standards – CellML, FieldML	Sustainability through interaction with open source communities.
HAMAM	Open XNAT toolkit. Multimodal processing tools. Multimodal data sets Standards – XML structuring of database, patient data.	Sustainability through commercialisation of *Workstation*
IMPPACT	Interventional Planning System (IPS) Physiological Organ Simulation Toolkit (POST) Real time visualisation Thermal modelling	Sustainability through community and clinical involvement
NeoMark	Trained algorithm for cancer reoccurrence prediction. Oral cancer database Software for data entry and analysis Image feature extraction Real-time PCR chip	Sustainability through adoption of tools in the wider community.
PASSPORT	VR-Render image viewer VR-Planning VR-Anat Open source FW4SPL Open source SOFA Open databases Standards - DICOM	Sustainability through community involvement - WebSurg, SOFA, vtk/itk, conference events etc. VR-Anat commercialisation
PreDICT	Open source cardiac simulation – CHASTE Validated ion channel models Myocardial tissue models (meshing) Standards – ontological annotation, CellML, FieldML	Sustainability linked to commercial interest. Hence acute awareness of the needs of pharma.
PredictAD	Heterogeneous data and software tool for early detection of AD.	Sustainability through community links (ADNI)
VPH2	A suite of tools that include… Data mining Functional assessment of left ventricle Functional assessment of mitral valve performance Predictive tool for scoring of therapeutic strategies Visualisation tool	Sustainability through community awareness, (BioMedTown, MAF). Potential (commercial) exploitation of tools.
VPHOP	Community support – BioMedTown, PhysiomeSpace Open source MAF Hypermodel platform Standards – ontology annotation	Sustainability through community (BioMedTown, PhysiomeSpace and NMSPhysiome/SIMBIOS)

Table 5. VPH-I ToolKit Summary

The table below identifies the technologies underpinning many of the VPH-I activities.

Name	Description
AHE	Application Hosting Environment – grid middleware
Chaste	Cancer, heart and soft tissue simulation environment
BioMedTown	Community provision for biomedical science
cmGUI	Continuum mechanics GUI. Field representation/visualisation
GIMIAS	Graphical interface for medical image analysis and simulation
GSEngine	Workflow tool
itk	Insight segmentation and registration toolkit
LIFEV	Finite element library exploiting state of the art numerical methods
MAF	Multimodal application framework
MIASE	Minimum information about a simulation experiment
OpenCMISS	Open source redevelopment of CMISS (Continuum Mechanics, Image analysis, Signal processing and System Identification)
OpenFEM	Finite element simulation
SOFA	Simulation open framework architecture
vmtk	Vascular modelling toolkit
vtk	Visualisation toolkit
XNAT	Extensible neuroimaging archive toolkit

Table 6. Underpinning technologies

Content Source 4: External Tools, Models & Data

Biomedical integrative research is not a concept unique to the VPH, nor is computer simulation of aspects of human physiology. Many tools of relevance have been developed, or are being developed elsewhere, without any affiliation to the VPH. In order to maximise the return on investment (both effort and funding) and avoid wasted effort in replicating work already done, it is desirable to include such tools within the VPH ToolKit wherever possible. Potentially, this is a massive task and raises unanswered questions about scope… what are the boundaries of the ToolKit?

In addition to VPH applications and their underpinning technology frameworks, the scope could extend to core scientific programming libraries (e.g. PETSc, HDF5), parallel computing libraries (e.g. OpenMPI), Grid/middleware efforts, and many US biomedical initiatives (e.g. CaBIG, Simbios), etc. Even within Europe, significant research funding has been expended on biomedical research, and there are many research outputs either already available or becoming available. The investigation and curation of these alone would require investment beyond the capacity of the VPH-NoE (both in funds and time).

The NoE is adopting a pragmatic, “on demand” approach to sourcing external content. Initial efforts will concentrate on adopting content that VPH researchers are already using or interested in using, in order to ensure that the ToolKit is addressing the current needs of the community. This will also be aligned with the expertise of VPH scientists involved in the curation of that content. Consequently, partners within each VPH project are being encouraged to submit relevant technologies, and to help with the curation process. However, it is recognised that this should not only apply to tools being proffered externally for ToolKit inclusion.

Evaluation and curation of tools generated internally is equally valuable and naturally precipitates discussions of metrics for Tool quality, with consideration of location, documentation, licensing, negotiation, training and maintenance important factors. Furthermore, perhaps there are aspirational qualities that can help drive ToolKit quality, and whose explicit stipulation might encourage delivery of exceptional quality tools, designated 'gold standard ToolKit content'. Currently, these are only in concept form, but they may become central to the longer-term sustainability of the ToolKit. They are examined further in the Discussion section.

By way of example, with image processing a core requirement for the VPH NoE ToolKit it is readily appreciated that a key task is the identification and assessment of imaging tools that may as yet have no equivalent in any tool sets developed within the NoE’s established communities. Or which may not have been appreciated as generally applicable to a wide range of VPH users. Examples of such external systems include:

- The Vascular Modelling Tool Kit (VMTK). An open-source toolkit developed by David Steinman (Univ. Toronto) and Luca Antiga (Mario Negri Institute, Italy). Luca Antiga is involved in the VPH-I project ARCH and VMTK has already been identified as a candidate for NoE inclusion.
- The @neurIST project (IST-027703) had similar goals to the VPH initiative (see vision document case study http://bit.ly/b5PX3M) within the arena of intracranial aneurisms, and has produced a number of useful but specific image-processing tools that may be of use to the wider VPH community.

Finally, however, it is appreciated that care is required in the management of NoE resources to ensure that ToolKit content is optimised without disproportionate devotion of effort to any one category of tool source.

Discussion

Even at this embryonic stage of the VPH, it is clear that the community is nurturing a wealth of potentially powerful and relevant tools (Section 6.20). The ToolKit portal can usefully bring some order and structure into this collection, with the NoE clarifying direction and recommending strategies that promote interoperability and sustainability. The tools range from software for data exchange (eg. ArchFTP client) to physiological modelling (eg. euHeart) and also include database and analysis support. Many imaging related tools are being developed, which reflects the importance of imaging to the VPH community, but perhaps, also increases the danger of overlapping functionality. Standards are not widely professed, but they are in evidence, from the use of XML to structure data (eg. HAMAM) to the use of CellML/FieldML in modelling work. Also, it is encouraging to see active uptake of underpinning technology frameworks like MAF, itk etc. Such action helps to consolidate the respective user communities, accelerates tool development and encourages standardisation.

From this analysis, it is possible to begin to identify centres of expertise; groups with experience that can assist new users, helping those in difficulty and clarifying the merits of the technologies to potentially interested parties. This might also be considered a pragmatic approach to standardisation because experts will tend to offer opinions that are necessarily biased towards formats that have been effective for their own use, and therefore encourage their uptake. In summary, the underlying message is one of community as an empowering force for sustainability. This is echoed by the NoE, which has adopted community/engagement as one of its strategic pillars.

The VPH and Sustainability

Project Assessment

Many of the exemplar projects are now complete or nearing completion and their content is duly being added to the ToolKit. The table of Section 6.20 clarifies the tools status of the VPH-I, and this data is further illuminated by a WP3 engagement exercise in which projects were asked to clarify their sustainability priorities by completing a questionnaire, formulated as a sustainability matrix.

The matrix row/column headings were derived from an analysis of priorities identified in the original RTAE, and an unpopulated example (with an explanatory comment in each element) is shown in figure 1. The grid attempts to pair sustainability 'themes' (community, standards, exploitation etc.) with ToolKit related content (visualisation, simulation etc), so that all components of the latter can be examined in a sustainability context. Each element is scored in relation to project effort/commitment, using a 4-point scale:

- 0 – No effort invested in this area
- 1 – Little effort currently invested in this area
- 2 – Moderate effort invested in this area
- 3 – High priority with significant effort invested in this area

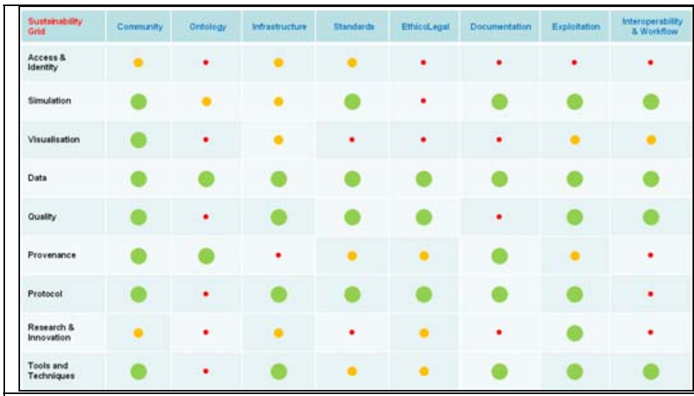
< - - - - - - - - - - Sustainability Concepts - - - - - - - - - - >

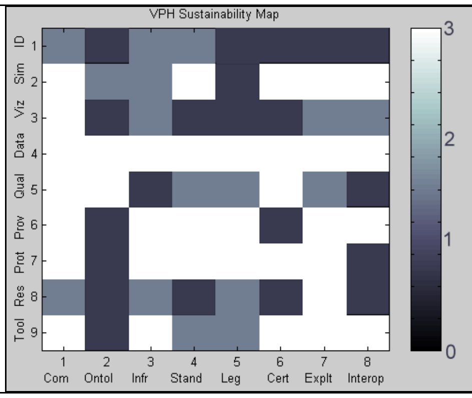
VPH ACTIVITIES

Sustainability Grid	Community	Ontology	Infrastructure	Standards	EthicoLegal	Documentation	Exploitation	Interoperability & Workflow
Access & Identity	Federated authentication	Ontology for authentication	Security infrastructure	Industry-standard authentication	Techniques will be precedents	All contributions must be optimally documented	Must meet industry standards	Must operate universally
Simulation	Simulation community	Shared ontology, multiscale, multidiscipline	Complex, to meet multiscale simulation needs	Establish and maintain standards for interoperation	VPH-wide standards for data and transfer	All contributions must be optimally documented	Systems to support collaboration	Services for 'all' platforms
Visualisation	Standards?	Terminology?	Bandwidth?	Standardised methods?	Legitimate?	All contributions must be optimally documented	Proprietary?	Consistent?
Data	Each discipline needs a data community	Entrench concept of data descriptors	Federated, searchable repositories	Gold standard for data formats	Attach Ethicolegal metadata to everything	All contributions must be optimally documented	Requires clear data usage definition	Aim for Gold Standard formats, but allow translation
Quality	QA community, Toolkit validation	Ontology of QA practices required	Infrastructure must facilitate QA	Use established standards	What are the implications?	All contributions must be optimally documented	Exploitation will favour HiQ content	QA practices for workflows
Provenance	Sources should be part of the community	Provenance should be in metadata	Facilities to control usage by permissions	Standardised formats	Strong control of usage	All contributions must be optimally documented	Is material suitable for industry?	Do limitations prevent chained activities?
Protocol	Community agreement on methods	All instructions defined	Toolkit use will also be protocol-driven	Wherever possible	Mutually supportive	All contributions must be optimally documented	Exploitable output will be protocol-driven	For workflows, protocols are essential
Research & Innovation	Support for IP, peer-review, outreach...	Mandatory for release?	Procedures required for IP-aware interactions	Consider standards from the outset	Licensing...	All contributions must be optimally documented	Framework for exploitation essential	IP etc services must integrate and interoperate
Tools & Techniques	User groups, community support	Tools to aid ontological annotation of tools	Infrastructure must support tool deployment	Adherence and development essential	Ground rules for use of tools in clinic applications	All contributions must be optimally documented	Not just academic use – need industry/clinic	Only interoperable content is gold standard?
Interoperability & Workflow	Consider having new workflow forum	Essential for automation	Infrastructure must support workflows	Perhaps the greatest challenge?	Permitted? Protected?	All contributions must be optimally documented	Licensing?	N/A

Figure 4: *Unpopulated sustainability matrix. In this example, each element is annotated with a supporting comment to help clarify interpretation of the row/column heading pairs.:*

There are many cases where the pairing of terms is unambiguous and straightforward and these elements are easy to populate (eg. simulation + documentation; ie. to what extent is your project allocating effort to documentation of simulation tools?). However the grid also frequently raises issues that participants may not have considered in respect of sustainability (eg. standards + protocol; ie. to what extent is your project committed to the standardisation of end-user protocols?). Note that the title 'interoperability and workflow' features as a heading for both rows and columns, since it was deemed to be equally valid as a sustainability theme and a ToolKit entry.





5a. Priority:
Blank – 0
Small red dot – 1
Medium amber circle – 2
Large green circle – 3

5b. Priority:
Black - 0
Dark grey – 1
Light grey - 2
White - 3

Figure 5: *The image on the left (5a) illustrates the sustainability matrix for one of the VPH-I projects (ARCH). The image on the right (5b) reinterprets the data as a sustainability intensity map*

An example of a project-specific, populated grid is shown at the left of figure 5 (for VPH-I project ARCH). However, for improved visualisation, the same grid is replicated on the right of the figure (5b), but where the numerical value of each response is represented as an intensity to create a sustainability intensity map (SIM). The value of this approach is shown in figure 6, in which the intensity maps from the projects have been superimposed and summed to create a VPH-I intensity map, with white normalised to the highest sum/value in the matrix. The result is snapshot of the sustainability focus of the VPH.

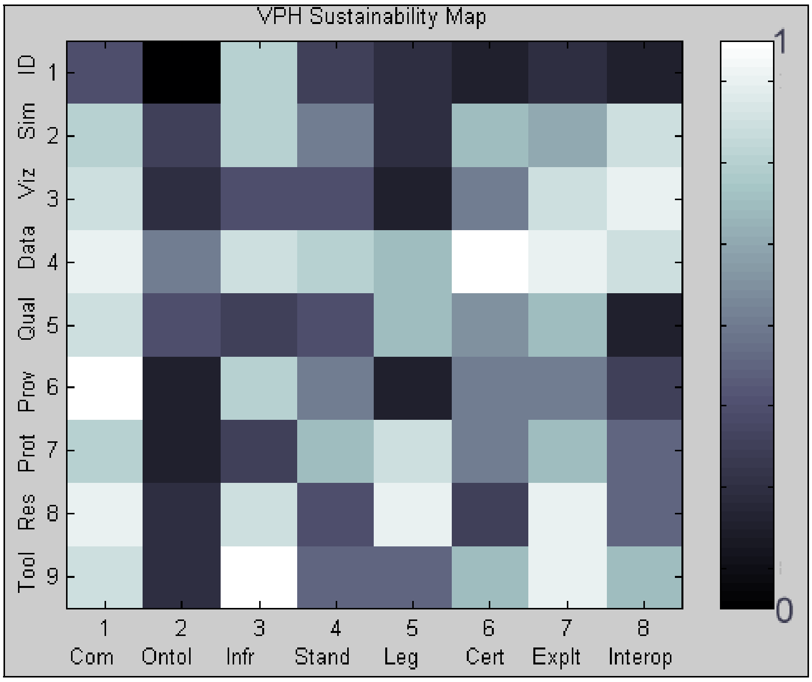


Figure 6: *Sustainability intensity map for the VPH-I as a whole*

As anticipated a broad spectrum of contributions is apparent, providing for a well populated matrix. Broad strengths (ie bright areas) are apparent in areas of community and Data/Visualisation/Simulation/Interoperability. Notably, there are several weaker areas too, identified as:

1. A dim column (ie. under-representation) in ontologies (labelled as Ontol)
2. A dim row in security (labelled as ID)

Other localised weaknesses relate to legal aspects, quality and provenance. As a whole, this data supports the significant NoE effort currently being invested in ontologies, but it also provides justification for closer scrutiny of security issues across the VPH. This is fundamental to interoperability and must support rigorous but flexible authentication/authorisation, based on a secure infrastructure for information and modelling exchange. Indeed, these sentiments echo some of the priorities of FP7-Call6 and appropriate solutions may emerge from projects that are funded.

Essential Attributes

Many project outputs are predominantly in the form of software tools, but if longer-term uptake is a genuine goal, then all should embrace a sustainability analysis that includes elements such as community, ethico-legal implications, certification/documentation, interoperability, training, user-friendliness etc.

- Community – Projects should operate in an environment that encourages user-community interaction, perhaps even pre-allocating resources for support and outreach throughout the life of the project, based on estimates of the software user-base.
- Ethico-legal –It is important to identify ethico-legal implications relating to use of the software tools and data, highlighting the magnitudes and types of effort that might be required to overcome hurdles to interoperable implementation (for example, explicit advice on licenses and terms of use of the software).
- Documentation – This is a key element of sustainability, since adequate documentation (including tutorials/examples) is essential if broader, long-term uptake of the project software tools is to be achieved. In particular, this effort would benefit from the concept of self-certification in which a project not only documents the functionality of its tools, but also includes a critical self-evaluation, characterising its adherence to professed capabilities, performance, standards etc. This would be invaluable to end-users and would assist external evaluation by 3rd parties. The latter

might be an important hurdle for those tools aiming for 'Gold standard ToolKit content'.

- Interoperability – The adoption of a standards mentality early within the projects would help to ensure that tenets of interoperability are foremost in the tools development process. An assessment of similar and competing technologies can help to inform interoperability objectives.
- Training and User-friendliness – This is complementary to the documentation effort and includes the possibility of project teach-ins and workshops in collaboration with the NoE.

Unfortunately, adherence to the sustainability principles proposed above does not necessarily ensure long term sustainability. However, the process will encourage acceptance and use within the VPH community and thereby help to promote its continuing existence beyond the end of the project. The presence of end users and early adopters, can create an enthusiastic user community that is willing to sustain the software and its concepts. This can also be achieved through affiliation with a ready-made community, relying on underpinning frameworks like MAF, itk etc. Finally, 'Gold standard content' is a quality concept that may be of value to the NoE ToolKit, since it not only encourages quality, but also opens a natural avenue for dialogue with the VPH.

Strategy

This document is tasked with assessing the first year of ToolKit progress. It is pleasing to report that there is evidence of steady growth and particularly that there is a plentiful selection of tools under development with which to populate it. These can be expected to be delivered over the next couple of years. Note however, that a collection of isolated tools spread across Europe gains little advantage if the only influence of the NoE is to enable them to become a collection of isolated tools residing under the roof of the ToolKit portal. The metric of success must not be the number of tools hosted by the ToolKit, but instead must be one of sustainability. In the near term this imposes many requirements, like that of community (creating inertia for continued tool use), interoperability (encouraging tool re-use and their interaction with other tools for novel solutions), documentation and training (providing maximum user benefit for minimum user effort)... the list goes on.

The work of the past 12 months has built on initial developments in which the technical landscape was charted through the RTAE and the ToolKit portal established. The more recent work of WP3 (since the 2009 review) has been an exercise in VPH engagement (addressing the spectrum of potential ToolKit content providers) in which the message of sustainability has been promulgated and accepted as a focus for ToolKit development. This acceptance must now be turned into action as the ToolKit becomes populated with the wealth of tools from the VPH and offerings from other EU projects (eg. the FP7 VPH calls) and beyond .

The NoE WP3 ToolKit strategy for the future is underpinned by sustainability and imposes numerous demands on VPH ToolKit content providers...

(i) Each VPH project will be required to declare its project's aims and possible ToolKit provision. This will be documented in summary form, similar to the project descriptions presented in this document. This will become a companion volume to the portal, growing yearly, outlining content contributions and offering a snapshot of ToolKit health at any instant. (The updated document with additional critique will constitute the NoE WP3 ToolKit deliverable of future years).

(ii) ToolKit content delivery from a VPH project will involve interaction with a designated individual from WP3. This contact will act as a champion for a core sustainability theme relevant to the project (eg. Grid, community), promoting integration of sustainability concepts and guiding delivery of sustainable-compliant tools to the ToolKit.

(iii) These efforts will be supported by a requirement for each project to complete a sustainability matrix for purposes of clarifying overall ToolKit and VPH status, direction and audit.

(iv) These efforts will be supported by a requirement for each project to complete a modelling matrix for purposes of clarifying the scientific/modelling status of VPH participants. The modelling matrix is an extension of the sustainability map concept and borrows from the modelling grid of WP2.

(v) These efforts will be supported by a requirement for each project to submit example application workflows that can be subjected to analysis for ontology and standards support.

It is proposed that these requirements are mandatory for all projects funded under the VPH umbrella, with relaxed requirements being applicable for other content sourced outside the VPH. Together, this assembly of initiatives can provide a snapshot of ToolKit status that can inform ToolKit strategy/development and deliver interoperable ToolKit content that offers improved sustainability for the future.

Conclusion

The second year of NoE activity has seen a change in the direction of activity for the VPH NoE. Year 1 overwhelmingly consisted of investigational work, in which a large series of studies into the current state of play of the technology, the participants, and the need for particular technologies, tools, data, models and training were conducted. In this second year the various developmental activities began to bear fruit, and the serious structuring has taken place to allow the production of detailed standards for performance, and to equip the NoE with a team approach to cooperation with those colleagues in the wider community who will provide the ToolKit with desirable content.

Significantly, the key evolution in thinking that has occurred has concerned the need for the NoE to become engaged in sustainable activities that will endure beyond the end of the allotted project duration. The key activities now under way are all concerned with identifying, and feeding the ToolKit with sustainable content that will endure, whether by its overwhelming desirability and utility, or its longevity by virtue of the established development teams and in some cases the user bases.

In this document has been presented the NoE's response to the need for enduring outputs:

- The formation of Working Groups to operate with VPH-related projects
- The establishment of further Working Groups to build standards for content
- The steps being taken to develop generalised tools for imaging and fusion
- The development of models and modelling aids including an array of MLs
- The emphasis on the workflow concept of model integration to support combinatorial operation and interactions across spatial and temporal scales
- The identification of the specification for a data-hosting environment and the ways in which sister projects can be enabled to make their data available
- Similar activities to examine application-hosting for the purpose of streamlined operation of models in a GRID environment
- A major set of activities to bring ontological annotation firmly into the VPH domain, with a series of initiatives to educate, inform and enable
- Development of the Portal, through which all VPH NoE ToolKit transactions will pass
- A comprehensive assessment of, and engagement with, the Exemplar projects and, equally importantly, the existing and emerging VPH-I projects
- Above all, the concepts of sustainability, as applied to VPH tools models and data

Year three will now see the sustained application of effort in these areas, to equip the ToolKit with content that meets the required standards, to engage further with the VPH Community, and to cooperate with the forthcoming central Infrastructure activities to build a secure and sustainable platform for continued VPH operation.

Application for an Account on the NoE/USFD SFTP Server

Please send an email request to Richard.Knight@sth.nhs.uk;steven.wood@sth.nhs.uk containing the username you would like to use on the system, you partner name as defined in the DoW (eg. USFD) and the work packages with which you are involved. The email address you send from should be the one that you use for all other NoE correspondence – your details will be checked for validity. We will respond with your login details.

Please note that only the SFTP protocol is allowed to connect, SCP will be rejected. Also please allow up to one week for your account to be processed. It should not take this long but at times we have other priorities and are unable to action these requests as quickly as we would like.

Usage:
The system is set up so that there is a folder for each project partner, and all users from this institution will have read/write access over the data within this folder. The remainder of the NoE area will have read only access for any partner.
There are no quotas or restrictions on file sizes on the server but we ask that you use it responsibly and that if you are transferring particularly large data sets that you delete them as soon as they are no longer required.
Please email the above address with any problems or requests for audit information.

Client software:
Linux has a huge amount of clients capable of connection to a SFTP site so we assume these users have the necessary skills to make a connection to the site without help. For windows users we recommend use of the free client software called WinSCP. This has never given us any problems. The client can be downloaded from the link below:
http://winscp.net/download/winscp400setup.exe

The installation is straightforward, but you may need to be an administrator on your PC. Please also note that for some institutions (eg. hospitals) this type of connection may be blocked by the main firewall. This is an issue you will have to resolve internally, but if you have any specific requests for information (so that you can forward the information to your IT departments to assist with the making of connections) we will be happy to try and answer them.
The host you will be connecting to is:
MPHNOESFTP1.sheffield.ac.uk

The remainder of the connection details will be given on application for an account. By using the login provided you are agreeing to the terms and conditions of use laid down below.
....

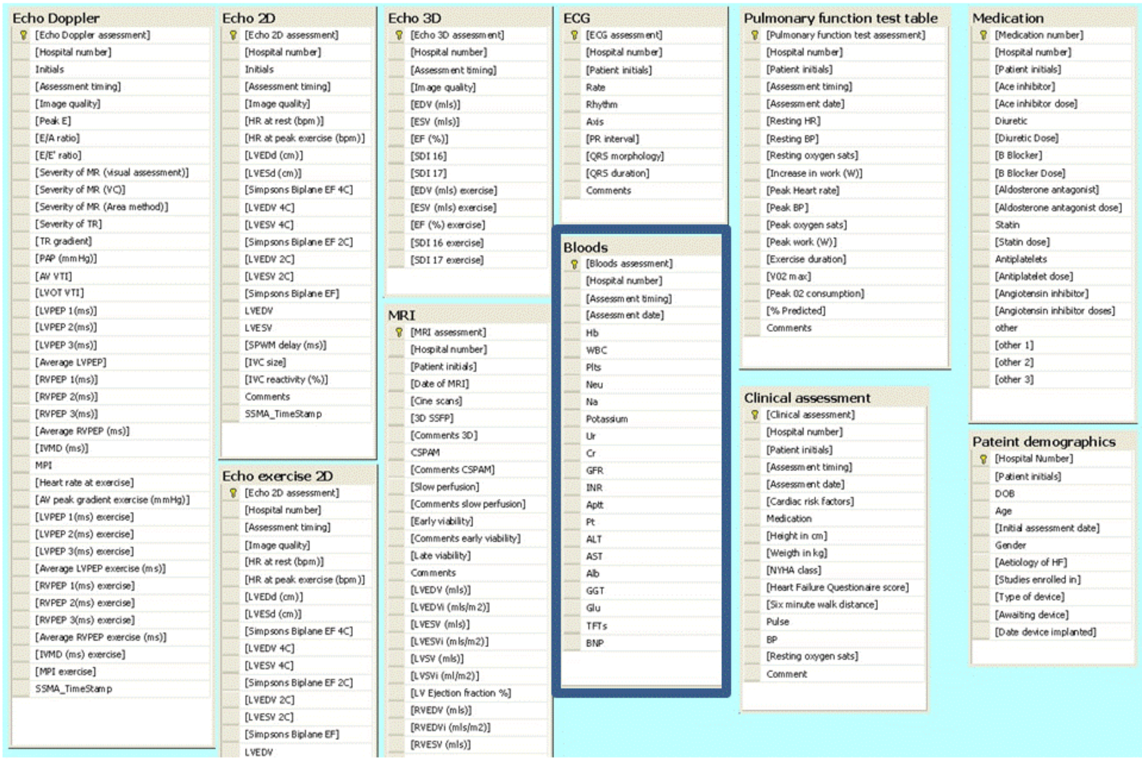
Sheffield SFTP Server: terms and conditions of use

This document outlines the appropriate use of the Sheffield SFTP server and by using the login provided, the user agrees to abide by its terms.

- All data saved on this server must be anonymised before hand. Whilst every effort has been made to secure this server the University of Sheffield will not accept any responsibility for the loss/exposure of data stored on this server if it is compromised.
- The default policy for the server is that all users in the NoE consortium can read all data stored under the /data folder. If this is unsatisfactory in your case, then you will have to make other arrangements for the transfer of your data.
- All access to the server is logged and NoE users are only permitted to use the /data folder. If any attempt is made to move outside this area your account will be disabled until the breach of these terms is explained.
- It is our policy to audit the data and the users that have accessed it. You are forbidden from passing any data downloaded from the server to any third party, even members of the project. If another member of the consortium requires data, they must apply for an account themselves and download it separately.
- You must not let other people use your account. They are able to apply for their own and you may be implicated in any irresponsible actions.
- Only data which is explicitly related to the NoE project may be stored on this server. It is forbidden to use this for the transfer of personal data or data associated with other projects you may be involved with. If there is any doubt as to what is appropriate the DoW will be used as the defining reference for the intended use of this server.
- Finally the server is not backed up in any way, so please ensure that any data you copy to it is also retained within your organisation.

Ontological mapping of clinical concepts.

CLINICAL DATA: SHEFFIELD HOSPITAL CVS Database Schema



The above figure shows a typical database schema used in handling patient data in a clinical cardiovascular hospital unit. The collaboration of the VPH NoE partners (EMBL-EBI and Sheffield) with the allied RICORDO project has carried out a case study of the ontology-based cross-product annotation of the blood test table schema (highlighted in a blue frame above). The list below is an illustration of a direct and consistent mapping of the above clinical concepts onto individual terms originating from key reference ontologies. These ontologies are being considered as part of a core standard for VPH-wide annotation. The annotation strategy investigated in this work will inform the ontology annotation pipeline discussed in WP3 (see Technical Developments section above: BluePrint for Ontology-based Data and Model Integration).

[Term]
id: Hb
intersection_of: PATO:0000033 ! concentration of
intersection_of: inheres_in FMA:9670 ! Blood
intersection_of: towards CHEBI:35143 ! hemoglobin

[Term]
id: WBC
intersection_of: PATO:0000033 ! concentration of
intersection_of: inheres_in FMA:9670 ! Blood
intersection_of: towards CL:0000738 ! leucocyte

[Term]
id: Plts

```
intersection_of:  PATO:0000033 ! concentration of
intersection_of: inheres_in FMA:9670 ! Blood
intersection_of: towards CL:0000233 ! platelet


[Term]
id: Neu
intersection_of:  PATO:0000033 ! concentration of
intersection_of: inheres_in FMA:9670 ! Blood
intersection_of: towards CL:0000775 ! neutrophil

[Term]
id: Na
intersection_of:  PATO:0000033 ! concentration of
intersection_of: inheres_in FMA:9670 ! Blood
intersection_of: towards CHEBI:26708 ! sodium

[Term]
id: Potassium
intersection_of:  PATO:0000033 ! concentration of
intersection_of: inheres_in FMA:9670 ! Blood
intersection_of: towards CHEBI:26216 ! potassium

[Term]
id: ur
intersection_of:  PATO:0000033 ! concentration of
intersection_of: inheres_in FMA:9670 ! Blood
intersection_of: towards CHEBI:16199 ! urea

[Term]
id: cr
intersection_of:  PATO:0000033 ! concentration of
intersection_of: inheres_in FMA:9670 ! Blood
intersection_of: towards CHEBI:16919 ! Creatine

[Term]
id: GFR
intersection_of: PATO:0000161 ! rate
intersection_of: inheres_in GO:0003094 ! glomerular filtration

[Term]
id: Alb
intersection_of:  PATO:0000033 ! concentration of
intersection_of: inheres_in FMA:9670 ! Blood
intersection_of: towards PRO:000003918 ! serum albumin

[Term]
id: AST
intersection_of:  PATO:0000033 ! concentration of
intersection_of: inheres_in FMA:9670 ! Blood
intersection_of: towards PRO:000022172 ! aspartate aminotransferase


[Term]
id: GGT
```

intersection_of: PATO:0000033 ! concentration of
intersection_of: inheres_in FMA:9670 ! Blood
intersection_of: towards PRO:000007963 ! gamma-glutamyl transpeptidase-related enzyme

[Term]
id: glu
intersection_of: PATO:0000033 ! concentration of
intersection_of: inheres_in FMA:9670 ! Blood
intersection_of: towards CHEBI:17234 ! glucose

[Term]
id: TFT
intersection_of: PATO:0000033 ! concentration of
intersection_of: inheres_in FMA:9670 ! Blood
intersection_of: towards FMA:74640 ! Thyrotropin

[Term]
id: BNP
intersection_of: PATO:0000033 ! concentration of
intersection_of: inheres_in FMA:9670 ! Blood
intersection_of: towards PRO:000011375 ! natriuretic peptides B

[Term]
id: pt
intersection_of: PATO:0000161 ! rate
intersection_of: inheres_in GO:0007598 ! blood coagulation, extrinsic pathway

[Term]
id: appt
intersection_of: PATO:0000161 ! rate
intersection_of: inheres_in GO:0007597 ! blood coagulation, intrinsic pathway