Model Driven Paediatric European Digital Repository

Call identifier: FP7-ICT-2011-9 - **Grant agreement no**: 600932

Thematic Priority: ICT - ICT-2011.5.2: Virtual Physiological Human

Deliverable 13.3

Complete list of functionalities for compliance and system functionality

Due date of delivery: 28-02-15

Actual submission date: 06-03-2015

Start of the project: 01-03-13
Ending Date: 28-02-17

Partner responsible for this deliverable: HES-SO

Version: 1.0

SEVENTH FRAMEWORK PROGRAMME

Dissemination Level: Public

Document Classification

Title	Complete List of functionalities for compliance and System Functionality
Deliverable	D13.3
Reporting Period	Month 24
Authors	Ranveer Joyseeree, Henning Müller
Work Package	WP13
Security	Public
Nature	Report
Keyword(s)	Technical requirements, compliance

Document History

Name	Remark	Version	Date
D13.3v1.0	First draft	1.0	02.02.2015

List of Contributors

Name	Affiliation
Ranveer Joyseeree	HES-SO
Henning Müller	HES-SO
Marcello Chinali	OPBG
Harry Dimitropoulos	Athena
Steven Wood	University of Sheffield

List of reviewers

Name	Affiliation
Rod Hose	USFD
Bruno Dallapiccola	OPBG

Abbreviations

MD-PAEDIGREE	Model-Driven European Paediatric Digital Repository
VPH	Virtual Physiological Human
OpenAIRE	Open Access Infrastructure for Research in Europe
3D	Three-dimensional
SNP	Single Nucleotide Polymorphism
ANA	Antinuclear Antibody
RF	Rheumatoid Factor
API	Application programming interface

Contents

1. Executive Summary

Using the Description of Work (DoW) of MD-PAEDIGREE, the foundations for a cross-discipline, modern, state-of-the-art digital medical image repository were laid. Executing an idea that looked good on paper and creating a tool that really adds value to the experience of clinicians and clinical researchers is not an easy step and requires much inter-disciplinary hard work. The first step in that journey from a project proposal to a finished product required that technical user requirements be gathered on the ground. Subsequently, several meetings and interviews were held to collect desired functionalities and to prioritise them. The list obtained published in Deliverable D13.1. This was then refined through further interview sessions and experience with first system use.

This deliverable presents the most up-to-date list of prioritised technical user requirements. The team behind the implementation of the infostructure then reallocated the gathered functionalities into specific use-case scenarios, thereby making it easy to split the work among the partners involved in the development process and rendering it possible to have a functioning prototype with all basic features at the earliest opportunity.

Next, compliance issues are explored. More specifically, the vital need for interoperability between MD-PAEDIGREE and VPH-Share[1] technology is emphasised. That is accompanied by a discussion regarding the dissemination of scientific output connected to the infostructure being developed through OpenAIRE[2].

That is followed by highlights of the links between the current project and other past or contemporary projects. Re-use of existing state-of-the-art techniques, infrastructure, and know-how is a means to ensure that MD-PAEDIGREE builds on solid experience instead of starting from scratch. The fact that it seeks to embrace cutting-edge advances and incorporate them into a harmonious whole is a testament of the considerable efforts being made to make and keep it relevant, modern, and up-to-date in the fast-changing world of medical informatics and digital healthcare. It can be said with confidence that, given the current state and rate of progress, MD-PAEDIGREE is right on track to reach the forefront of smart data repositories and personalised digital medical care.

[1] VPH-Share. http://www.vph-share.eu/. Retrieved: 10.02.2015.
[2] OpenAIRE. https://www.openaire.eu/. Retrieved: 10.02.2015.

2. Introduction

This document builds on Deliverable D13.1 and Deliverable D13.2 of the MD-PAEDIGREE project and aims to provide a complete list of outcomes from the analysis of technical requirements and to demonstrate links with other initiatives that will help ensure a large distribution and usability of the proposed infostructure.

Over the past 2 years of the project, an initial list of technical user requirements was created, published in D13.1, and subsequently improved upon through regular interaction with the eventual end-users of the MD-PAEDIGREE infostructure. Compliance challenges relating to VPH-Share and OpenAIRE were also explored and an in-depth review of the methods to employ in order to ensure compatibility with them was published in D13.2.

Quite naturally, the next step in the lifetime of the project is to combine the findings of the previous deliverables into a solid plan of action in order to ensure good progress in the final two years of the project. In particular, the reinforced list of prioritized requirements will serve as an ordered checklist for the developers who are in charge of implementing the functionalities asked for by clinical and research partners. The current deliverable will also serve to highlight the efforts made and to be made in order to make MD-PAEDIGREE a very attractive proposition to all stakeholders to the detriment of any other competing tool.

In the remainder of this deliverable, the outcomes of the analysis of technical user requirements will be discussed. Then, the link between technical and clinical user requirements will be highlighted. Subsequently, the compliance challenges surrounding MD-PAEDIGREE will be expounded upon and measures to address them will be described. That will be followed by an exploration of links between MD-PAEDIGREE and other past and present initiatives. Finally, the outlook for the future will be described based on what has been achieved so far.

3. Outcomes of analysis of technical requirements

Assessing the level of priority for the development of each technical user requirement of MD-PAEDIGREE gathered is vital to the progression of the project since the quality of the early prototypes provided to the end-users will have a major impact on whether or not the infostructure will have any success. In fact, it is necessary to make any initial prototype of the system as useful to as many end users as early as possible. In other words, if the prototype implements the most frequently desired functions of MD-PAEDIGREE, it will quickly have a heavy positive impact on the work carried out by stakeholders and would, thus, quickly be able to integrate usual clinical and/or research workflows. Thus, the direction of integration and development of MD-PAEDIGREE needs to be dictated by the most pressing user needs and dependencies of those.

Deliverable D13.1 presented an initial list of MD-PAEDIGREE requirements for the information management structure of the project after consultation with stakeholders via a number of surveys. It also presented priority domains within that list. It is important to note that that document did not include all dependencies of the full system requirements. It rather gave an initial overview of the views of the stakeholders. Between the publication of D13.1 and the present, however, the list of technical requirements was updated at regular intervals through frequent contact with participating partners. Through D13.3, the latest iteration of that list will be presented.

First, individual requirements will be presented in descending order of priority, as determined by a rigorous mathematical analysis detailed in D13.1. Then, those requirements will be ordered according to specific use-case scenarios that were identified by the partners involved in the development of the infostructure.

3.1 Specifications in descending order of priority score

After regular consultation with stakeholders through surveys, the following list of prioritized requirements was drawn:

1. Ability to search using pathology as a criterion.

2. Easy upload of images and associated metadata.

3. Automatic pseudonymisation or anonymisation of uploaded data.

4. Secure data-sharing between users.

5. Ability to search using keywords.

6. Ability to search using age as a criterion.

7. Unified repository for clinical and research systems.

8.
- Ability to search using sex as a criterion.
- Ability to search using anatomical structure as a criterion.
- Automatic curation of uploaded data.

11. Use of MD-PAEDIGREE does not add to the workload of users.

12.
- Support for unstructured text and free-text reports.
- Support for data modelling and simulation in 3D.

14. Ability to search using image modality as a criterion.

15. A rating system to assess the quality of cases stored in the infostructure.

16. Ability to search using image similarity as a criterion.

17. Learning from usage pattern of users to provide more relevant search results.

18. Automatic annotation of anatomical regions and detection of modalities.

19. Support for search in multiple languages.

20.
- Data extraction from .c3d files[3].

[3] The 3D Biomechanics Data Standard. http://www.c3d.org/. Retrieved: 10.02.2015.

• Patient-specific simulation and prediction.
• Patient-specific workflow.
• Search for similar cases based on information in gait analysis time series.
• Ability to find similarity in all data (3D kinematics, kinetics, muscle activity) instead of just in images.
• Support for management of standard non-image data formats used in project (e.g. genetic SNP).
• API for easy access to heterogeneous data collection for data modelling/research.
• Annotation and extraction of structures: support for uploading and viewing annotation and extraction outcomes.
• Ability to upload and visualize simulation results.
• Ability to run simulations on data.
• Ability to incorporate information from and for patient advocacy groups.
• Ability to view detection, segmentation, and simulation results.
• Ability to display a structured overview of the complete patient files including results of medical examinations.
• Ability to find context-specific similar patients and visualize similarity connections.
• Ability to run computationally-intensive processes on distributed machine clusters or the cloud.

35. • Ability to include dietary information of patients in the repository.
• Ability to add details regarding treatment in MD-PAEDIGREE.
• Ability to download data easily from MD-PAEDIGREE.
• Ability to search using clinical features (e.g. ANA, RF, uveitis, and morning stiffness).
• Online viewers for data.
• Ability to create/upload/run custom-made processing algorithms e.g. Python and Javascript.
• Ability to classify motion-capture curves (e.g. joint angles) and to export classifications (code available at KU Leuven).
• Ability to edit detection, segmentation, and simulation results.
• Ability to export results of diagnostics from the platform.
• Ability to find similar patient files to compare patient history.
• Ability to use GPU processing to run computationally-intensive processes.

46. • Ability to place annotations on curves and to export annotated graphs.
• Ability to edit layout for data presentation.
• Ability to create custom work-flows for processing blocks (e.g. http://noflojs.org/).
• Ability to browse through patient list to compare similar patients.

50. Ability to send and receive data via the RESTful API.

Note: Scepticism around the value/quality of searching by image similarity has lowered its rank in the list above. It was, in the preparation of the proposal, considered a high priority as it is a demonstrably functioning and valuable tool in concrete scenarios.

3.2 Priority list of use-case scenarios

Partners involved in the development of the infostructure collaborated on finding suitable use-case scenarios that would help break down the initial list of technical requirements into coherent blocks. That allows an efficient distribution of tasks among partners in charge of developing the infostructure. The scenarios were validated as a correct representation of clinical and research needs by the stakeholders involved in MD-PAEDIGREE. In this section, those scenarios (numbered from 1 to 5) are presented in no particular order as all of them need to be worked on simultaneously in order to deliver a working prototype as fast as possible. Within each scenario, however, the individual requirements are in descending order of priority.

1. Decision Support - Search through internal data:

a) Search using pathology as a criterion.
b) Search using keywords.
c) Search using age.
d) Search using gender.
e) Search using anatomical structure.
f) Search using image modality.
g) Search using image similarity.
h) Learning from usage pattern of users to provide more relevant search results.
i) Support for search in multiple languages.
j) Find similarity in all non-imaging data (3D Kinematics, Kinetics, muscle activity).
k) Search using clinical features (e.g. ANA, RF, uveitis, morning stiffness).

2. Data collection, curation, annotation and representation:

a) Easy upload of images and associated metadata.
b) Automatic pseudonymization or anonymization of uploaded data.
c) Automatic curation of uploaded data.
d) Unified repository for clinical and research systems.
e) Streamlining MD-PAEDIGREE to avoid significant increase to workload of users.
f) A rating system to assess the quality of cases stored in the infostructure.
g) Support for unstructured text and reports, including details such as dietary habits, and treatment.
h) Automatic annotation of anatomical regions and detection of modalities.
i) Data extraction from .c3d files.
j) Support for management of standard non-image data formats used in project.
k) API for easy access to heterogeneous data collection for data modelling/research.
l) Support for annotation/extraction of features in data, for exporting/uploading/viewing them.
m) Incorporate information from and for patient advocacy groups.
n) Easy download of data from MD-PAEDIGREE.
o) Online viewers for data.
p) Classify motion-capture curves (e.g. joint angles) and export classifications (KU Leuven code).
q) Edit layout for data presentation.
r) Create custom workflows for processing blocks (e.g. http://noojs.org/)
s) Export results of diagnostics from the platform.
t) Display a structured overview of the complete patient file including results of medical examinations.
u) List of patients that can be browsed for ease of comparison between similar patients.
v) Find context-specific similar patients and visualize similarity connections.
w) Send and receive data via the RESTful API.

3. Data Protection:

a) Secure data-sharing between users.

4. Decision Support - Search through external information:

a) Search through offline records.

b) Elicit second opinions from fellow experts.

c) Access to online search engines used often by clinicians.

5. Modelling and Simulation:

a) Support for data modelling and simulation in 3D.

b) Patient-specific simulation and prediction.

c) Patient-specific workflow.

d) Upload and visualize detection, segmentation, and simulation results.

e) Edit detection, segmentation, and simulation results.

f) Create/upload/run custom-made processing algorithms e.g. Python and Javascript.

g) Run computationally-intensive processes on distributed machine clusters or the cloud.

h) Use GPU processing to run computationally-intensive processes.

4. Link with clinical user requirements

Deliverable D2.1 was prepared by clinicians and detailed the clinical user requirements for MD-PAEDIGREE and represents a great resource in terms of implementing the most sought-after features first. Its general message was that there is a need to uncover complex interactions between clinical variables in order to gain a deeper understanding of the diseases being studied. In addition, understanding how the diseases progress in time is also an important objective, as is an insight into the influence of therapeutic intervention on the evolution of the patient's condition. To achieve those goals, D2.1 lists a wide range of clinical variables that might help reveal the complex underpinnings of the diseases being studied.

Detailed clinical workflows were produced for the four disease areas and the one gathered for cardiovascular diseases (Work Package 3) is illustrated in **Figure 1**. It shows that MD-PAEDIGREE can potentially add value to the prognosis prediction and patient follow-up tasks. The clinical requirements, as presented in D2.1, would, therefore, fit very well with the 'modelling and simulation' use-case. The modelling and simulations framework to allow for finding relations between those variables and clinical outcomes would seem to be well covered by the technical requirements listed in the previous section. There is a clear need, however, to accommodate for, within the MD-PAEDIGREE infostructure, the large set of clinical variables to be collected.

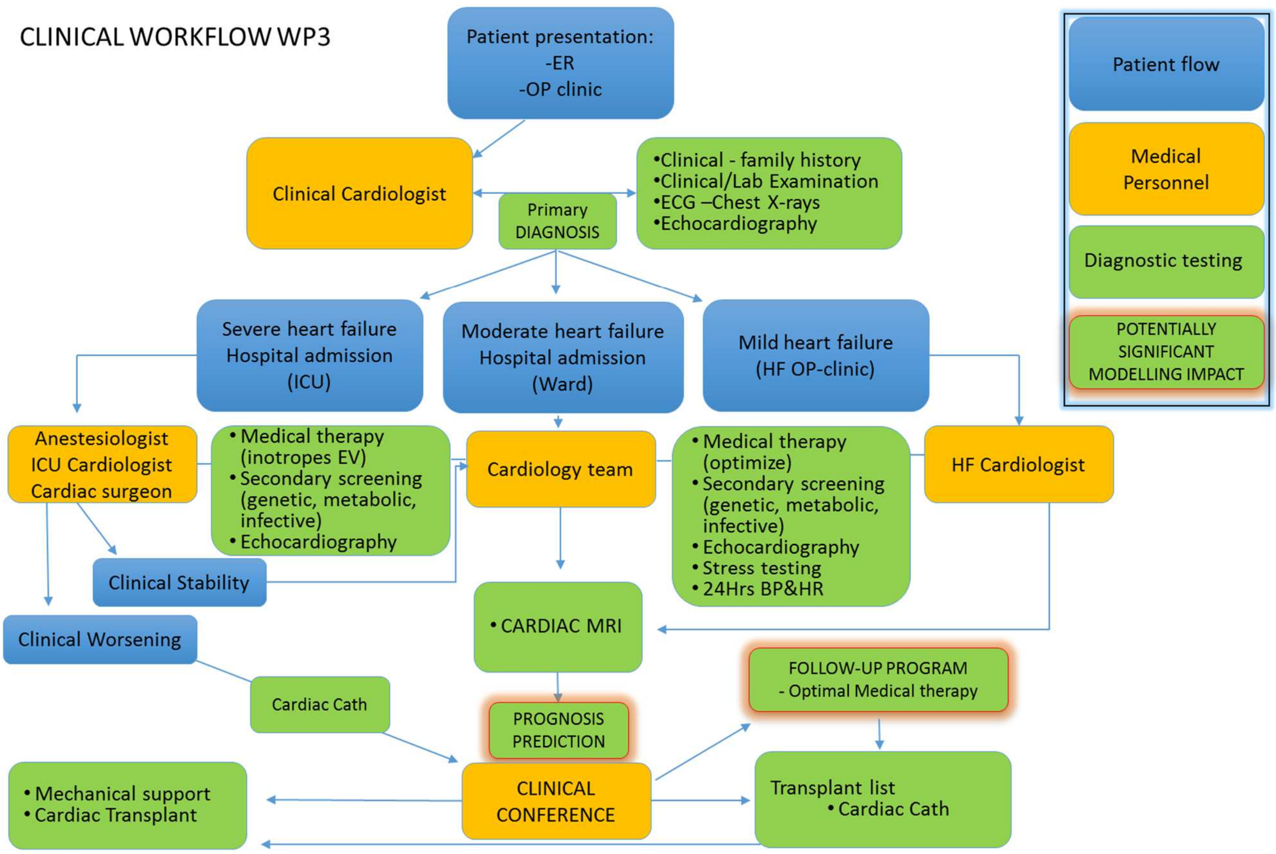

***Figure 1**: Clinical workflow for Work Package (WP) 3, which deals with cardiovascular diseases. The significance of each coloured block in the flow chart is described on the top right block.MD-PAEDIGREE has the potential of improving the 'prognosis prediction' and 'follow-up program' blocks.*

5. Compliance considerations

Compliance of the MD-PAEDIGREE infostructure with existing standards is essential if one aims to make a useful, sustainable clinical tool out of it. In the DoW of the project, inter-operability with both OpenAIRE and VPH-Share was identified as a priority. The objective of this section is to make sure that requirements related to compliance that are likely to influence the development of the proposed infostructure are clearly identified and taken into account for its development.

Deliverable D13.2 started with a description of the compliance guidelines associated with VPH-Share. This was then followed by a similar description for OpenAIRE. Subsequently, a concise analysis of the impact of compliance requirements on the infostructure was carried out. From the latter, it can be noted that constraints lie in the area of data access, data security, and publishing of novel work based on MD-PAEDIGREE. One can also conclude that despite the fact that the list of compliance constraints is relatively short in D13.2, the resulting compliance with the standards is important for the project. This compliance will likely increase the visibility of research results, increase data reuse (if data can be shared), and thus also potentially increase the number of citations. That type of impact is an important factor for the MD-PAEDIGREE project.

Regarding the VPH-Share-related constraints on data access, the task of making MD-PAEDIGREE compliant is relatively straightforward. That is because mastery of well-documented technologies is required and the latter do not present any significant challenge for system design. In any case, semantic representation of a large majority of the data was planned and would thus favour a full integration into VPH-Share. Using the VPH-Share data acquisition tools will also help to ensure an optimal integration.

The same is true of making the infostructure inter-operable with OpenAIRE, which is the European Union initiative for an Open Access Infrastructure for Research in Europe that promotes open research and provides access to the research output of European funded projects and Open Access content from a network of institutional and disciplinary repositories. It also gathers links from Open Access publications to associated datasets, and using text mining, can infer links between publications, research funding and research data, making this publically available. Results of such processing might include: “other recommended publications”; “publication and dataset cited by a publication”; “datasets used / mentioned in a publication”; “cross-discipline relationships”, etc., which can be provided via the OpenAIRE portal. The steps for MD-PAEDIGREE to follow to achieve the intended goals of compliance with OpenAIRE are well-defined and pose no problem. That will increase visibility of the project and will make sure that scientific reproducibility is guaranteed.

Security is, however, an entirely different matter. VPH-Share has some limiting aspects with regards to its security policy. There is a stringent requirement that any data stored on it should be stripped of all sensitive data behind the firewall of the data-gathering centre. However, it is planned that the system be used for all data, as described in **Table 1**. That points to considerable work that needs to be done to equip such centres with the necessary tools to achieve this goal. The use of VPH-Share data acquisition tools is planned, which should actually make this task easier.

Data Privacy Levels	VPH-Share Compatibility	OpenAIRE Compatibility
Clinicians-only	Yes	No
Consortium-only	Yes	No
Public*	Yes	Yes
Open Access	Yes	Yes

**subject to terms and conditions*

***Table 1**: Planned compatibility of different levels of data access within MD-Paedigree*

In addition, the single username-password security layer might prove inadequate for a pan-European project of the size of MD-PAEDIGREE where security levels are expected to be quite high. This might be the biggest challenge that inter-operability is likely to cause during the development of the MD-PAEDIGREE infostructure, depending on which data really need to be shared and whether this can be limited in the entire project to data stripped of all potentially identifying information. On the positive side, if required, a seamless gateway between MD-PAEDIGREE and VPH-Share can be set up to enable live integration of queries and workflows. Data can then be transferred easily between them for analysis.

6. Link with other initiatives

As is apparent from the previous sections, considerable effort has been and will be made to render MD-PAEDIGREE compatible with VPH-Share. That represents a major step in terms of luring existing users of the latter to the infostructure being proposed.

Moreover, adherence to OpenAire will ensure full visibility of publications and research data that are connected to its development or use, making it a more attractive proposition to potential users. To guarantee a large distribution and usability of the infostructure, links with initiatives such as ELIXIR[4] are also currently being explored. In the same breath, compliance with openAIRE2020[5] (which is a continuation of the openAIRE initiative beyond the EU FP7 framework) is being considered. To achieve compliance with that new initiative, however, all MD-PAEDIGREE-related research data have to be made public and that is considerably more demanding than simple compliance with current openAIRE guidelines.

Furthermore, re-use of Health-e-Child and Sim-e-Child resources and acquired experience has allowed and will continue to allow developers to start all work from a reliable platform instead of starting from scratch. Furthermore, interoperability is sought with p-Medicine[6]. That will be achieved through the use of the grid Gateway concept [1] [2] [3], which is based on a service-oriented architecture (SOA). The infostructure will also implement the service-oriented knowledge utility (SOKU) [4] vision to facilitate the design and development of innovative predictive models as reusable and adaptable workflows. Another added benefit would be that previous users of those tools will benefit from instant familiarity with MD-PAEDIGREE. They would, therefore, find it easy to switch to the more modern and state-of-the-art proposed infostructure.

Work is also being carried out to make the infostructure compliant with the best practices from biomedical semantic web. High-level semantic interoperability with such projects as SemanticHealthNet[7], and DebugIT[8] will enable the wider acceptance of the proposed tool. In the same breath, the use of epSOS[9] terminology and interchange standards will only reinforce the usefulness of MD-PAEDIGREE.

7. Discussions and Conclusion

MD-PAEDIGREE aims to be a state-of-the-art tool for clinicians and researchers. To achieve such a level of sophistication, it is being developed using the best available techniques in the field of medical semantics, clinical workflow representation (VPH-Share), and research dissemination (OpenAIRE). It also reuses existing infrastructure, knowledge, and experience that were obtained through past projects such as Sim-e-Child and Health-e-Child.

Gauging the pressing needs of the intended end users of the proposed infostructure was a top priority at the very beginning of the project. Through interviews with stakeholders early on and throughout the time between the start and the current point the project, technical user requirements were gathered and prioritised. The most up-to-date version was presented in this deliverable. Developers may henceforth use it as a reference in order to implement the most desired features. That will ensure that even the earliest prototype will cover for the most immediate wishes of clinicians and clinical researchers.

Throughout this deliverable, the central message was that the development of MD-PAEDIGREE was based around the philosophy of excellence and durability of the product being created. Many cutting-edge tools in

[4] ELIXIR Data for life. http://www.elixir-europe.org/. Retrieved: 10.02.2015.
[5] openAIRE2020. https://www.openaire.eu/news-events/openaire2020-press-release. Retrieved: 10.02.2015.
[6] p-medicine. http://p-medicine.eu/. Retrieved: 10.02.2015.
[7] SemanticHealthNet. http://www.semantichealthnet.eu/. Retrieved: 10.02.2015.
[8] DebugIT. http://www.debugit.eu/. Retrieved: 10.02.2015.
[9] epSOS. http://www.epsos.eu/. Retrieved: 10.02.2015.

the field are either being incorporated directly into the infostructure or being made interoperable with it. That will help MD-PAEDIGREE gain a good foothold in the field once it is released as a prototype. Subsequently, that will be improved even further through the continual release of new functionalities.

8. References

[1] Amendolia, S. Roberto, et al. "Grid databases for shared image analysis in the mammogrid project." Database Engineering and Applications Symposium, 2004. IDEAS'04. Proceedings. International. IEEE, 2004.

[2] Skaburskas, Konstantin, et al. "Health-e-Child: a grid platform for european paediatrics." Journal of Physics: Conference Series. Vol. 119. No. 8. IOP Publishing, 2008.

[3] Manset, David, et al. "Gridifying Biomedical Applications in the Health-e-Child Project." Chapter XXIV of the Handbook of Research on Computational Grid Technologies for Life Sciences, Biomedicine and Healthcare. ISBN: 978-1.

[4] Future for European Grids: Grids and Service Oriented Knowledge Utilities – Vision and Research Directions 2010 and Beyond, European Communities, 2006.