SHARE the journey

A European Healthgrid Roadmap

SHARE PROJECT

SUPPORTING AND STRUCTURING HEALTHGRID ACTIVITIES & RESEARCH IN EUROPE: DEVELOPING A ROADMAP

Project report October 2008

European Commission
Information Society and Media

About the Report

This report has been prepared on the basis of work undertaken for the European Commission, Directorate General Information Society and Media in the context of a specific support action awarded to the Share team and executed under contract number FP6-2005-IST-027694.

This compact version, compiled by Mark Olive, Hanene Rahmouni and Tony Solomonides, is based on I. Andoulsi; I. Blanquer; V. Breton ; A. Dobrev; C. Van Doosselaere; V. Hernandez; J. Herveg; N. Jacq; Y. Legré; M. Olive ; H. Rahmouni; T. Solomonides; K. Stroetmann; V. Stroetmann; P. Wilson SHARE Integrated Road Map II (SHARE Deliverable D6.2 – http://roadmap.healthgrid.org)

Ref.: Plan-Publi 2008.2321
SHARE the journey - A European Healthgrid Roadmap
Language : EN
Catalogue number : KK-30-08-624-EN-C
ISBN-13 : 978-92-79-09668-6
DOI : 10.2759/27550

Design & Printed by OIB

eHealth

Information Society and Media

SHARE the journey

A European Healthgrid Roadmap

Compiled by : Mark Olive, Hanene Rahmouni and Tony Solomonides

based on I. Andoulsi; I. Blanquer; V. Breton ; A. Dobrev; C. Van Doosselaere; V. Hernandez; J. Herveg; N. Jacq; Y. Legré; M. Olive; H. Rahmouni; T. Solomonides; K. Stroetmann; V. Stroetmann; P. Wilson SHARE Integrated Road Map II (SHARE Deliverable D6.2)

European Commission
Information Society and Media

Summary

Grid technology, one of the key technologies for the 'European Research Area', offers rapid computation, large scale data storage and flexible collaboration by harnessing together the power of large numbers of computers, from end-users' desktops to powerful workstations and clusters of more powerful machines.

The grid was devised for use in scientific fields, such as particle physics and bioinformatics, in which large volumes of data, or very rapid processing, or both, are necessary. The impact of this concept has already reached beyond eScience, to eBusiness, eGovernment and eHealth. However, a major challenge is to take the technology out of the laboratory to the citizen. The term 'healthgrid' is used to describe the application of this technology to biomedical and healthcare informatics. This domain of application presents some difficult challenges.

Europe has already played a major part in the development of grids and healthgrids. The European association HealthGrid (http://www.healthgrid.org) has set the agenda in this domain with its 2005 White Paper (http://whitepaper.healthgrid.org) urging the concept, the opportunities and the likely benefits on senior decision makers. With that vision as its initial point of reference, the SHARE project (http://www.eu-share.org) was asked to identify the key developments needed to achieve wide adoption and deployment of healthgrids throughout Europe. The project was asked to organise these as milestones on a road map, so that all technical advances, social actions, economic investments and ethical or legal initiatives necessary for healthgrids would be seen together in a single coherent document.

The full road map (http://roadmap.healthgrid.org) includes an extensive analysis of several case studies exploring their technical requirements, full discussion of the ethical, legal, social and economic issues which may impede early deployment, and concludes with an attempt to reconcile the tensions between technological developments and regulatory frameworks.

Table of Contents

1 Introduction

Grid technology, one of the key technologies for the 'European Research Area', offers rapid computation, large scale data storage and flexible collaboration by harnessing together the power of a large number of commodity computers or clusters of other basic machines. The grid was devised for use in scientific fields, such as particle physics and bioinformatics, in which large volumes of data, or very rapid processing, or both, are necessary. The impact of this concept is expected to reach far beyond eScience, to eBusiness, eGovernment and eHealth. However, a major challenge is to take the technology out of the laboratory to the citizen.

The concept of grids for health was born in Europe in 2002 and has been carried forward through the HealthGrid initiative. This European collaboration has edited a White Paper setting out the concept and benefits of emerging grid technologies in different applications in healthcare.

The White Paper defines the concept of a healthgrid as follows:

Healthgrids are grid infrastructures comprising applications, services or middleware components that deal with the specific problems arising in the processing of biomedical data. Resources in healthgrids are databases, computing power, medical expertise and even medical devices.

The EU-funded SHARE project has identified some important challenges towards wide deployment and adoption of healthgrids in Europe. The project has devised a strategy to address the issues identified in the Action Plan for European e-Health and has devised roadmaps for technological developments, legal and ethical initiatives and socio-economic investments needed for successful uptake of healthgrids in the next ten years.v

The SHARE roadmaps express certain measurable goals and objectives for the HealthGrid community, provide an analysis of the technical gaps to be bridged in order to achieve a number of staged technical objectives, explore the ethical, legal, social and economic (ELSE) conditions of such developments, analysing the extent to which technology and its environment will need to be reconciled, and articulate a strategy for the concurrent achievement of these goals and objectives subject to realistic contextual conditions.

These roadmaps have been developed from three major inputs:

a) an analysis of user requirements in a carefully triangulated set of domains through current projects and scripted use-cases;

b) a technical road map which sets out the key objectives for a viable 'knowledge healthgrid' to be achieved in a span of 10-15 years;

c) a conceptual map of ELSE conditions, constraints and requirements which must be addressed before a knowledge healthgrid can be deployed in a real healthcare setting.

The conceptual map of ethical, legal, social and economic issues considered the regulatory challenges that any real healthgrid must meet, including: legal challenges concerning rights to privacy and confidentiality, 'right to know' and duty of care; ethical challenges concerning primary and secondary use of data whether individual or aggregated; legal and ethical challenges concerning provenance and quality of information; legal, ethical and economic challenges to the use of healthcare data in commercial and public research, including questions of ownership of data; legal and ethical challenges in the communication of genetic information and the resultant 'lateral leakage' of information; legal and ethical challenges to the communication of medical data across borders; social and legal challenges concerning the formal professional competencies of different healthcare actors; and legal, ethical and socio-economic challenges of 'exceptional cases', such as assisted reproduction, organ donation and transplantation.

2 Benefits of healthgrids

2.1. For the healthcare professional/biomedical researcher

Healthcare systems both in developed and in developing countries face major economic and capacity challenges to maintain quality of care in the face of the growing demands of ageing populations and the increasingly sophisticated treatments available. Add to this the desire to improve access to new care methods, and the challenge of delivering care becomes significant. In an attempt to meet these demands, health systems have increasingly looked at deploying information technology to scale resources, to reduce queues, to avoid errors and to provide modern treatments into remote communities, for example.

From the individualised care point of view, in order for clinicians to make the best diagnosis and decide on treatment all the relevant health information of the patient needs to be available and transparently accessible to them regardless of the location where it is stored. Moreover, computer-aided tools are now essential for interpreting patient-specific data in order to determine the most suitable therapy from the diagnosis.

To store and process medical images, genetic information and other patient data, a large amount of computing power is needed. Large computing resources are also needed for keeping statistics of patient records, for knowledge extraction using data mining, and for the simulation of organisms and diseases using complex biomedical models. Grid technology has undoubtedly much to offer medical professionals, as illustrated by the following examples.

The delivery of medical information and certain services through the internet is familiar. In healthgrid computing, we seek an extension of the concept to consider how to provide large scale services to the user on demand. Some examples will serve to illustrate:

1. Consider a radiologist who needs to manipulate an image: we want to provide a set of services, some of which may require heavy processing, making them available on her desktop 'transparently', as if they were programs simply running on her computer.

2. Consider a public health service which monitors certain infectious diseases and has to trigger an alert in case of a suspected epidemic. The identification of unusual patterns would in many cases be the critical step to halting the problem.

3. Consider a surgical simulation prior to maxillofacial surgery, to determine how the patient's face may appear after one manoeuvre versus another, the presence of sufficient tissue to allow the operation or to demand transplantation, and even to involve the patient in the decision.

4. Consider a 'neglected disease' like malaria. Malaria is neglected by the pharmaceutical industry because there is no prospect of profit in it. Relatively little progress has been made towards the eradication of this well understood disease, notwithstanding substantial investments of public funds in research projects. In silico lead generation may possibly be coupled with investment in plant by the poorer nations that suffer from it to lead to a locally sustainable solution.

5. Consider the possibility of linking genomic information to imaging in diseases like juvenile idiopathic arthritis. The genome will indicate susceptibility long before the disease is expressed, but equally, signs picked up from imaging may obviate the need for genetic screening, thus avoiding some of the most acute problems associated with it.

6. Consider more abstractly the nature of evidence-based practice, the volume of scientific literature that provides the evidence base and the accumulation of evidence from practice that occurs as a matter of routine healthcare. How can these be integrated?

How can they be used without violating any ethical restrictions on use of data, confidentiality, privacy, security? How can they be shared without violating any data protection laws?

These are simple examples of foreseeable beneficial advances within the next generation of developments. For the radiologist or the maxillofacial surgeon, the services described would be important new tools, while for the provider of such services the underlying grid would be an ideal e-marketplace. For the public health service concerned with infectious diseases or the academic unit concerned with neglected diseases, healthgrids would provide power and flexibility beyond what could be achieved using traditional approaches. For a group of physicians seeking to improve treat¬ment through research, as described in the fifth scenario above, the possibility of collaboration and concerted scientific work through a grid offers a completely new concept. Knowledge management, our concluding example, is an issue of increasing concern in the medical domain; a solution of the kind described here would be of immense benefit, even without the additional service of automated policy bridging and compliance.

Of course, there are challenging problems even among these scenarios. Standards have not stabilised in the grid world any more than in current healthcare systems, so data exchanges will present problems straight away. Codes and coding languages are also still not universally adopted, while the application provider will wish to protect investments in software licence rights. These are none the less tractable problems, ones in which the very nature of grids would make easier to address. It is arguably the broader modernisation process that faces the more complex challenges:

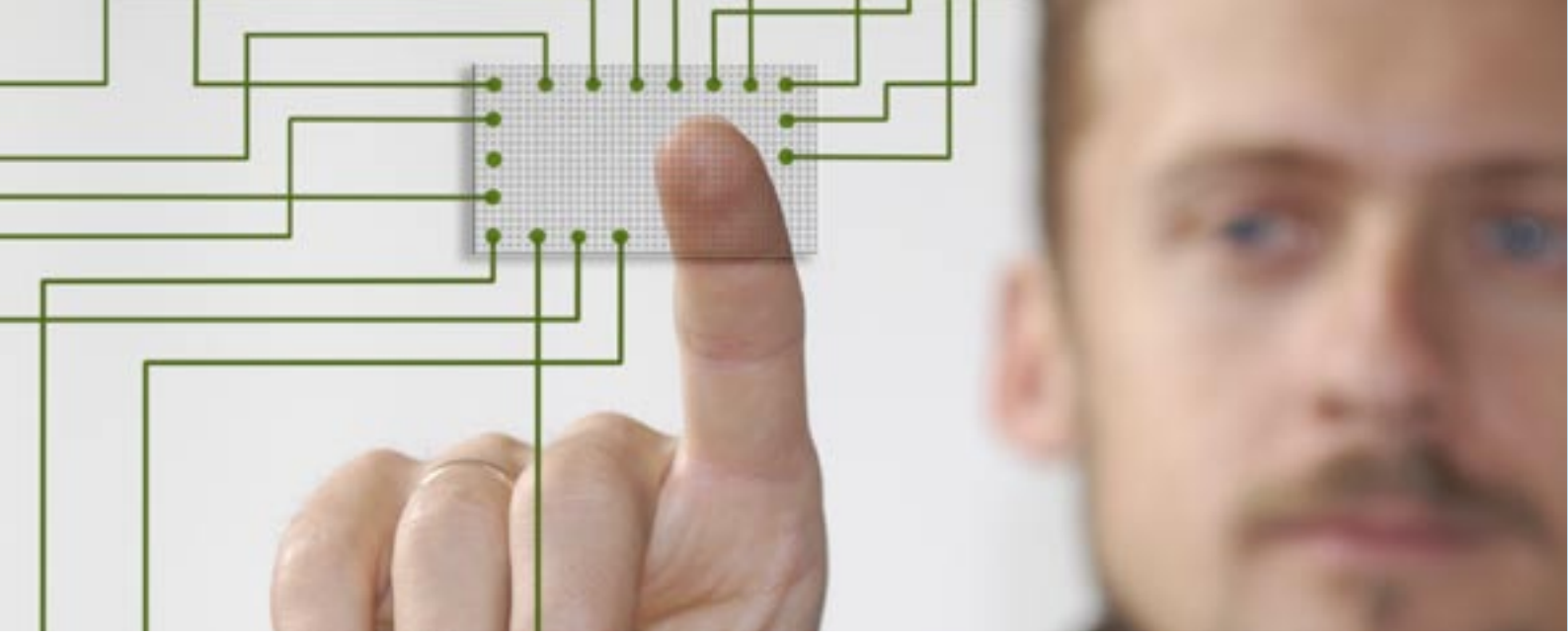
- Connecting and understanding patient records across organisation structures and national borders.
- Ensuring that information is secured and is accessed in compliance with a multilayered regulatory framework.
- Discovering trustworthy sources of information.
- Handling the enormous volume of data involved, for example, in genetic medicine.
- Incorporation of 'legacy' traditional information networks and technology in healthgrids.
- In view of these questions, it is essential to understand also the socio-economic advantages of grids.

2.2. Socio-economic benefits

Modern healthcare services are expected to be available around the clock, seven days a week, so that systems with pervasive access and near-absolute fault tolerance are indispensable. However, it is difficult for these applications to run non-stop with a high quality of service. Grids could help by providing a platform of collaboration, allowing the linking centres which co-operate to achieve better continuity and quality of service. Medical staff will then be able to share experience, knowledge and 'second opinion' with other internal and external staff. The distributed architecture of grids with the availability of high-bandwidth networks responds well to the requirements of healthcare provision. There are also optimistic stakeholders' views towards medical research, healthcare and computing capabilities combined to better satisfy the patient.

Healthgrids promise many benefits to mobile patients as well as citizens. It could help a travelling individual to receive the right treatment in an emergency situation, thanks to the ability of the grid to facilitate communication between the local hospital of the patient and the admitting hospital abroad in order to exchange necessary heath related information.

In addition, healthgrids enable the mobility of a patient within EU states and allow them to receive medical treatment in a country of their choice. This could help solve problems of long waiting lists in states with busy hospitals and lack of medical staff. In economic terms, the



grid could provide an optimal solution for healthcare. It allows a better use of resources and maintenance of tasks, an improved global IT organisation, scalable costs, and a large and consolidated IT business within the healthcare organisation.

Health tourism is a growing concept which can enrich the economy of countries where modern medical treatments (plastic surgery, dental surgery, reproductive medicine, laser surgery for vision correction, etc.) are evolving and having higher success rates than others. This domain can benefit from healthgrid technology as it facilitates the exchange of patient health records and the communication between foreign hospitals and health insurance companies to facilitate the referral and payment process.

Transferring medical images for the purpose of a second opinion to another hospital requires high bandwidth connections between hospitals. Healthgrid technology can provide automated workflows that could be considered a better alternative to manual workflows, such as agreement over the phone and fax transmission of data. These manual workflows are still used by clinicians at present, but are labour-intensive and can cause errors.

3 Use case scenarios

3.1. Grid paradigms

Grids are often differentiated into computational, data and collaboration grids. The ideal grid, envisaged as a servant of a new paradigm of scientific research called 'e-science', must provide transparent processing power, storage capacity and communication channels for scientists who may from time to time join the grid, do some work and then leave, so that the alliances they form in their scientific endeavours might be described as 'virtual organisations' (VOs). Different sciences have different needs, and the grid concept has become differentiated: particle physics generates enormous amounts of data which must be quickly stored, but not necessarily instantly processed; on the other hand, data in bioinformatics is not large by comparison but requires intensive processing. In extending the application of grids to e-health, collaboration becomes pre-eminently necessary.

3.2. computing grids: an example from innovative medicine

Drug discovery is the long term, multi-stage and high cost process by which drugs are discovered and/or designed. The drug discovery goal is to find new molecules that bind with specific macromolecules known to play a key role in a disease process, in a manner that changes their function, either to increase resistance to or to reduce the virulence of some pathogen. Reducing the research time in the discovery stage and having enhanced information about substances affecting the selected target ('leads') are key priorities for pharmaceutical companies worldwide.
In silico drug discovery, including analysis of the gene expression data, target function prediction and target three-dimensional (3D) structure prediction, is one of the most promising strategies to speed up the drug discovery process, avoiding time consuming and costly in vitro and in vivo tests. It is estimated that the efficiency gains of this approach could result in 35% cost savings, or about US$300 million, and a 15% time reduction, or two years of development time per drug.[1]

In silico drug discovery requires advances in data integration (including data format standardisation, dataflow definition in a distributed system, services for data and meta-data registration, and development and sharing of ontologies and knowledge representations), workflow enactment to ease data management and data mining and to assist the scientist and the decision-maker in organising their work in a flexible manner, access to computing and data resources (computing 1 million docking probabilities or modelling 1000 compounds on one target protein requires a few TFlops for one day), and collaboration between public and private partners, involving the concrete sharing of data and knowledge, software and workflow, and infrastructures such as computing, storage and networks. Security and the effective protection of intellectual property and sensitive information are key challenges for pharmaceutical industries, but also for academic institutes in most cases.

3.3. data grids: examples from epidemiology and public health informatics

3.3.1. Epidemiology

The epidemiology use case is defined as a system able to link the information from distributed and heterogeneous databases, identify patients, complete episodes and automatically improving quality without interrupting clinical practice. With this data, complex epidemiological models are fed and simulated producing aggregated prospective results, in a reliable way. This use case is

1. This is discussed in greater detail in SHARE deliverable D5.2b The Innovative Medicine Case Study (http://eu-share.org/about-share/deliverables-and-documents.html)

representative of different applications and systems including ontological information systems, infectious surveillance networks, pharma-epidemiology analysis of efficiency and cost, and study of propagation models for diseases.

The main users (from the highest-concept level to lowest) are public health authorities, epidemiologists and pharmaceutical companies. The data is normally owned by clinical care (both public and private). In order to achieve this use case the user needs to go through different steps including automatic data gathering, data quality improvement, processing of the data, and presentation of results.

3.3.2. A surveillance network for avian flu

A use case in the field of public health informatics is a surveillance network of data repositories offering services to the research communities working on avian flu and issuing an alert in case of pandemic risk. Indeed, the ability of the international community to respond efficiently to the possible emergence of a human-to-human transmissible avian influenza virus depends on its capacity to quickly assess any evolution of the disease. Many countries have set up very efficient national networks for collecting data and monitoring outbreaks. However, there is presently no international surveillance network that allows the sharing of data collected at a national level.

The starting point would be to set up a data grid collecting public and private information on avian flu. Public data would be made available to all registered users while access to private data would be strictly controlled through grid authorisation and authentication mechanisms. The repositories would share a common model allowing distributed queries.

3.4. collaboration grids: examples from tertiary healthcare and vph

3.4.1. Virtual Physiological Human (VPH)

The EuroPhysiome initiative has led to the concept of the Virtual Physiological Human (VPH). This concept goes beyond the idea of a model to an altogether more ambitious methodological and technological framework to enable the investigation of the human body as a single complex system. Just like the biological human, the VPH will be made up of interacting subsystems which need to be understood at a hierarchy of levels, from the molecular to the cellular, through organs and systems to the whole individual. These are often associated with a different discipline of scientific knowledge and have been served by a different specialism within informatics. The metholodological aspect of VPH as a project is to integrate these into an appropriate framework within which observations made in laboratories, hospitals, and in the field all over the world can be collected, catalogued, organised, shared and combined in a variety of ways. It should also allow experts collaboratively to analyse observations and develop systemic hypotheses that involve the knowledge and insights from multiple scientific disciplines. The ultimate scientific goal would be to interconnect predictive models defined at different scales, with different methods, and with different levels of detail, into systemic networks that allow complex hypotheses to be formulated as models which can be tested through simulation and experiment. This may be illustrated by the study of tumour growth in cancer, in which molecular and cellular processes may be studied together to understand first what happens in the individual cell and how these proliferate to form a malignant colony.

There is a great deal of similarity in the conception and ambition of the VPH and of healthgrids. The vast scope and integrative approach of the VPH call for the kind of resource sharing mechanisms provided by a grid infrastructure. The need for a medium of collaboration for a large but heterogeneous community of scientists and the need for coherent multi-scale modelling likewise argue in favour of a grid approach.

3.4.2. Tertiary and Specialist Healthcare

The Health-e-Child project dealing with paediatrics, and the MammoGrid and e-DiaMoND projects dealing with breast cancer, are examples of projects that have made use of grid technologies to benefit healthcare professionals and patients in a tertiary healthcare setting.

Paediatrics is an area where not only the disease in question is changing, but the child as well, as he or she grows. The Health-e-Child project addressed specific diseases within certain medical domains; cardiology, rheumatology and neurology. Through consultation with these communities, a number of requirements for the project were identified. The primary requirements from cardiology are related to imaging and integrated disease modelling. Requirements from rheumatology include the construction of homogenous juvenile inflammatory arthritis subtypes (which will require vertically integrated data), and models for predictive disease outcome, and for progressive organ damage. The project determined that a variety of models using integrated data would be useful for neurologists, including models for surgical decision-making, post surgery treatment, and models supporting

automatic tumour detection and change quantification. The potential for individualised brain models has also been recognised, which will require deforming a generic brain atlas in order for it to match the geometry of a patient.

Breast cancer as a medical condition, and mammograms as images, exhibit many dimensions of variability across a population. Likewise, the way diagnostic systems are used and maintained by clinicians varies between imaging centres and breast screening programmes, as does the appearance of the mammograms generated. A distributed database that reflects the spread of pathologies across a broad population is an invaluable tool for the epidemiologist; understanding the variation in image acquisition protocols is essential to a radiologist in a screening programme. Exploiting emerging grid technology, the aim of the MammoGrid project was to develop a potentially EU-wide prototype database of mammograms to be used to investigate a set of important healthcare applications and to explore the potential of the grid to support effective collaboration between healthcare professionals. In particular, the project aimed to prove that grid infrastructures can be practically used for collaborative medical image analysis. MammoGrid and e-DiaMoND effectively demonstrated the viability of the grid by harnessing its power to enable radiologists from geographically dispersed hospitals to share standardised mammograms, to compare diagnoses (with and without computer aided detection of masses and microcalcifications) and to perform sophisticated epidemiological studies across national boundaries.

3.5. knowledge grids: an example from general healthcare

For any given domain, a distinction is often drawn between declarative knowledge ('know what') and procedural or operational knowledge ('know how'). In the domain of healthcare, both kinds occur. What is often referred to as 'the scientific basis' of medicine, that which must furnish the evidence in so-called 'evidence-based practice', is present in research publications to which different standards of credibility are attached. For example, research results based on a randomised, double-blind, controlled clinical trial are held to be the gold standard, provided they were also submitted to adequate peer review. Evidence based on one physician's own practice, although not negligible, would be considerably less reliable. On the other hand, 'best ways' of treating patients – in a particular context – are often described in integrated care pathways (ICPs). It is not unreasonable to claim that declarative knowledge in medicine tends to be disseminated through peer-reviewed publication and operational knowledge through such things as guidelines and care pathways. In a healthgrid environment, these are brought together for the better treatment of patients and at the same time to improve research; indeed, the interplay between healthcare and research, e.g. through appropriately controlled 'secondary use', would be an important element in a full healthgrid environment.

Chronic obstructive pulmonary disease (COPD) refers to an airway obstruction caused by chronic inflammation. It is usually progressive, not fully reversible, and often occurs as a result of smoking but other factors such as air pollution can also contribute to the development of COPD. In the UK almost 900,000 people have been diagnosed with the disease, and the true number of people suffering from the condition is estimated to be around 1.5 million.

According to NHS guidance, the management of the disease should be tailored to the individual, with adjustments being made based on responses to treatment. The guidance includes a large number of drugs, including some off-label drugs. In the UK, the National Institute for Health and Clinical Excellence (NICE) has issued national guidelines for the treatment of the disease, but these are frequently modified to account for local variations and priorities. As a result, the procedure for assessing and treating the disease will vary even within a single country, let alone between countries. The evidence on which this guidance is based comes from a variety of sources, such as national studies by NICE and systematic reviews with an international scope.

Two main concerns exist for general healthcare; supporting the travelling patient, such as migrating elderly populations, and enabling decision support systems that can account for local variations in best practice and clinical evidence.

4 Technical, ethical, legal and socio-economic issues

4.1. Grid technology

A 'grid' – not the grid – is now understood to mean an Internet-like infrastructure which extends the concept of the Internet in several significant ways:

- like the Internet, a grid would provide access to information services but in addition would provide pooled storage, processing power and collaboration in so-called 'virtual organisations' (VOs);
- use of a grid will be reciprocal – while a user subscribes and takes advantage of services provided by a grid, the user's resources are pooled and are available to all grid subscribers;
- the process is transparent – the grid allocates resources and provides an interface to services which give the appearance that the user is accessing just one powerful machine.

Major IT companies have agreed to develop web services as the technology to enable the deployment of services on the Internet. It has been also adopted by the Open Grid Forum which is the acknowledged body to propose and develop standards for grid technology.

Moreover, web service technology provides the bridge between the grid world and the Semantic Web which is about common formats for interchange of data and about language for recording how the data relates to real world objects. Although many current grid infrastructures do not offer a web service interface to their services, we will concentrate our 'state of the art' on web services because it is the relevant technology for the future. We will then go on to discuss the status of existing grid infrastructures, the technologies they use and the services they offer.

The initial idea behind web services was to enable the World Wide Web increasingly to support real applications and a means of communication among them. Thus, a set of standards and protocols have been proposed to allow interaction between distant machines over a network. These interactions are made possible through the use of standardised interfaces which describe the available operations in a service, the nature and form of messages exchanged (requests and responses), and the physical location of the service on the network. A language, Web Service Description Language (WSDL) has been devised to describe such interfaces.

One of the subtle advantages of a so called Service Oriented Architecture (SOA) is that it leads to loosely coupled components that may be substituted by better services so long as they comply with the interface specification. Thus in a grid, services can be offered in a way that approximates an ideal marketplace, a market with perfect information.

4.2. Technical issues

Basic technical requirements include the adequacy of network infrastructures for the effective operation of a grid, the usability of grid interfaces and simplicity of administration of grid nodes. Other healthgrid-specific issues can be divided into standardisation, communication and security.

4.2.1. Standardisation issues

Standards are essential for the deployment of services which integrate data in bioinformatics and medical informatics, but are also vital for data coming from different medical disciplines or from different countries

in Europe. These standards are needed for building data models, producing ontologies and the development of knowledge management services. The adoption of standards for the exchange of biological and medical information is still limited to a few specific fields. Moreover, they need to be compatible with grid standards so as to allow their implementation on healthgrids. Of course, many of these issues arise in any data integration programme, but they appear all the more sharply in grids where the heterogeneity of the underlying system is expected to be transparent, or at least not to impact on the user.

4.2.2. Communication issues

Lack of information about grids and grid technologies is frequently identified as one of the key reasons why there has been very little interest in them from the field of medical research. It is essential that all relevant actors to be kept well informed by the HealthGrid community of the potential benefits of the technology to them. Success stories demonstrating the impact of grids for medical research will be vital for convincing medical researchers of these benefits. As a result, there is a need for a suitable demonstration environment, offering very easy access to the grid for non experts and providing services that will help convince the medical research community. On this dissemination environment, dedicated efforts to promote the technology can then be developed.

4.2.3. Security issues

Deployment of a data grid for medical research will only be possible when the middleware can provide all the necessary guarantees in terms of management of personal data. We perceive the specific technical requirements related to the handling of medical data on the grid to be as follows.

- Manipulation of personal data on the grid must obey strict regulations. These regulations vary between European member states.
- Services for the anonymisation and pseudonymisation of medical data must be provided.
- Medical data is the property of the patient. A mechanism must be set up to allow individuals to access their data on the grid.
- For healthcare purposes, the authentication of healthcare professionals on the grid cannot be handled by requesting all of them to get a grid certificate. A mechanism must be set up so that professional cards can be used to provide authentication on the grid.

4.3. Ethical issues

Ethical issues in healthgrids correspond to well known ethical principles, autonomy, beneficence and justice. Each has individual value, but the three must be taken as a whole offering a system of ethics in which the needs of the individual are balanced against those of society.

4.3.1. Autonomy and Healthgrids

Most common belief systems give a special place to the autonomy of the individual - the right of the individual to control his or her own person. Autonomy is intimately tied up with the legal duties of consent and confidentiality, both of which could prove difficult in the context of healthgrids. Thus, in healthgrids, one of the key ethical issues will be in the possible compromise of the patient's autonomy that will arise from sharing his or her data with people who are yet to be identified. It is worth noting that it has been argued, notably by the European Article 29 Data Protection Working Party, that consent may have only a very limited place as a justification of the sharing of health related data in the electronic age. The limitation is based on the argument that to exercise autonomy one must be able to make decisions unfettered by coercion. If

it is accepted that sharing health data allows doctors to provide better care, then a patient who refuses permission to share such information will be de facto opting for a lower quality of healthcare, arguable therefore he or she is not able to withhold consent freely and is therefore not able to act autonomously. It is argued therefore that robust system of security of information and ethical practice should be adopted in which patients will be able to trust, notwithstanding that their information is shared, and providing for special opt-out possibilities when the nature of the information is especially sensitive.

4.3.2. Beneficence and Non-Malfeasance

The ethical duty of Beneficence and Non-Malfeasance is the duty to do good, or in the words of the Hippocratic Oath at least to do no harm. This ethical duty is frequently used to justify the adoption of specific health technologies to improve treatment. The argument with respect to healthgrids is that in order to act ethically, a healthcare professional is obliged to use suitable grid applications if they are available. A healthcare professional refusing to use standard medical technology or refusing to prescribe antibiotics would be considered in breach of his or her duty of beneficence, thus, as the sophistication of grid aided diagnosis develops we will one day arrive at a time when a practitioner not linked to the appropriate grid networks will be in breach of his or her duty.

However, until we have reached a time when grid applications are stable, well 'fed' with data and fully integrated into the evidence base of good clinical practice such arguments will not apply. At present, in the more experimental stages of the healthgrid it will be important to ensure that the use of the applications does no harm, but perhaps most importantly to ensure that the patient is aware of any possible medical and social damage.

4.3.3. Justice

The ethical principle of justice concerned with the duty to achieve a fair distribution of resources as well as the need to develop an overall just medical system in which the greatest health benefit of the greatest number is achieved is the principle of justice. In most legal systems this ethical principle is used to support social systems of distributive justice which provide for tools as taxation to distribute wealth on such a way that all may be afforded an acceptable minimum of social care. The developments of applications such as MammoGrid have established that the sharing of a very large number of mammogram images across a wide network that allows radiologists to test suspect images against a known and tested database of cases significantly contributes to the early detection of breast cancer. The healthgrid in this case not only acts to the benefit of the known patient whose suspect image is submitted to the tool, but to the overall health of the population.

4.4. Legal issues

The legal issues presented here and further analysed in SHARE deliverables D4.1 and D4.2 were chosen for their relevance to healthgrid technologies. Other legal concerns such as competition issues are of relevance for grid technology, but in order for a full and complete analysis to be made, these were intentionally omitted. Regarding competition issues in relation to eHealth, please see the reports and analysis from the Legally eHealth project.

4.4.1. Data Protection

The ethical principle of autonomy is legally underpinned by the duty of data protection. This is taken very seriously at EU level; as well as the inclusion of privacy protection in the European charter of fundamental rights, the EU has developed the robust directive on data protection to promote privacy.

However, the current legislation is not adequate to support most of the longer running research initiatives around which healthgrids are based. As the current EU level legislation stands, member states can enact specific legislation covering specific tools such as healthgrids in order to exempt scientists and medical practitioners using healthgrids from some of the more onerous duties of the directive.

No member state has addressed legislation to this particular issue and so healthgrids are burdened with onerous data protection requirements which could deter scientists from adopting healthgrid technology and using its enhanced computational and data acquisition power.

4.4.2. Liability

Using grids blurs the liability issues in terms of medical practice. While the EU has a range of legislation designed to protect citizens from harm resulting from goods offered on the market, the construction of healthgrids makes it difficult to ascertain at which EU level legislation would apply to each part of the system.

This is particularly the case with the law on medical devices, which is unclear with respect to healthgrids. In September 2007, the European Parliament and the Council adopted Directive 2007/47/EC of 5 September 2007 amending Council Directive 90/385/EEC on the approximation of the law of the Member States relating to active implantable medical devices, Council Directive 93/42/EEC concerning medical devices and Directive 98/8/EC concerning the placing of biocidal products on the market (OJ, L 247/21, 21.09.2007). In particular, it is stated in Recital 6 of this directive that "it is necessary to clarify that software in its own right, when specifically intended by the manufacturers to be used for one or more of the medical purposes set out in the definition of a medical device, is a medical device. Software for general purposes when used in healthcare setting is not a medical

device." An amended definition of what a medical device is also to be found in the directive.

4.4.3. Intellectual Property

In the EU, legislation dictates that the owner of copyrighted software running a healthgrid has exclusive rights to reproduce his work, produce derivative works, distribute copies to the public, and perform and/or display the work publicly. Under these circumstances any natural or legal person would have to pay to use the computer programs that constitute one of the most important components of healthgrids. Conversely, an open standards approach to software co-development could help in the development and implementation of healthgrids. The open source licensing model actually uses copyright and contract principles to retain control of the work while enabling its use effectively for free, and could thus encourage use and development.

4.5. Organisational, social and cultural issues

Both at the individual and the societal level, issues like universality of availability of full healthcare services to all citizens, equal access to healthcare, and equally high quality of services are key issues. Geographical factors relate mainly to equal access to quality care independent of location. ICT-based systems pose new problems like access to EHRs by insurance companies or employers, and even police and prosecutors. The opinions and attitudes of patient and citizen associations and lobbying groups, often magnified by the media, could have a strong impact through public (policy) discussions of these topics on the implementation and diffusion of healthgrids.

The organisational level is always complex. Perspectives, confirmed by two recent research studies, include:

- Changing care pathways that need new information, skills, knowledge and processes for healthcare providers
- Changing roles of healthcare professionals, teams and healthcare organisations
- The transfer of roles between healthcare professionals, teams and healthcare organisations
- Increased collaborative working and exchange of information between providers
- New relationships between citizens and healthcare professionals and organisations
- New strategic partnerships for third party financers and healthcare providers.

Cultural issues are a key factor in health services, including the great diversity of attitudes, behaviour and knowledge exchange among professional and non-professional staff involved in healthcare, and the impact this has on the quality, efficiency and organisation of services. Education and training, professional standards and bodies, rules and regulations, attitudes and behaviour all have an influence here.

5 Requirements from use cases

An analysis of user requirements from the use case scenarios outlined in chapter 3, with reference to the technical, ethical, legal and socio-economic issues identified in chapter 4, lead to the formulation of the following challenges and requirements for each use case.

5.1. Innovative medicine

The European Federation of Pharmaceutical Industries and Associations (EFPIA) identified pre-competitive barriers to innovation. The objective for the future would be to identify as soon as possible in the pre-clinical phase the reasons for lack of efficacy, despite promising pre-clinical data, and the potential for adverse drug reactions and pre-clinical toxicity. The identified key bottlenecks in the R&D process are:

- predictive pharmacology at the discovery research stage;
- predictive toxicology at the preclinical development stage;
- identification of biomarkers at the translational medicine stage;
- patient recruitment and validation of biomarkers at the clinical development stage;
- risk assessment with regulatory authorities at the pharmacovigilance stage.

The knowledge management area is identified as key to leveraging the potential of new technologies such as genomics and proteomics and to analyse the huge quantity and diversity of information in an integrated way, such as the capture, analysis and interpretation of knowledge generated for one potential drug candidate from discovery, non-clinical and clinical development all the way to lifecycle management. One of the major bottlenecks is the lack of availability of databases across R&D that might facilitate data integration.

5.2. Epidemiology

The main challenges to this use case include the general problem of access to distributed, critically sensitive and heterogeneous data, resulting in costly computing processes. Patient-centric analyses normally deal with smaller amounts of data and require a pre-existing knowledge of models of healthy and diseased organs or tissues. Population-level analyses normally deal with the integration of larger, poorer-quality data. Semantics are especially relevant in those approaches.

The research requirements for the epidemiology use case can be summarised as follows:

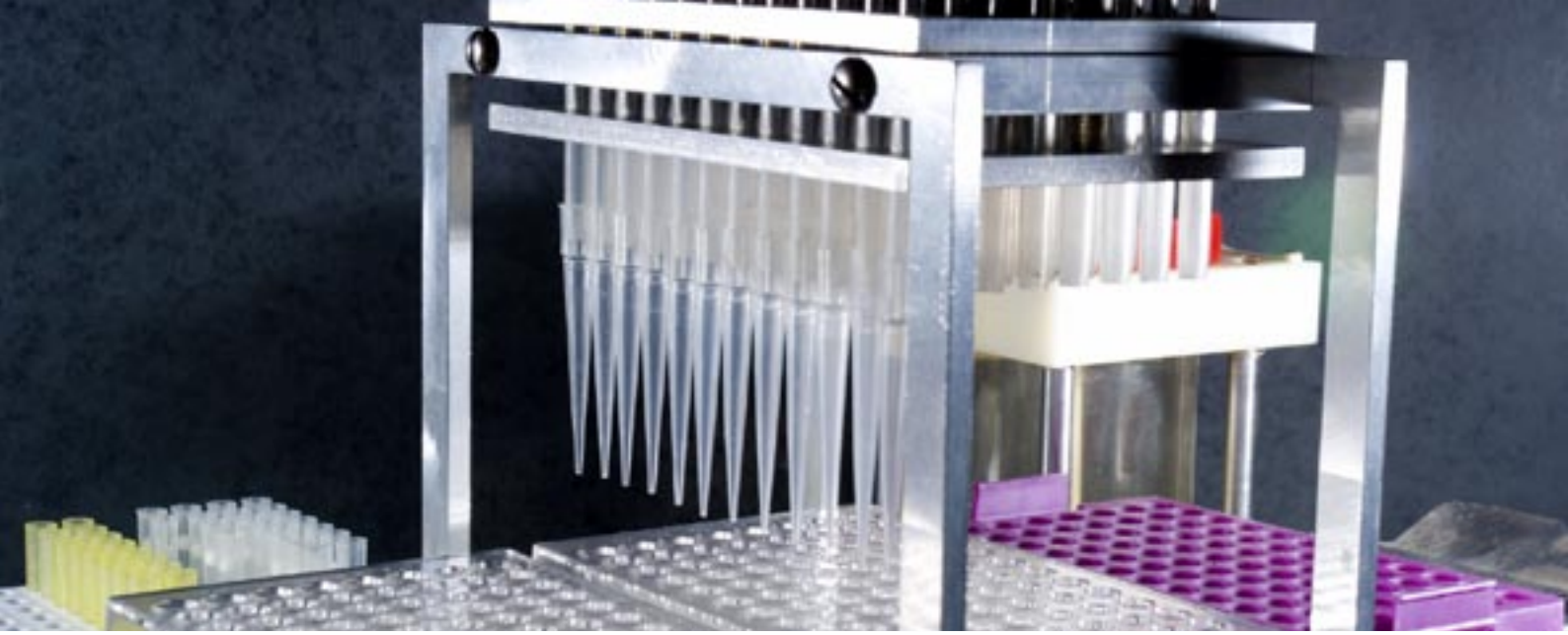
Effective semantic annotation of data. Data is poorly coded and interoperability of coding is not trivial. Extracting knowledge from medical data, however, is a main objective.

Effective integration of distributed and heterogeneous data. Integrating distributed resources requires exchange protocols, secure mechanisms, patient de- and re-identification, and automatic data analysis services.

Availability of efficient infrastructures and usage policies. Applications will require sufficient resources and a reliable infrastructure to work on under a clear Quality of Service (QoS) promise.

Robust security mechanisms insuring that the different components of the infrastructure behave in way that conforms to the European norms controlling the handling of personal data.

Usability of applications and services. The tools should be available through protocols and interfaces similar to those used in the users' normal research. Not only must the applications be as compliant as possible with current systems and interfaces, but so must the technologies.



Ensuring that the research is done in a secure, ethically and legally compliant framework.

Reliability, scalability and pervasiveness. All the previous services must be robust and should be scalable without reducing performance.

5.3. A surveillance network for avian flu

The two main concerns in this use case are the evolution of the virus and its capacity for human to human transmission.

Services for molecular epidemiology. Once a new virus stream has been sequenced, its comparison to the previously identified streams allows the evolution of the virus to be measured. Genome sequences can also be used to guide the search for vaccines and drugs. The data must be properly integrated and its exploitation must build upon expert skills in epidemiology.

Mathematical modelling and computer based simulations for the testing of outbreak hypotheses. Disease epidemiology requires measuring the evolution of the disease in time and space using GIS (geographic information systems) and environment data as well as understanding the disease transmission, reservoir, immunity and treatment.

The reliability, relevance, and completeness of the data stored. Reliability depends on the mechanism to collect data. Relevance depends on the mechanism to update data while completeness is influenced by the capacity to collect data from multiple scientific disciplines and multiple countries.

The usability of the environment. Scientists are so busy they will not take time to contribute information and operate services if these services are not easy to use.

The integration of several types of data for disease monitoring: medical / epidemiological data on human and animal cases, geographical data for each outbreak (location, number of cases, geographic distribution of the casualties, environment, and population density), virus genome sequences, philogenetic trees, and proteomics.

Text mining services. In order to extract information from this literature on demand, advanced text mining services will be required.

5.4. Virtual physiological human (vph)

With respect to grid computing, the following challenges and requirements have been identified:

Knowledge management. The accumulative nature of VPH requires that everything is organised within solid knowledge management models. This should make it possible to keep even very large information spaces organised and usable.

Fostering grid adoption. To foster grid adoption in the VPH community, it is highly recommended to identify a few VPH CPU intensive applications which could benefit immediately from deployment on existing grid infrastructures like EGEE or DEISA. The deployment of these applications will allow the identification of missing services on the existing infrastructures, and will raise the awareness of grids in the VPH community.[2]

2. Some FP6 related projects, such as VIROLAB, Health-e-child and @Neurist, may serve as models.

Access to resources. Researchers require access to all available resources in a uniform way, from those provided by their own department to specialised HPC resources. Access to these should be as seamless as possible, with simulations at different scales being automatically migrated and appropriate resources being used as required.

The usability of interfaces. Current grid portals require the user to specify parameters such as memory to be allocated at execution time; this would not be appropriate for VPH users.

Grid usage models. The nature of VPH simulations means that processing is time critical, and current models of HPC use would not be appropriate. Instead, models which permit a large number of grid nodes to be used for a relatively short time ('burst mode') with little or no waiting time should be established. Resource co-reservation will also be required, particularly where multiscale simulations that run over multiple sites are concerned.

Shared storage for large data/model repositories. The imaging datasets concerned can be several hundred megabytes in size, but after pre-processing to generate a predictive model, the modelling and simulation data can be as much as a hundred gigabytes in size. With potentially thousands of these, there is a clear need for multi-terabyte storage facilities, connected to distributed HPC resources via high speed networks. These must incorporate the required security and confidentiality measures.

Methods to solve multiple predicted models, in a coupled way. As the coupling of predictive models at different scales is central to VPH's description of human physiology, coupled methods to solve multiple models will be required. This is considerably complicated by the fact that the models concerned may be very different both in conceptual nature and mathematical nature. Even relatively simple VPH problems can be considerably complicated by variations between individual subjects and treatment procedures.

Direct prediction from medical images. Methods for transforming a medical imaging dataset into a subject specific predictive model that do not require the costly pre-processing phase are being developed, such as the Boltzmann Lattice in haemodynamics and voxel meshes for hard tissue simulations. However, these are enormously computationally intensive, requiring fifty or more teraflops of computational power to solve in less than a day.

5.5. Tertiary healthcare – the mammogrid, e-diamond and health-e-child projects

Several technical issues were identified by MammoGrid and e-DiaMoND, including the standardisation of mammograms, design of an appropriate clinical workstation and distribution of data, images and clinician queries across a grid-based database while respecting patient confidentiality and security protocols.

Image quality. While clinically significant signs are subtle, many parameters also affect the appearance of an image. For mammograms, these include image acquisition parameters, such as degree of breast compression, tube voltage and beam intensity, and anatomical and physiological data, which show marked variation across the population, at different times in the menstrual cycle and throughout the course of a woman's life. The way diagnostic imaging systems are used and maintained by clinicians also varies between imaging centres and breast screening programmes. In order to study the epidemiology of breast cancer, it is necessary to understand this variability. This is also a prerequisite for the integration of computer-aided detection (CADe) tools and quality control in the process, as well as for comparison between images and for radiological training programmes.

Image standardisation. There are a large number of standards that are sometimes conflicting or overlapping with worldwide accepted standards like DICOM and the de-facto SMF standard for the standardisation of mammograms. These must be disambiguated and adhered to for a production grid to be deployed even at a national level.

Storage of medical images. Images have to be stored unaltered and uncompressed for legal reasons. Also, several copies of the mammogram need to be stored which increases the storage requirements. The e-DiaMoND project decided to use expensive high speed disks for storing the currently active data (those currently undergoing treatment), and cheaper lower performance disks for storing data that is being migrated to or from long term slower storage. For the MammoGrid project only metadata about the images were stored in the shared database.

The following additional requirements were identified by the Health-e-Child project:

Vertical integration of clinical data. Practitioners within different medical disciplines require information at different levels and with different granularity. Integrating genomic, imaging and proteomic data could provide medical practitioners and researchers with a unified, coherent view of the patient's current and past health.

Storage and sharing of biomedical information. Biomedical information should be accessible from geographically distributed sources, including information from the local hospital intranet and databases such as gene databanks on the internet. Access to data must be secure, with appropriate levels of encryption, and access rights must be enforced. Additionally, data sources must have sufficient availability and responsiveness to assure an appropriate quality of service.

Biomedical query processing. Clinicians from different disciplines will require different views of the integrated record provided by the project. Given the vertically integrated nature of the patient records, these queries will be considerably complex, but must be executed sufficiently fast for use by clinicians during consultations or treatment.

Integrated disease modelling. Disease modelling is required at all levels addressed by the project, from molecular modelling such as searching for gene defects, to in silico physiological models of the whole body at a high level. These models should aid clinicians when attempting to determine if a patient has or is likely to develop a certain disease, and what the best treatment might be.

Tools to support image annotation. Image annotation tools should support the creation and modification of annotations by cardiologists, rheumatologists and radiologists, and should be stored as metadata. For second readings by radiologists, standard ways of annotating images are needed for better collaboration between experts.

Queries to find records of similar cases. Based on similarity criteria specified by the user, a mechanism should exist to find and display similar cases, with any accompanying annotations, metadata and records of clinical decisions taken.

Macroscopic computational models for key organs and diseases. Generic models should be constructed, which can then be adjusted according to patient-specific clinical data in order to produce individualised models.

5.6. General healthcare - chronic obstructive pulmonary disease (copd)

Evidence from national studies, such as the aforementioned NICE study, may not be available to a doctor from a different locale. In order to continue treating the patient concerned, the doctor (or decision support system) must be aware of the evidence and guideline/pathway that informed the plan of care for that patient, and any deviations from that plan that have occurred to date. This may not a trivial matter of simply retaining a link to the relevant material, as there may be language barriers, and local reasons why the guidance followed in one country would not be appropriate in another.

The guidance also mentions drugs that are not certified for the treatment of COPD (off-label) despite this evidence coming from high quality systematic reviews. Different drugs will be certified for the treatment of the condition in different countries, complicating the process of following a single guideline or pathway regardless of travel. In fact, the patient concerned may be travelling for the express purpose of receiving different or less costly treatment in another country.

Prior history of exacerbations and smoking are essential for properly treating the disease, and therefore the doctor concerned must be able to access, comprehend and update the patient's record. This requires standardisation of electronic health records (EHRs) and electronic integrated care pathways (eICPs). When it comes to decision support, a standard interface format, such as the proposed HL7 vMR, will also be a necessity.

6 Recommendations and roadmaps

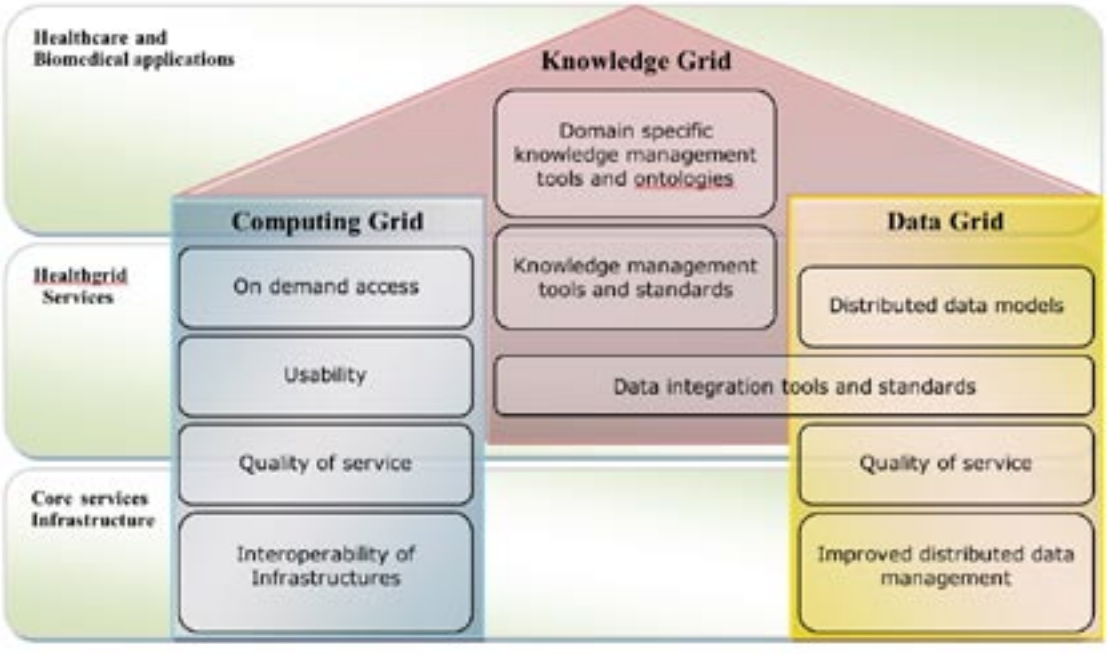
In order to translate the broad range of issues highlighted in our discussion above into a manageable programme of necessary developments, we have adopted certain broadly acceptable assumptions: first, that the fundamental architecture will be service-oriented; second, that it will be necessary to distinguish and layer, from the bottom up, core infrastructure services, generic healthgrid services and specific biomedical or healthcare application services; and third, that the typology of grids will be reflected in the nature of the challenges. Moreover, there are interdependencies between challenges. Thus challenges have been organised according to a number of key criteria:

- Is the proposed healthgrid essentially a computational grid, a data grid or a collaboration grid? Could it potentially develop into a knowledge grid?
- Is the necessary development to achieve any given stage likely to be delivered by generic grid research or is it particular to healthgrids?
- Is some prerequisite standard or other agreed framework necessary for the achievement of any particular milestone?

6.1. Proposed roadmaps

In this section, we present a technical roadmap for the adoption of the grid technology for healthcare. In the previous section, for three families of grids, computing, data and knowledge grids, we have identified a number of research challenges which have been characterised by key words. The following diagram represents how research challenges address different layers of services from core infrastructure to applications.

Figure 1: research challenges and their relationship to healthgrid service layers and grid family



Reading the diagram from left to right, we note two pillars of challenges for computational and data grids, with knowledge grids supportedon these two pillars. For simplicity, the challenges of collaboration grids have been subsumed under those of computational and data grids. The image conveys the sense that knowledge grids will be built at a higher level (and, necessarily, at a later stage) than computational and data grids.

Reading the diagram from the bottom up, we have depicted challenges according to their closeness to the underlying infrastructure, or in terms of our criteria above, whether they are likely to be addressed by generic grid services, by healthgrid service or by specific applications.

The following comments can be made from the picture:

- Interoperability as well as improved distributed data management must be core functionalities of the infrastructure
- Quality of service is required from both core and healthgrid services for successful healthcare / biomedical applications
- Healthgrid services should be accessible on demand, with a high degree of usability. Distributed data models must also be provided.
- Data Integration Tools and Standards are healthgrid services which stand at the interface between data and knowledge grids.
- Knowledge Management Tools and Standards require the availability of proper job and data management tools. They stand at the interface between generic healthgrid services and the application specific developments.
- Domain Specific Knowledge Management and Ontologies are the responsibility of individual research communities. Their interface to the knowledge grid is achieved using Knowledge Management Tools and Standards.

On the basis of this analysis, we have represented in figure 2 the research challenges according to their complexity and an estimated time when they should be overcome. The figure inspired from The Innovative Medicine Case Study[3], also indicates the level of adoption by the research communities. As can be seen clearly from the picture, we identify several distinct roadmaps:

- Research and development for computing grids should allow offering the quality of services needed for biomedical research and healthcare at a 5-year horizon.
- Data grids are expected to reach maturity at a 10-year horizon as the core technology is not yet mature.
- Collaboration grids are achievable with different levels of sophistication at different stages.
- Knowledge grids depend on the quality of services for distributed data integration and the capacity of the research communities to agree on standards and ontologies. As a consequence, their maturity is not expected before 15 years.

3. SHARE deliverable D5.2a

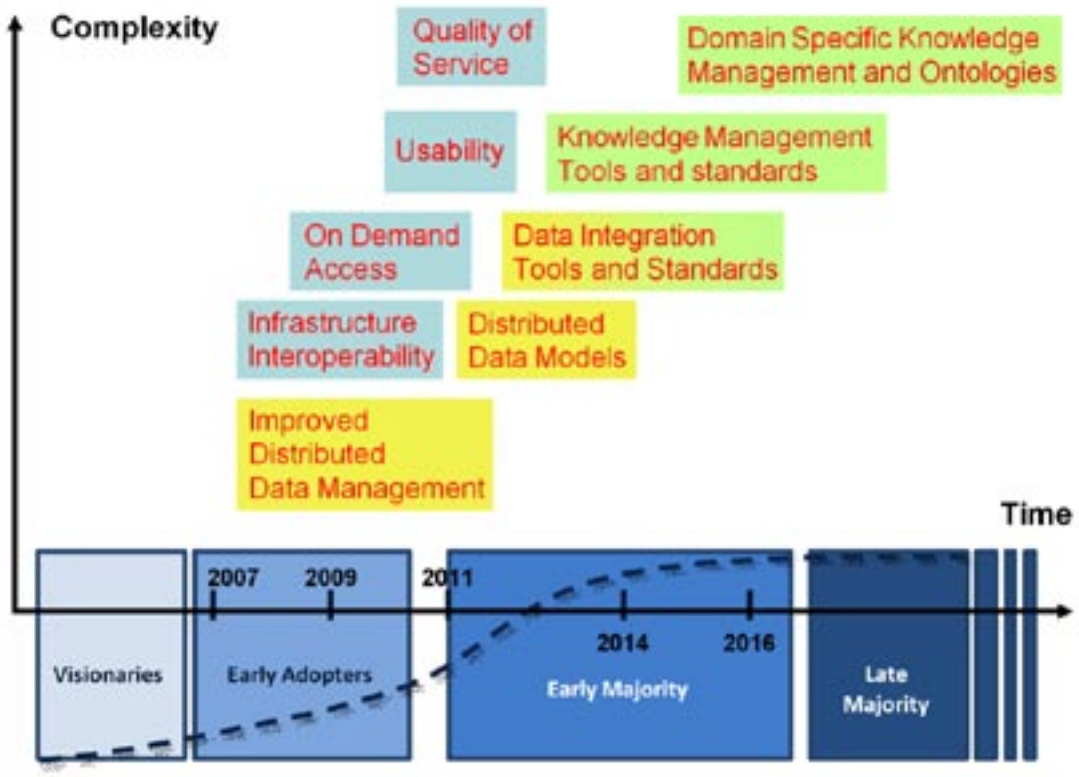


Figure 2: research challenges as a function of time and complexity

This model arises in the field of innovation studies, and distinguishes between: visionaries, pioneers both in research and applications; early adopters, who recognise the potential for rapid benefits and take up the technology quickly, often introducing further innovation; early majority who incur relatively little risk in adopting the technology; and late majority, those who take virtually no risk and finally adopt a technology because there is no other option.

The diagram depicts priorities: even for early adopters, infrastructure interoperability and distributed data management are already necessary; on demand access, usability and quality of service are at the first point of inflection, before rapid expansion, while with sophisticated AI tools in the later stages, a second inflection occurs and the technologies become routinely accepted.

6.2. Mapping else requirements:

- Liability Issues. The current state of EU legislation does not cover liability issues that are specific to healthgrids. The following tasks could help minimise liability concerns for healthgrid usage. Effort need to take place at the organisational level to incorporate strategies and guidelines in order to minimise liability risks.
- Trust and Acceptance. Pilot projects and prototype applications are essential in order to demonstrate the usability of grid services in clinical and research workflows
- Data Protection. Adopting evolving privacy enhancing technologies in order to insure the enforcement of data protection regulation within data management processes and clinical workflows.
- Sustainability Guarantees. The development and deployment of data grids will benefit from more focused prospective assessments of the socio-economic impact in order to identify existing and potential barriers.
- Education and New Skills Requirements. Training and educational programs to increase users' confidence in the use of healthgrid products and services.
- • Intellectual Property. There is a contradiction between intellectual property rights and the needs of grid technology, which will require that access to databases, knowledge and software is free of rights. Contract law and agreements could be an option to regulate the IP issues related to knowledge integration, ontologies and software reuse.
- Governance and Delegation. Some analysis need to be performed to explore the requirement of data controller for data stored and processed on a European grid environment.
- Policies and Codes of Conduct. Discussions should take place between different healthgrid stakeholders to decide on the importance and benefit of applying for new legislation to address healthgrid related legal and ethical issues.
- Dissemination and Publicity. Dissemination and publicity programmes need to precede the deployment of knowledge grids. This includes workshops, conferences, and magazines to attract the user community and build awareness of healthgrid facilities.

7 Concrete recommendations

7.1. Technical recommendations

It is important that technical research and development be conducted in close collaboration with user communities. At certain stages it must be driven and validated by user groups, although there is always scope for innovators to introduce unforeseen possibilities to users. The research communities involved in the definition of the roadmap expressed their interest and support for the deployment of proto¬types and test cases on existing grid infrastructures. We recommend that these infrastructures and tools continue to be made available to applications requiring computing services and data management.

Indeed, some projects are already using the DEISA and EGEE infrastructures for scientific production in the fields of epidemiology, medical imaging and drug discovery. However, these initiatives come from pioneers and are not sufficient to achieve a wider adoption in these research communities. We recommend that:

- More attention be paid to such initiatives so that they may influence the evolution of the technology to make it better fit the needs of the community;
- Two projects within the framework of the EuroPhysiome initiative be identified that could directly benefit from the computing and data management resources of the EGEE and DEISA infrastructures; these should be deployed in parallel on the two infrastructures in order to investigate interoperability issues and identify bottlenecks.

In terms of encouraging biomedical applications to fully exploit grids, we recommend:

- Linking certain advanced health domains to an e-science infrastructure;
- The adaptation of epidemiology data sources to grid models and grid-enabled gateways to epidemiological data, using medical informatics-related connectors such as HL7, DICOM, ENV13606, or similar.

In the same spirit, in order to foster the uptake of grids in the biomedical research and healthcare communities, we recommend:

- The release of open-source components for medical data interfacing;
- Building a core reference database of validated experimental and clinical research data extracted from the literature in innovative medicine and to explore whether a grid infrastructure could support this activity;
- The creation of disease-specific European imaging networks towards the establishment of standards, validation of imaging biomarkers, development of regional centres of excellence in innovative medicine and exploration of grid infrastructures to support such activity.

We recognise that there are a number of concerns (for example: security and standards) in which problems exist irrespective of the use of grids. It is important to understand the nature of these problems and the extent to which the use of grids complicates them. Results could be concrete implementation recommendations (for example: security improvement) and a suggested list of health applications requiring security which may be able to be deployed on a grid. In the field of standards, we believe that the HealthGrid initiative provides the right framework to coordinate the development of the different standards in collaboration with the OGF and the various medical informatics standardisation bodies. We recommend:

- The active pursuit of standards for the sharing of medical images and electronic health records on the grid within the already existing medical informatics standardisation bodies;

- The active pursuit of ontology matching and development for healthgrids;

We believe that technology transfer between EC projects should receive more prominent and active encouragement. In particular, we recommend:

- The commission implements collaboration measures in the funding mechanism for projects;
- Targeted capacity building so that projects may access grid resources on demand, without previous agreement or request; European grid infrastructures should be freely access¬ible to European projects;
- Porting of one or two biomedical grid applications, already successfully deployed on grid infrastructures, to e-science environments using OGSA-compliant grid toolkits.

Finally, to return to a frequent theme in this analysis, we recommend:

- The encouragement of cross community interaction, in order to build meaningful dialogue between grid developers and health researchers.

7.2. Legal recommendations

Liability in a healthgrid System. Using grids blurs the liability issues in terms of medical practice. A stepwise approach should therefore be taken to develop the liability framework, distributing legal responsibility appropriately across healthgrid users. Such an approach would help to favour the reliance on the system while providing legal certainty for all stakeholders, including patients. Moreover, the European Commission should consider supporting the adoption of EU level guidelines that would identify the various parties involved in delivering healthgrid services and annex services and establish the various liabilities that each party must accept. Such guidelines should be widely disseminated in order to develop users' confidence in the use of healthgrids in general. In particular it should be investigated whether specific guidelines on those specific services could be drafted under the provisions for a code of conduct established in directive 2000/31 on eCommerce.

Product safety. As mentioned in D4.2, in the framework of the European level legislation applicable to product safety, national authorities have been established to monitor product safety and to take appropriate measures to protect consumers. Under these circumstances, an information system has been put in place that imposes collaboration between distributors, producers and the national authorities but also between member states and the European Commission (RAPEX).

At present, this system is not used at all for products used in the composition of grid systems. The European Commission should thus adopt policy tools encouraging the use of the RAPEX system for such products.

Healthgrid as a medical device. The law on medical devices is very unclear with respect to healthgrids. While it may be argued that a healthgrid could fall within the ambit of the current medical devices directive in that it is a software tool that impacts on a medical act, the whole construction of the directive is based upon physical goods (which might have a software component) that are placed on the market for purchase or lease. In this situation, many of the currently available monitoring devices are covered only by general product liability, but not by specific liability provision.

In this framework, special guidelines should be issued in order to clarify the application of medical devices legislation to specific tools used in healthgrids.

Patient consent. In February 2007, the European working party on data protection, established under article 29 of the directive issued a working paper looking at the applicability of data protection legislation to Electronic Health Record (EHR) systems. In its report, the working party noted in particular the limitation of the use of consent to permit the processing of heath data. The working party notes that if processing health data in

an EHR system is the primary way of processing health data in a given health system, then a patient's care may be compromised if he or she opts-out of such a system by not giving his or her consent to the creation of an EHR. Accordingly, consent should not be used as it cannot be said to be truly and freely given.

The remaining provisions setting aside the general prohibition on article 8 of the directive 95/46/CE can also be said to pose some problems – notably the idea that a patient ought to know the full finality of the use of data before his or her data may reasonably be used. But, as noted by the data protection working party there are some problems in using consent as a valid basis for processing data in eHealth applications. Indeed, if the creation of, for example, electronic medical records is a necessary and unavoidable consequence of the medical situation, withholding consent may be to the patient's detriment.

Specified and explicit purposes. According to the data protection directive, data may only be collected for specified and explicit purposes. If healthgrids can be used for risk detection, disease monitoring and preventive care, legal guidelines should be established that clarify the circumstances in which professionals can make further use of personal data related to health in the interests of public health. Such guidelines should allow for secondary uses even where such uses could not have been foreseen at the time of data collection.

Technical and organisational security measures. Efforts should be made to harmonise national standards on the technical and organisational measures of data security. While the data protection directive calls for such standards to be adopted, little has been done at a regulatory level to harmonise guidelines across the EU.

Intellectual property rights. It might be desirable for the commission to develop guidelines for the use of open licensing and open standards, which could address the tension between the intellectual property rights of developers and the needs of the grid technology. Such an open standards software approach could then be a solution to help the development and implementation of healthgrids.

On the other hand, the use of healthgrids in the drug discovery sector raises the issue of the ownership of both methods used to discover the medicines and the results achieved. Indeed, all the grid nodes that contribute resources to compute the docking probabilities could claim some ownership of the results and the designers of the software used in the process would certainly be in position to claim ownership of the method. In this context, one may ask whether it is important to know, say, which grid node was the one to identify a particular candidate molecule.

In this context, it is of essential interest, notably in patents, to determine guidelines that would determine, in case of collaboration in the research, what every actor is entitled to according to his contribution to the system.

Privacy policies and codes of conduct. A directive or code of conduct on privacy and health information infrastructure should be developed within the context of directive 95/46/EC and could take the form of either a dedicated directive or could be an EU-level code of conduct to be approved by the European working party on data protection set up under article 29 of the directive. This could help to solve the problem of data processing legitimacy. In particular, it could provide possible bases of legitimacy other than the data subject's consent. It could also provide the following solutions:

- Appropriate safeguards to allow for the further processing of personal data (and especially of medical data) for substantial public interests (without requiring the data subject's consent) like scientific research. An example of appropriate safeguard would be a first coding by the initial data controller and a second coding by a trusted third party gathering all the data from the data controllers before sending them to the researchers,
- Appropriate safeguards to allow keeping the data for longer periods for scientific use; terms under which identification numbers or other identifiers may be used; terms under which (coded) personal data may be transferred to third countries for scientific research.

7.3. Socio-economic recommendations

Trust and Acceptance. Trust is a very important element in any interaction between the different members of a society. In the market context, trust is crucial for successful business to business collaborations. Similarly, in a healthgrid domain a good collaboration will not be achieved unless a trust relationship exists between the different users and stakeholders. Pilot projects and prototype applications, which are an inherent part of the technology roadmap, need to be future oriented in the sense that the ultimate routine operation users have to be persuaded both of their value and their applicability, i.e. their ability to fit into real clinical or research workflows. This has to be taken seriously from the very beginning, even in proof-of-technology demonstrators: the goal should always be to give users, especially clinicians, tools that they would consider using with patients in real healthcare situations. Trust and acceptance can be greatly enhanced by the establishment of appropriate ethics committee structures to advise on the observance of ethical principles.

Estimation of Costs and Benefits. Ex-ante analyses over time, based on initial pilot experience, have to focus on ensuring acceptance, technical and regulatory certainty, and sufficient private incentives in the steps to follow. An

inherent part of such assessments should be to estimate potential net benefits (i.e. expected benefits less expected costs over time), accounting for different risks and for optimism bias in estimations. Such studies will facilitate access to initial funding, but can also be beneficial in the necessary dissemination work among the health sector.

Sustainability Guarantees. Work towards achieving the next milestone in complexity – data grids – will benefit from more focused prospective assessments of socio-economic impact in order to a) identify already existing, as well as potential barriers, and b) build convincing business cases ensuring sustainability. The analysis of alternative resource allocation options from a societal perspective, but also on organisational level, becomes necessary.

An organisational milestone can be defined here in the move from technology science towards service provision. By that stage, a notable amount of legal and regulatory certainty has to be achieved, so that private incentives can be assessed and adjusted (including via government intervention) if necessary.

Cross-Organisational Interoperability. The effective deployment of knowledge grids will crucially depend on collaboration between institutions, meaning more than "simple" access to each others' data and computing resources. This collaboration requires the utilisation of human resources and in some cases a significant strategic re-orientation and re-organisation of working processes and even management structures. As the health sector, including clinical research and public health, is (and should be) highly regulated, policy makers on regional, national, and EU level should review the existing regulatory framework against the requirements arising from the exploitation of knowledge grids. Particular attention should be given to flexibility of government regulated budgets and reimbursement schemes. The latter should encourage cross-organisational collaboration, including such beyond national borders, by means of using knowledge grids.

7.4. Further reading

A more detailed account of the proposed road map (deliverable D6.2) for healthgrids is at:
http://eu-share.org/about-share/deliverables-and-documents.html

where the full set of public SHARE documents is also to be found:

D3.1	Healthgrids Framework
D3.2	Baseline on technological and security aspects of healthgrids
D3.3	Bottlenecks & challenges and RTD responses for technological and security aspects of healthgrids roadmap
D3.4	Technology component roadmap II
D4.1	Baseline on legal, ethical, social, and economic aspects of healthgrids
D4.2	Bottlenecks & challenges and RTD responses for legal, ethical, social & economic aspects of healthgrids roadmap I
D4.3	Legal, social & economic challenges component roadmap II
D5.1a	Use case scenario for epidemiology
D5.1b	Use case scenario for innovative medicine
D5.2a	Epidemiology roadmap
D5.2b	Innovative medicine roadmap
D6.1	Integrated roadmap I
D6.2	Integrated roadmap II

Terminology

Abbreviation List

DICOM	The Digital Imaging and Communications in Medicine Standard
EC	European Commission
EHR	Electronic Health Record
ELSE	Ethical, Legal and Socio-Economic
EPAL	Enterprise Privacy Authorisation Language
EU	European Union
HL7	The Health Level 7 Standard
HPC	High Performance Computing
IPR	Intellectual Property Rights
OGSA	Open Grid Services Architecture
QoS	Quality of Service
SOKU	Service Oriented Knowledge Utility
VPH	Virtual Physiological Human
W3C	World Wide Web Consortium
WPx	Work Package x
WP3	SHARE Technology and Security Activity
WP4	SHARE Health Policy, Legal, Social and Economics Activity
WP5	SHARE Applications Activity
WP6	SHARE Roadmap Synthesis and Validation Activity

Definitions

Authentication: Verifying and confirming the identity of a grid user.

Authorisation: Restricting access to resources based on what a user has been granted access to.

Data: Any and all complex data entities from observations, experiments, simulations, models, and higher order assemblies, along with the associated documentation needed to describe and interpret them.

Data controller: The person or organisation responsible for the manner in which any personal data is processed.

Data mining: Automatically searching large volumes of data for patterns or associations.

Data model: A model that describes in an abstract way how data is represented in an information system. A data model can be a part of ontology, which is a description of how data is represented in an entire domain.

Data processor: Any person who processes data on behalf of a data controller.

Data subject: An individual who is the subject of personal data.

Grid: A fully distributed, dynamically reconfigurable, scalable and autonomous infrastructure to provide location independent, pervasive, reliable, secure and efficient access to a coordinated set of services encapsulating and virtualising resources.

Informed consent: A legal term referring to a situation where a person can be said to have given their consent based upon an appreciation and understanding of the facts and implications of an action.

Metadata: Data about data; may be regarded as a subset of data. Metadata summarise data content, context, structure, inter-relationships, and provenance (information on history and origins). They add relevance and purpose to data, and enable the identification of similar data in different data collections.

Middleware: A software stack composed of security, resource management, data access, accounting, and other services for applications, users and resource providers to operate effectively in a grid environment.

Ontology: The systematic description of a given phenomenon, which often includes a controlled vocabulary and relationships, captures nuances in meaning and enables knowledge sharing and reuse. Typically, ontology defines data entities, data attributes, relations and possible functions and operations.

Processing: Obtaining, recording or holding the data, or carrying out any operation on the data, including organising, adapting or altering it. Retrieval, consultation or use of the data, disclosure of the data, and alignment, combination, blocking, erasure or destruction of the data are all legally classed as processing.

SOAP: A protocol for exchanging XML messages over a network. It defines the structure of the XML messages (the SOAP envelope), and a framework that defines how these messages should be processed by software.

The Article 29 Data Protection Working Party: A working party established by article 29 of directive 95/46/EC. It is the independent EU advisory body on data protection and privacy. Its tasks are laid down in articles 30 of directive 95/46/EC and 14 of directive 97/66/EC.

Virtual Organisation: A group of grid users with similar interests and requirements working collaboratively and/or sharing resources regardless of location.

Web Service: A software system designed to allow inter-computer interaction over a network to perform a task. Processes interact with a web service, in a manner prescribed by its interface, using messages which are enclosed in a SOAP envelope. Software applications can use web services to exchange data over a network.

Workflow: A set of components and relations between them, used to define a complex process from simple building blocks. Relations may be in the form of data links which allow the output of one component to be used as the input of another, or control links which state some conditions on the execution of a component.

XML: An annotation technology used to describe structured data within a document using mark-ups and tags, similar to HTML. The main difference between the two is that the elements in XML can be given a definition depending on their usage which may be semantic rather than presentational. XML is a text format and can be read easily either by humans or machines.

Notes

Notes

Additional information:

Document ID:	SHARE Compact Road Map
Date:	2008/08/14
Authors:	All Partners
Activity:	WP6: Integrated Road Map
Document status:	PUBLIC
Document link:	http://roadmap.healthgrid.org
Keywords:	Healthgrid; Biomedical Informatics; Grid Computing – Research Challenges; Innovation and Technology Management; Ethical, Legal, Social and Economic Issues (ELSE); Epidemiology; Innovative Medicine; Genomic & Individualised Medicine; Security; Regulatory Compliance; Collaboration

KK-30-08-624-EN-C

SHARE the journey

A European Healthgrid Roadmap

For further information, please contact:

European Commission
Information Society and Media Directorate-General
Directorate H - ICT adressing Societal Challenges
Unit H1 - ICT for Health
Phone: +32 2 296 41 94
Fax: +32 2 296 01 81
eHealth@ec.europa.eu
http://ec.europa.eu/information_society/ehealth

ISBN 978-92-79-09668-6

Design & Printed by OIB

European Commission
Information Society and Media