



SHARE

INNOVATIVE MEDICINES ROADMAP

Document ID:	SHARE-D5.2b_v4.0
Date:	15/06/07
Authors:	V. Breton, A. Dobrev, N. Jacq, T. Solomonides
Activity:	WP5: Applications
Document status:	Final
Document link:	http://eu-egee.org/deliverables.html
Confidentiality:	Public
Keywords:	Innovative medicines, Roadmap, Health grid

Abstract: This document presents an analysis of current technologies in grids and the ethical, legal, social and economic (ELSE) framework of grid technologies for their adoption in innovative medicines. Based on this analysis, the document proposes milestones to be achieved for the uptake of health grids in this application domain. The proposed roadmap for health grid adoption is compared to the technology platform of the strategic research agenda of the Innovative Medicines Initiative. As a result, it is recommended that the health grid roadmap explicitly addresses the development of enhanced knowledge representation models and data exchange standards for complex systems, as well as the definition of standards for the federation of data bases. We also suggest that some of the money to be spent on the Innovative Medicines Initiative Knowledge Management effort should be spent on health grid developments.

Document Log

Issue	Date	Comment	Author
0	02/05/07	First version	Nicolas Jacq
0.1	10/05/07	Second version	Vincent Breton
1.0	28/05/07	First release for comments	Vincent Breton
1.1	28/05/07	Comments	Nicolas Jacq
1.2	31/05/07	Contribution to chapter 7 and comments on the document	Alexander Dobrev
2.0	01/06/07	Second release for review	Nicolas Jacq
2.1	06/06/07	Changes and minor corrections throughout the document; contributions to Chapter 7.	Tony Solomonides
2.1.1	12/06/07	Minor corrections and critical comments throughout the document	Jörg Artmann, Veli Stroetmann
3.0	14/6/07	Revision including the different comments	Vincent Breton
3.1	18/6/07	Revision including Howard Bilofsky comments	Vincent Breton
3.2	05/07/07	Revision especially of Chapter 7 following Howard Bilofsky's comments	Tony Solomonides
4.0	15/07/07	Final version	Nicolas Jacq

Document Change Record

Issue	Item	Reason for Change
0	General structure and general items	Preparing first release
1	Complete version	First release
2	Improved version and chapter 7 completed	Second release
3	Improved version including comments from T. Solomonidès, V. Stroetman and Jörg Artmann	Third release
4	Final version	Fourth release



CONTENT

1. INTRODUCTION	4
1.1. PURPOSE OF THE DOCUMENT	4
1.2. APPLICATION AREA.....	4
1.3. REFERENCES	4
1.4. TERMINOLOGY.....	5
1.5. SOURCES	6
1.6. ACKNOWLEDGEMENTS	6
2. EXECUTIVE SUMMARY	7
3. HEALTHGRID VISION	8
4. BACKGROUND	12
5. USE CASE	16
6. ASSESMENT OF GRID TECHNOLOGY FOR <i>IN SILICO</i> DRUG DISCOVERY	19
6.1. ASSESMENT IN TERMS OF DATA INTEGRATION	19
6.2. ASSESMENT IN TERMS OF WORKFLOW ENACTMENT	19
6.3. ASSESMENT IN TERMS OF ACCESS TO RESOURCES	20
6.4. ASSESMENT IN TERMS OF COLLABORATION BETWEEN PUBLIC AND PRIVATE PARTNERS	20
7. ASSESMENT OF ELSE ISSUES	22
7.1. ETHICAL ISSUES	22
7.2. LEGAL ISSUES	23
7.3. SOCIAL ISSUES	23
7.4. ECONOMIC ISSUES	24
8. ROADMAP	26
8.1. RESEARCH TOPICS	27
8.2. DEPLOYMENT ACTIONS	28
8.3. TIME PLANNING.....	29
9. CONCLUSION	30

1. INTRODUCTION

1.1. PURPOSE OF THE DOCUMENT

The purpose of the document is to analyse the requirements, needs and aims of the innovative medicine use case defined in the deliverable D5.1b with respect to the technological and ELSE issues identified in D3.3 and D4.2. Those documents presented the bottlenecks and challenges for Technology, Ethical, Legal, Social and Economic Issues in the adoption of Grids in health. This document will review the issues raised in those documents and analyse how they meet the requirements expressed in the use case scenario. From this analysis, the evolved versions of the technology and security roadmap (D3.4) and the ELSE roadmap (D4.2) will be elaborated. This document will also be used for the integrated roadmap (D6.2).

This document has a corresponding version for Epidemiology identified under the code D5.2a. This document is self-contained.

This document is intended for both users (government agencies, pharmaceutical companies, researchers in academia or industry) and developers (technologists and legal experts). It tries to outline the milestones that must be covered for the adoption of health grids. The level of technical detail is presented in two steps, trying to attract the interest of policy makers, but providing the technical details that are necessary for the implementation.

All deliverables can be downloaded from the SHARE web-site at www.eu-share.org.

1.2. APPLICATION AREA

The document is intended for internal and external use. It will be used as a dissemination tool for the SHARE project, and for consultation in the process of elaborating the second versions of the deliverable and the integrated roadmap.

1.3. REFERENCES

- [1] L-M. Birkholtz, O. Bastien, G. Wells, D. Grando, F. Joubert, V. Kasam, M. Zimmermann, P. Ortet, N. Jacq, N. Saidani, S. Hofmann-Apitius, M. Hofmann-Apitius, V. Breton, A.I Louw and E. Marechal, Integration and mining of malaria molecular, functional and pharmacological data: how far are we from a chemogenomic knowledge space?, *Malaria Journal*, Vol. 5, (2006) 110.
- [2] N. Jacq, <http://members.healthgrid.org/njacq/>
- [3] V. Breton, K. Dean and T. Solomonides, editors on behalf of the Healthgrid White Paper collaboration, "The Healthgrid White Paper", Proceedings of Healthgrid conference, IOS Press, Vol 112, 2005
- [4] Doman, T.N., et al., Molecular docking and high throughput screening for novel inhibitors of protein tyrosine phosphatase 1B. *J. Med. Chem.* 45 2213–2221 (2002).
- [5] The Innovative Medicines Initiative (IMI) Strategic Research Agenda. Creating Biomedical R&D Leadership for Europe to Benefit Patients and Society, 15 September 2006 available at, http://www.efpia.org/4_pos/SRA.pdf, <http://www.imi-europe.org/DocStorage/PublicSiteAdmin/Publications/Innovative%20Medicines%20Ini>

- tiative%20SRA%20Version%202.0.pdf, separate chapters also available at:
<http://www.imi-europe.org/Publications.aspx?viewCategory=Researchx20Agenda>
- [6] The Dengue docking project: <http://www.biozentrum.unibas.ch/grid/projects.html>
- [7] The Swiss Bio Grid initiative: <http://www.swissbiogrid.org/>
- [8] A. Dobrev, The economic and business cases for HealthGrid solutions, contribution to eHealth Week, Berlin, April 2007
- [9] Boston Consulting Group, BCG Estimate – A Revolution in R&D – The Impact of Genomics, (2001).

1.4. TERMINOLOGY

Glossary

Term	Definition
Innovative Medicine	IM means either a new process i.e, an innovative process for producing drugs faster, cheaper and better, or it means the development of new drugs i.e, innovative drugs with greater efficacy, reduced side-effects or other unique features.
QSAR	Quantitative Structure Activity Relationship is a parametric representation of the correlation of chemical structure with chemical behaviour using well-defined methods.
Drug Discovery	Drug discovery is the long term, multi-stage and costly process by which drugs are discovered (i.e. identified from among a multitude of candidates) and/or designed.
Screening	Measurement of the chemical activity of a set of compounds on a given target molecule
Target	An enzyme or a protein which is believed to be associated with a desired change in the behaviour of disease processes and on which drugs usually act
<i>In silico</i>	Computer-based simulation of a process.
Docking	A research technique for predicting whether and how one molecule, usually small, will bind to another, usually a protein
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 virtualizing resources (computing power, storage, instruments, data, etc.) in order to provide a collaboration medium, to facilitate effective computation and to generate knowledge.
Health grid	A grid infrastructure with services dedicated to healthcare and biomedical research.
HealthGrid	An organization originally established in Europe, but now with world-wide associates, whose aim is to promote the development and adoption of health grids.

EFPIA	European Federation of Pharmaceutical Industries and Associations (http://www.efpia.org/)
IMISRA (or SRA)	Innovative Medicines Initiative Strategic Research Agenda (http://www.imi-europe.org/)

1.5. SOURCES

This deliverable contrasts the first draft HealthGrid roadmap described in SHARE deliverables D3.3, D4.2 and D6.1 to the requirements of biopharmaceutical research and development. A use case scenario has been described in SHARE deliverable D5.1b in the specific area of *in silico* drug discovery. This document extends its considerations beyond this use case and analyses how a grid infrastructure could contribute to revitalizing the biopharmaceutical research and development (R&D) environment in Europe.

At the request of the European Commission, the European Federation of Pharmaceutical Industries and Associations (EFPIA) has identified the main barriers to innovation in Life Sciences research in Europe with the objective of establishing a European technology platform for innovative medicines. A document was produced by all relevant stakeholders in July 2005 describing the Strategic Research Agenda for the Innovative Medicines Initiative [5]. This document has been used as the basis for our analysis of the present bottlenecks in the biomedical R&D process as well as the recommendations on how to address these bottlenecks.

We also collected feedback from key users. We interviewed researchers from academic laboratories but also from pharmaceutical laboratories as well as service providers to biotech and pharma SMEs.

1.6. ACKNOWLEDGEMENTS

We gratefully acknowledge the work of the internal reviewers who made very relevant suggestions to improve the document content: Jörg Artman, Howard Bilofsky and Veli Stroetmann.



2. EXECUTIVE SUMMARY

The innovative medicine use case has been selected because of its relevance to healthcare and biomedical research. This document analyses the roadmaps proposed for health grid from technical and ethical, legal, social and economic (ELSE) perspectives in view of the Strategic Research Agenda (SRA) of the Innovative Medicines Initiative (IMI) and more specifically a specific use case relevant to drug discovery.

The discovery and development of new drugs is very costly and attrition rates are high. Initiatives to reduce the rate of attrition during later phases of development are clearly desirable and if successfully implemented will reduce costs. EFPIA's Research Directors Group has identified four key areas, addressing the key bottlenecks in the R&D process. The knowledge management area is identified as being key to leveraging the potential of new technologies such as genomics and proteomics and to analyse the huge amount of information in an integrated way. The *in silico* drug discovery use case described in deliverable D5.1b shows how grids can start to be used today for biopharmaceutical research and development but proper management of the knowledge generated is needed to really bring added value to the drug development process.

The first outlines of the technical and ELSE parts of the health grid roadmap developed in deliverable D3.3 and D4.2 identify useful and necessary steps toward building the flexible, secure and scalable IT infrastructure required for biopharmaceutical research and development. However, on the technical side, further, specific steps are needed to enable sophisticated knowledge management. In the biopharmaceutical R&D process, an enormous quantity and diversity of data is created. A critical factor is the ability to turn these masses of siloed information into actionable knowledge. The need to develop enhanced knowledge representation models and data exchange standards for complex systems, as well as to design standards for the federation of data bases must be properly addressed in the roadmap.

3. HEALTHGRID VISION

The vision is to create an environment where information at all levels of the biological hierarchy (molecule, cell, tissue, individual, population) can be associated to provide individualized healthcare.

Most healthcare systems in the developed world are facing multiple challenges in their attempt to maintain an acceptable level of care for their citizens. The principal challenges are often experienced and expressed in economic terms, e.g. as issues of

- total cost,
- capacity and responsiveness, and
- allocation of limited resources.

Underlying these economic constraints is the moral challenge of priorities, as governments seek to balance the demands of

- changing demographics, with an ageing population both surviving and remaining active longer, yet increasingly suffering from chronic conditions;
- increasingly effective treatments for both acute and chronic conditions;
- sophisticated – and sometimes unproven – novel treatments for conditions which not very long ago were considered untreatable;
- expectations of ever increasing opportunities and quality of care;
- Equity or equal access to adequate healthcare by disabled citizens.

In an attempt to meet these demands, health systems have increasingly looked to information technology to help, among other things, to optimize the distribution and use of resources, to reduce queues and waiting times, to record and so avoid errors, and to provide modern treatments into underserved or remote communities.

Beyond these essentially resource-oriented uses, information technology is also seen as an essential ingredient in a change in medicine itself. In the course of the last two decades, the practice of medicine and healthcare provision in general have moved away from reliance on the doctor's personal knowledge and skill to requiring a scientific basis for diagnosis and treatment, in what has come to be known as 'evidence based practice'. The evidence a doctor or nurse must now take into account is

- published medical knowledge;
- their knowledge of the patient; and
- practical knowledge of what is available by way of procedures, protocols, and so on, in their environment.

In this context also, governments have initiated programmes to create information-driven healthcare systems.

However, these modernization processes face a number of challenges:

- creating and populating, connecting and understanding patient records across organization boundaries and, in due course, across different national health systems;
- increasing the openness and accessibility of systems - e.g. providing patients with ownership of their healthcare record – while
- ensuring privacy, confidentiality and ethical compliance in the socio-legal plane, and
- developing and maintaining data integrity, security and authenticity (e.g. provenance and semantics) in the technical plane;
- developing and maintaining interoperability between systems – technical as well as semantic and cultural;
- providing appropriate levels of authorization and authentication of users across all the services and the citizen;
- discovering, grading and certifying trustworthy sources of knowledge and information to guide future action; finally,
- winning the trust and commitment of the medical professions and the public at a time of immense change and economic pressure.

One immediate limitation is in the application of traditional information networks and technology in healthcare. Governments have naturally focussed on the technical issues that are reasonably well understood, even if solutions are not always easy to obtain: robustness of networks, scalability of systems, readiness to handle a very large volume of data. However, in many respects, these reproduce in the technology some of the problems of the traditional paper-driven systems: inflexibility, maldistribution of resources, failure to understand the needs of medical practitioners, failure to support effective collaboration, and an ultimately simplistic equation of quality with ‘choice’, while minimal provision is made for just-over-the-horizon future technologies such as genomic medicine and individualized prescribing.

In the face of these challenges, a computational innovation, ‘grid’ technology, or *the grid*, has become available to clinicians in the last few years, first as a research tool and then, in the not-too-distant future, as a serious healthcare infrastructure. The grid is not one technology but many, and the use of the singular is somewhat misleading, but it is convenient inasmuch as it echoes ‘the Internet’ to which it is closely related and in fact upon which it is built. Just as the Internet, or more precisely the World Wide Web, has provided a massive information platform whose exploitation is limited only by economics (and, in some cases, politics) grid technology promises to scale this up to the provision of unprecedented computational power, online storage and collaboration opportunities. The informatic grid approaches the provision of computational, information and communication services through resource sharing in a seamless and transparent manner, much as the electricity ‘grid’ provides power to any device plugged into it, irrespective of its purpose or design. Grid computing aims at the provision of a global ICT infrastructure that will enable a coordinated, flexible and secure sharing of diverse resources, including computers, applications, data, storage, networks, and scientific



instruments across dynamic and geographically dispersed organizations and communities (sometimes known as ‘virtual organizations’ or ‘VOs’). Grid technologies promise to change the way organizations tackle complex problems by offering unprecedented opportunities for resource sharing and collaboration. Just as the World Wide Web transformed the way we exchange information, the grid concept takes parallel and distributed computing to the next level, providing a unified, resilient, and transparent infrastructure, available on demand, in order to solve increasingly complex problems.

A ‘health grid’ is an innovative use of this emerging information technology to support broad access to rapid, cost-effective and high quality healthcare. In particular, the areas of healthcare provision and research that can be beneficially affected by health grid technology include:

- medical imaging and image processing;
- modeling the human body for therapy planning;
- simulation of disease-related processes;
- pharmaceutical research and development;
- epidemiological studies;
- genomic research and treatment development; and
- biosurveillance

In all these areas, grid technology can either significantly reduce the cost or time to produce results and evidence, or even provide resources that are able to deliver services that cannot be economically delivered using conventionally networked information systems. Moreover, the emergence of this technology enables interdisciplinary research at the crossroads of medical informatics, bioinformatics and system biology to impact healthcare.

A growing number of grid applications are under development, with several completed and deployed in life sciences and medical research. Within the European Union and its member states, many applications have benefited and still benefit from substantial funding from the European Commission and some individual state funding bodies. Among the current projects, those relevant to health can be roughly classified into three categories:

- Infrastructure projects that aim to offer a stable distributed environment for the conduct of collaborative scientific research; the currently deployed infrastructures in Europe (DEISA, EGEE) offer a generic multidisciplinary environment where biomedical applications can be deployed.
- Technology projects aimed at developing new grid-enabled services and environments relevant to the needs of life sciences and healthcare.
- End-user projects that focus on specific life science or healthcare issues and integrate grid technologies wherever they appear relevant.



Born from discussions between grid application developers and medical computer scientists, the concept of 'health grid' is now over three years old. The HealthGrid community gathers individuals from the public and private domain world-wide who are actively exploring the beneficial impact of health grid technology on healthcare provision and research. The annual HealthGrid conferences are an opportunity to evaluate the growing usage of grids for life science and medical research. Adoption of grids for healthcare is expected to follow their adoption in the life sciences and medical research, provided the legal and ethical framework of member states allows their deployment.

In the next chapter, we provide some background information on pharmaceutical research and development before focussing on the innovative medicine use case described in SHARE Deliverable D5.1b.

4. BACKGROUND

This document focuses on the innovative medicine use case described in deliverable D5.1b, which has been selected due to its relevance to healthcare and biomedical research. However, it is worth considering more globally the biopharmaceutical research and development (R&D) environment in Europe which is described in the Strategic Research Agenda for the Innovative Medicines Initiative [5].

Among other issues, the discovery and development of new drugs is very costly and attrition rates are high. Initiatives to reduce the rate of attrition during later phases of development are clearly desirable and if successfully implemented will reduce costs.

EFPIA's Research Directors Group has identified pre-competitive barriers to innovation. The objective for the future would be to identify as soon as possible in the pre-clinical phase:

- Reasons for lack of efficacy, despite promising pre-clinical data.
- The potential for adverse drug reactions and pre-clinical toxicity.

The identified key bottlenecks in the R&D process are the following:

- 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.

In these areas, scientific and technological advances would be of direct benefit to the pharmaceutical industry by improving efficacy of tests and containing costs. In addition, a more efficient R&D process will bring more efficacious and safer drugs to the market, resulting in a direct benefit for the patients. In certain fields, such as reproductive medicine, attention has already begun to be focussed on the individual patient; it is anticipated that this will be a general trend in the relatively near future. To accelerate the development of more effective medicines, safety and efficacy evaluation of new molecular entities needs to be improved. In order to achieve these goals, the proposed Strategic Research Agenda is organised around four key areas, addressing the key bottlenecks in the R&D process. These key areas are:

- Safety, addressing the bottlenecks of predictivity in safety evaluation and pharmacovigilance with the authorities,
- Efficacy, addressing the bottlenecks of predictive pharmacology, biomarkers identification and validation, patient recruitment and risk assessment with the authorities,
- Knowledge Management, leveraging the potential of new technologies to analyse a large quantity and diversity of information in an integrative and predictive way.

- Education and Training, addressing certain gaps in expertise which need to be resolved in order to change and support the biopharmaceutical research and development process.¹

The expected long term benefits to this approach are the following:

- Reduce development times and costs. By bringing a drug to market faster, lives can be saved and suffering reduced.
- Discover and develop better medicines, which will be safer, have a better efficacy and will be better adapted to patients needs.
- Facilitate risk / benefit evaluation by the authorities to accelerate access of innovative medicines to the patients.

The EFPIA report stresses also the need to achieve the active participation of all relevant stakeholders (academia, clinicians, patient organisations, large industry, SMEs, regulatory and ethics specialists) in the development of new drugs. The collective impact is expected to come from the transparent, total-systems approach to the discovery and development process and in so doing enables each player to appreciate more fully the roles and needs of the others.

In relation to drug safety, the report identifies very important research needs in the development of *in silico* methods, in order to:

- Improve predictivity for endpoints characterized in late non-clinical safety studies (e.g. chronic target organ toxicity; reproduction toxicity),
- Provide tools to screen and select the best chemical lead at the discovery stage,
- Avoid specific structural (and activity) characteristics linked to safety issues,
- Judge “toxicodevelopability” in very early development,
- Help to tailor a specific toxicity testing program.

¹ The ‘four pillars’ are stated in the Executive Summary of the SRA to be:

- Predictivity of Safety Evaluation (Pillar I): Nine recommendations are presented. These include the creation of a European Centre of Drug Safety Research, and establishing a framework to develop biomarkers that will have human relevance and regulatory utility;
- Predictivity of Efficacy Evaluation (Pillar II): Five recommendations are presented related to each of the five disease areas that have been identified as priorities for Europe, based on unmet medical need. These recommendations include creating disease-specific European Imaging Networks, developing regional centres of excellence, creating disease-specific European centres for the validation of new biomarkers and enhancing collaborations with patients and regulatory authorities;
- Knowledge Management (Pillar III): Fifteen recommendations are presented. These include establishing a Translational Knowledge Management team to support Pillar I and Pillar II projects, and creating a Knowledge Management Platform to develop effective data integration and analysis tools;
- Education and Training (Pillar IV): Five recommendations are presented that include establishing a European Medicines Research Academy, and the implementation of multi-disciplinary programmes to develop skills in integrating biology and medicine expertise.

In terms of drug efficacy, the following needs are common to all disease areas:

- Develop better understanding of disease mechanisms,
- Develop in vitro and in vivo models predictive of clinical efficacy,
- Develop in silico simulations of disease pathology,
- Stimulate translational medicine in an integrated fashion across industry and academia,
- Create disease-specific European Imaging Networks for establishment of standards, validation of imaging biomarkers and development of regional centres of excellence,
- Create disease-specific European Centres for validation of “omics-based biomarkers”,
- Co-ordinate the development of national patient networks and data bases to develop a true pan-European organisation for patient selection and clinical trial analysis,
- Form a European stakeholder consortium to address value demonstration, including quality of life issues, patient reported outcomes and burden of disease,
- Develop a partnership with regulators to devise innovative clinical trial designs and analyses, to aid acceptance of biomarkers and to promote data sharing and joint consideration of ethical issues.

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.

The report identifies two levels of knowledge management that need to be addressed:

- The capture, analysis and interpretation of knowledge generated regarding the physiology and pathophysiology related to disease stage or toxicological targets. Here the aim is to improve the understanding of the underlying process including the impact of pharmacogenomics in order to predict successfully the validity of a drug target and risk management for patient populations
- 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. The aim here is to integrate all available knowledge at any given stage of the development process in order to make the best predictions possible for the chances of success of this molecule in the next stage. The know-how for an integrated model-based Drug Development tool is available in Europe but one of the major bottlenecks is the lack of availability of databases across R&D that allow easy access for data integration.

The scientific requirements are addressed by the following actions:

- Develop a strategy to identify the areas of interest to all stakeholders,
- Provide mechanisms for data federation across heterogeneous data sources,
- Provide a flexible and secure collaborative environment serving all stakeholders,
- Provide standards and mechanisms for consistent data integration and data sharing,
- Provide standards and mechanisms for consistent integration of complex scientific tools and computational models,



- Insure interoperability of computing services across organizations,
- Develop broad and generic research projects for bridging gaps in current technologies.

The main recommendations coming out of the report are:

- **Develop enhanced knowledge representation models and data exchange standards for complex systems,**
- **Build a core reference database of validated experimental data extracted from the literature,**
- **Design standards for and build an expert tool to allow the federation of local databases in a secured environment.**

This background information is now going to be used to further analyse the potential impact of a health grid infrastructure on the following use case scenario.

5. USE CASE

The use case described here is documented in detail in SHARE deliverable D5.1b. It focuses on *in silico* drug discovery.

Drug discovery is the long term, multi-stage and high cost process by which drugs are discovered and/or designed. [2]. Drug candidates are intermediate products of the drug development process. Drug Development manages preclinical safety studies and clinical phases. Registration and Delivery are the last steps of the full process. Reducing the research time in the discovery stage and having enhanced information about the leads are key priorities for pharmaceutical companies worldwide. Collaborations with academic laboratories and small biotechnology or pharmaceutical companies are crucial, mainly in exploratory research, then in the lead discovery stage and progressively less during the drug development phases.

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.

Recent progress in genomics, transcriptomics, proteomics, high throughput screening, combinatorial chemistry, molecular biology and pharmacogenomics has radically changed the traditional physiology-based approach to drug discovery where the organism is seen as a black box [10]. The approach is now to understand how disease is controlled at the molecular and physiological level and to target specific entities based on this knowledge.

In silico drug discovery is one of the most promising strategies to speed up the Drug Discovery process. It is important to understand and control the *in silico* process; this is described below.

Figure 1 shows the different phases of the drug discovery process with their approximate duration, their success rate and the corresponding *in silico* contributions.

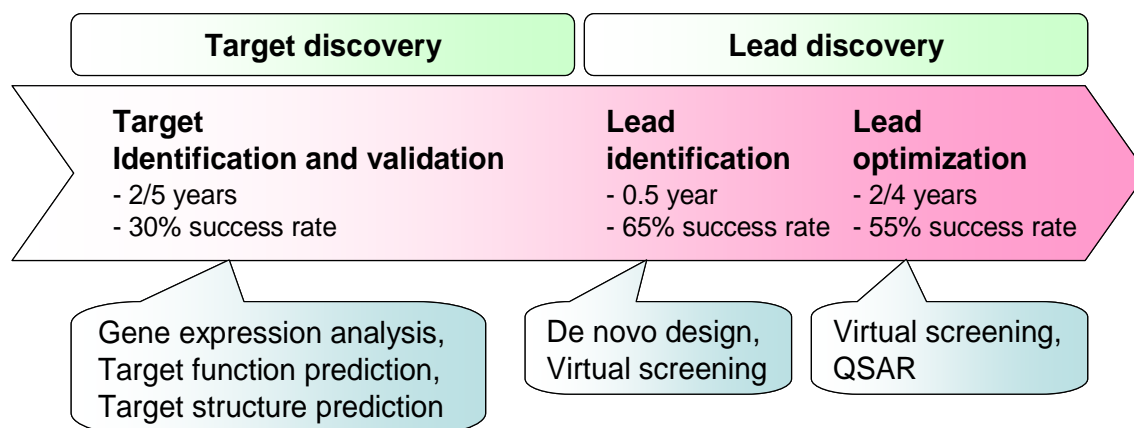


Figure 1: Representation of the different phases of the drug discovery process with their duration, their success rate and the corresponding potential *in silico* contributions.

A target is a cellular molecule which is believed to be associated with a desired change in the behaviour of a disease process and on which drugs usually act. Target identification and validation aims to isolate and select it. *In silico* drug discovery contributes to the target

discovery by analysis of the gene expression data, target function prediction and target three-dimensional (3D) structure prediction.

A lead compound is a substance affecting the selected target. Two different *in silico* pipelines can be used for identifying it: the *de novo* design and virtual screening. *De novo* design builds iteratively a compound from the structure of a protein active site. Virtual screening selects *in silico* the best compound from a molecule database. These methods speed up the process and reduce costs avoiding time consuming and costly *in vitro* tests.

Lead optimization addresses the development from the most promising lead compounds to a safe and effective drug. Instead of expensive and longer *in vitro* and *in vivo* tests, evaluation of the basic chemical properties can be achieved by virtual screening and using Quantitative Structure Activity Relationship (QSAR). QSAR can be used in a quantitative process correlating chemical structure with function in order to optimize pharmaceutical properties (Absorption, Distribution, Metabolism, Excretion and Toxicity) or efficacy against the target organism.

In silico drug discovery contributes to increasing biological system knowledge, to managing data in a collaboration space, to speeding up analysis and consequently improving the success rate compared with the traditional “wet” approach. The efficiency gains of such an integrated knowledge system could result in 35% cost savings, or about US\$300 million, and 15% time reduction, or two years of development time per drug [10].

Reducing the research time and cost in the discovery stage and enhancing information about the leads are key priorities for pharmaceutical companies worldwide. To achieve this goal, *in silico* drug discovery must meet the following requirements:

- **Data integration.** The *in silico* drug discovery process includes the management of a large variety and quantity of scientific data. For example: images, sequences, models, databases. Data integration is thus a challenge to increase knowledge discovery but also to ease the complex workflow. This implies data format standardization, dataflow definition in a distributed system, infrastructure and software providers for data storage, services for data and meta-data registration, data manipulation and database updates, development and sharing of ontologies and knowledge representations. .
- **Workflow enactment.** The *in silico* drug discovery process also includes the management of a large variety and quantity of software. Software integration is another challenge to build efficient and complex workflows and to ease data management and data mining. Experts in different areas are absolutely necessary to maintain and update software and workflows, to propose new methods or pipelines, to use remote services, exploit outputs, and finally to propose compounds for assay. A software workflow will assist the scientist and the decision-maker in organizing their work in a flexible manner, and in delivering the information and knowledge to the organization.
- **Access to computing and data resources.** Deploying intensive computing is a challenge for *in silico* drug discovery. For instance, computing 1 million docking probabilities or modelling 1,000 compounds on one target protein requires on the order of a few TFlops for one day. Extensive computing resources are also needed to accurately describe protein structure models by computational methods based on all-atom physics-based force fields including implicit solution. Computing power is also required for bioinformatics resource



centres where server access is saturated by the large number of short tasks requested by users.

- **Collaboration between public and private partners.** Joining the new Information Technologies with life science to enable *in silico* drug discovery requires strong remote collaboration between different public and private experts when addressing neglected and emerging infectious diseases. It also involves concrete sharing of resources: data and knowledge, software and workflow, and infrastructures such as computing, storage and networks. The collaboration space needs experts to maintain the resources. Having tools and data accessible to everyone in collaboration requires intuitive interfaces that need to evolve and be maintained. These interfaces reduce the development time of new methods. They also help the integration of data and software from *in silico* drug discovery but also from experimental processes. Of course, security is a key challenge for pharmaceutical industries but also for academic institutes in most cases. Effective protection of intellectual property and sensitive information requires, for instance, authentication of users from different institutions, mechanisms for management of user accounts and privileges and support for resource owners to implement and enforce access control policies.

In the next section, we are going to confront these requirements to the state of the art of grid technology described in deliverable D3.2.

6. ASSESSMENT OF GRID TECHNOLOGY FOR *IN SILICO* DRUG DISCOVERY

Once the use case and the requirements have been defined, this section reviews the status of the technology, as reflected in the deliverable D3.2 with respect to these requirements. The next section will analyse the requirements with respect to other issues studied in deliverable D4.2.

6.1. ASSESSMENT IN TERMS OF DATA INTEGRATION

The Strategic Research Agenda of the Innovative Medicines Initiative highlights the need to develop enhanced knowledge representation models and data exchange standards for complex systems, to build a core reference database of validated experimental data extracted from the literature, and to design standards for and build an expert tool to allow the federation of local databases in a secured environment.

The prerequisite for knowledge management is data integration. *In silico* drug discovery requires integration of data coming from all the “omics” experiments. These data are distributed in databases around the world.

Adoption of grids for biopharmaceutical research and development depends on the capacity of grids to enable the manipulation of data in a secure and efficient way. Biomedical data are complex and presented in multiple formats. .

From the description of data management services offered by grid infrastructures in deliverable D3.2, it appears that these services must be very significantly improved in order to allow such manipulations. The existing tools allowing the federation of databases in a secure environment (SRB, OGSA-DAI, AMGA) must be validated at a large scale on real biological and clinical data.

6.2. ASSESSMENT IN TERMS OF WORKFLOW ENACTMENT

In silico drug discovery requires a pipelined treatment of the compounds to select the most promising ones. The Strategic Research Agenda of the Innovative Medicines Initiative stresses also the importance of developing *in silico* models predictive of clinical efficacy and methods for predicting conventional and recently recognised types of toxicity. These different software tools need to be used in very early development in order to judge “Toxicodevelopability”. Indeed, the greatest need for the pharmaceutical industry is to detect the possibility of failure at the earliest stage possible,

As a consequence, *in silico* drug discovery workflow involves a large number of software tools which have to be used either sequentially or in parallel to assess the compounds’ chemical properties.

The status of grid-enabled workflow engines capable of handling licensed software is not discussed in deliverable D3.2.

6.3. ASSESSMENT IN TERMS OF ACCESS TO RESOURCES

European e-infrastructures like DEISA or EGEE provide the resources needed for large scale *in silico* virtual screening. The following steps of the drug discovery process are less CPU and storage demanding because they deal with a significantly reduced number of compounds.

Deliverable D3.2 documents the successful deployment of *in silico* drug discovery applications on e-infrastructures. Despite this success, there is still much room for progress in the design of these large scale docking deployment environments:

- Environments still need to be handled by grid experts. Progress is being made but significant improvements are needed to turn it into a service for the research community active in this domain. These improvements include the pipelining of the different relevant steps in the virtual screening process, the full automation of the fault tolerance system and the development of friendly user interfaces which allow easy job submission, job monitoring and data analysis
- Large scale virtual screening results should be integrated with relevant biological data on the selected targets to ease the comparison of *in vitro* and *in vivo* tests.
- Environments should allow easy access to the different grid infrastructures around the world. Interoperability of these infrastructures remains a major issue.
- Interoperability of the data produced by different *in silico* drug discovery projects must also be considered. This requires agreed standards for shared data repositories.
- Accounting is not yet available on e-infrastructures, although it is a necessary service for private customers.

6.4. ASSESSMENT IN TERMS OF COLLABORATION BETWEEN PUBLIC AND PRIVATE PARTNERS

The opportunity to address unmet medical needs has never been greater, but spiralling costs threaten to make the development of new drugs increasingly unaffordable for both developers and patients alike. Every effort must be made to make the drug development process more efficient, faster and more predictable. To be effective, the problem must be addressed by the active participation of all relevant stakeholders (academia, clinicians, patient organisations, large industry, SMEs, regulatory and ethics specialists). The collective impact is expected to come from the transparent, total-systems approach to the discovery and development process and in so doing enable each player to appreciate more fully the roles and needs of the others. As discussed in more detail in the next chapter, public organisations and even governments themselves must be engaged.

Grid technology offers very interesting opportunities to enable collaboration between public and private partners. Indeed, the grid is an excellent platform for information and knowledge sharing. Some pioneering initiatives like the Dengue docking project [6] of the Swiss Bio Grid initiative [7] involve both public and private research laboratories. However, the proper



tools must be enforced to provide the needed trust so pharmaceutical laboratories are willing to share information on the public grids.

7. ASSESSMENT OF ELSE ISSUES

This section analyses the use case described briefly in chapter 5 with respect to the Ethical, Legal, Social and Economic issues described in Share deliverable D4.2, “Bottlenecks and Challenges and RTD responses for legal, ethical, social and economic aspects of HealthGrids – Roadmap I”.

7.1. ETHICAL ISSUES

Deliverable D4.2 in section 5.1 identifies four principles of ethical behaviour in medicine and biomedical sciences:

- Autonomy
- Beneficence
- Non-Malfeasance
- Justice

Of these, autonomy will be of least concern to us here since no individuals are involved. We will address, in turn, beneficence and non-malfeasance, and then justice.

It is natural to assume that in a civilized society all drug development is intended to do good: to assume that no company sets out to develop drugs which do harm. However, ‘doing good’ may not – indeed, need not – be the main purpose, as perhaps when development takes place in order to market a ‘me-too’ drug, or in order only to extend patents and protect profits, so that in some circumstances there is at least a risk that the criterion of beneficence may fail. There are also cases, such as mental illness, where the nature of the good the drug will do may be at the very least debatable; the case of ‘hyperactivity’ in children may be a case in point. What is the health grid specialist’s share of the ethical burden in such circumstances? Like any other scientist working in a drug development programme, he or she would have to assess their own position.

Non-malfeasance presents an even thornier problem in drug development. From the point of view of a health grid worker, much of the grid-based work would be directed towards *positive* outcomes, such as ligand identification and molecular matching. How should undesirable side-effects be guarded against, especially if they might also arise and be identified in a similar fashion – through an unwanted ‘docking’ of molecules? To what extent is there a moral imperative, indeed does it even make sense, for health grid workers to attempt to address undesirable matches while seeking desirable ones? Must every experiment consider both positive and negative outcomes on pain of failing the criterion of non-malfeasance?

The ethical principle of justice is 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 of the greatest number is achieved. The proposed use case demonstrates the great potential ethical benefit of health grids. Indeed, grid-enabled *in silico* drug discovery allows researching new drugs at a much lower cost than traditional *in vitro* approaches. By doing so, it reduces the barrier to research and development on neglected diseases which suffer lack of R&D from pharmaceutical laboratories because there is no proper market.

Health grids are also promising environments for sharing information, expertise and resources with the academic laboratories of the developing countries which lack the needed infrastructure for successful research.

7.2. LEGAL ISSUES

Drug development raises a number of legal issues, particularly in late stages where clinical studies are conducted. Our use case focuses on drug discovery which is the initial stage of the drug development process, where no personal data have to be handled. Thus, data protection issues are not relevant at this stage.

However, as in all design work, a subsidiary claim of liability could arise if an experiment in docking, say, was badly designed and the experimenters considered to have been negligent. If at a subsequent stage some toxic or other undesirable effect is causally linked to poor experimental design, some liability might attach to those involved in the health grid stage of the research.

As pointed in deliverable D4.2, EU legislation on Intellectual Property Rights affects the uptake of grids. This is a major issue for drug discovery where the patenting of the potential drugs is needed before going further in the drug development process. *In silico* drug discovery using public grid environments raises the issue of ownership of both the methods used and the results achieved. Indeed, all the grid nodes which 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 a position to claim ownership of the method. This issue is presently addressed within the WISDOM collaboration by claiming no Intellectual Property Rights (IPR): All information, including analysis of potential hits, is made publicly available. If a group takes screening information and synthesizes the physical compound and tests it extensively in the wet-lab, it might establish IPR on their side and can establish claims on the physical compound and its behavior in biological assays as long as they cite the source for their initial analysis correctly. However, this leaves the social issue that the contribution of the wet-lab may itself be overvalued (see next section).

7.3. SOCIAL ISSUES

Most of the social issues identified in SHARE deliverable D4.2 deal with applications closer to actual provision and receiving of health services. Thus, their relevance to this specific use case is limited. On the other hand, the social impact of developing new approaches to drug discovery should not be neglected. Pharmaceutical companies are often criticized for investing only in research on drugs for widespread diseases. The scale allows a reasonable price and thus access for a large proportion of the affected people. Drugs for diseases that are not common/prevalent in rich countries either do not exist, or are extremely expensive. This is due to the enormous costs of drug discovery and development. In many cases, the cost of research in new, more effective drugs for such diseases is not likely to be covered by the expected revenue. For those who believe in the market mechanism as a means to address social issues, this should be an area of concern.

The changes associated with an eventual wide use of health grids for the purposes of drug discovery will change this situation significantly. By reducing the cost and increasing the effectiveness, i.e. probability of success, of drug discovery, health grids can make research financially worthwhile even for currently neglected diseases.

This brings us on the topic of equality, which is strongly related to the already discussed matter of IPR. It is worth considering an argument for promoting the current practice of using public resources and keeping results in the public domain to a certain extent. From a system point of view, this will be the best way to keep drug development costs, and thus drug prices at low levels, provided pharmaceutical companies are committed to the process, so that their wet-lab and clinical trial contribution is not so unreasonably priced that the drugs are finally too expensive, despite the public contribution. This is necessary given that those suffering most of neglected diseases are the poor – either in terms of poor countries, or in terms of the relatively poor segments of populations in

developed countries. Of course, this would require a significant commitment of political will and public resources.

An entirely different set of social issues around health grid participation in drug discovery arises in those cases where the ‘disease’ is of a social nature or at least has a significant social dimension. Consider, for example, drugs to combat obesity (when a healthy lifestyle might be the alternative) or against hyperactivity in children (said by many to be a social construct, not a disease category) or in elective medicine, such as assisted reproduction. The development of such drugs often calls for considerable public debate and health grid researchers may wish to take this dimension into consideration when determining choices or priorities among experiments.

The next section addresses this issue and gives some insights on how the usage of health grids for drug discovery can be made sustainable in the free market.

7.4. ECONOMIC ISSUES

The economic issues raised in deliverable D4.2 are fully relevant to the service discussed here:

- Cost-benefit and cost-utility analysis
- Benefits to patients, professionals, organisations and health systems
- Costs to patients, professionals, organization and health systems.

A major theme that needs to be addressed is the focus and perspective of such analyses. On the one hand, one can assess the economic impact of using health grids in drug discovery from a systemic perspective, including all stakeholders, and all possible costs, benefits, and utility. On the other extreme is a specific business case for one stakeholder, taking into account only the financial impact of health grids over a short period of time. Given that grid applications, by definition, require the participation of a large number of partners, and that drug discovery impacts on an even larger number of stakeholders (from organizations to individuals), these two kinds of analysis do not necessarily yield consistent results.

On the cost side, the main items are the set-up and maintenance of the technical infrastructure of grids and labour costs for the actual research activities. All other materials, licences etc. are covered regardless of whether grids are used or not. Thus, they are general sunk costs and do not enter the analysis. The health grid associated costs occur at the organisations involved in drug discovery and in the organisations participating in running the grids.

The benefits are spread wider. Drug researchers gain some time and efficacy in the process, at the time of use. Particular savings occur in the form of reducing the number of expensive physical compounds that need to be bought for *in vitro* tests, compared to traditional procedures achieving the same results. There is another group of stakeholders, namely citizens, that reaps significant benefits years after the utilisation of the grid. This is because the impact of a new drug, of course qualified by the probability of a new drug actually being brought on the market, only kicks in after all the other stages of drug discovery, between 5 and 10 years after the initial stages affected by this use case. As a satellite study to the German MediGRID project (<http://www.medigrid.de/>) commissioned by the Telematikplattform für Medizinische Forschungsnetze (TMF), Berlin, Germany (<http://www.tmf-ev.de/>) revealed, these benefits can constitute up to 50% of the total economic benefits from *in silico* drug discovery (based on preliminary data from an in depth analysis of the WISDOM initiative).

If all these factors are put together, the result is a cost-benefit comparison with a systemic focus and from an economic perspective. The above satellite study to the MediGRID project indicates, on the basis of analysing WISDOM, that this comparison is favourable, i.e. that benefits sufficiently outweigh the costs. An important feature of the WISDOM initiative is that the utilised grids already

existed. This means that only the costs proportionate to the use are taken into account. If the full costs of setting up the grids were to be incurred, the outcome would have been different.

The situation may also look different if the organisations participating in WISDOM did not have access to public funding. Moreover, the current success of WISDOM is based on project-based subsidies. Public financing of continuous activities, ensuring sustainability, is extremely difficult to organise. Thus, attention should be given to an issue already raised in D4.2 – private incentives.

Matching the realised benefits to the cost for each stakeholder will bring the appropriate insight on whether the market alone can sustain a grid-based service to drug discovery research. In that exercise, focus must be on financial items rather than including less tangible benefits like level of suffering. A health grid service will be sustainable if, and only if, the financial benefits, including indirect ones like improved market position or image, at least cover the costs of participating over a reasonable period of time for each individual participant. In other words, there must be a business case for each stakeholder.

Following the success of the WISDOM initiative, plans are being made for commercialising a grid-enabled drug discovery service based on health grids. The business case for such a service should be based around tangible benefits to potential customers, such as [8]:

- reducing time for testing the impact of chemical compounds on a target protein playing a vital role in a disease
- using existing resources instead of carrying significant set-up costs for computing power
- enabling focussed in-vitro testing, with a significantly higher probability of success
- reducing the cost to build focussed compound libraries.

The customers of such a service would be academic laboratories, SMEs researching targets and the pharmaceutical industry. These organisations must see the benefits for themselves and be prepared to pay for the service. The price of the service will be somewhere between the cost of providing the service and the value of the benefits reaped by the customer. The exact position depends on the bargaining power of buyer and supplier.

At the same time, the public sector should not be completely left out of the equation. It is perfectly valid to build a business case at least in part depending on public funding. Because of the sizable benefits to society, through the citizens' share of the benefits, the case of grid-enabled drug discovery being subsidised over a longer period of time is strong.

8. ROADMAP

To build its Strategic Research Agenda, the Innovative Medicines Initiative has set up a work package to provide input on technology required for establishing a knowledge management environment capable of supporting the scientific objectives of biopharmaceutical research and development, to identify gaps in current technologies and to offer recommendations on how to bridge those gaps. On page 73 of the SRA report, the following sentence has been extracted from the summary of the section on Knowledge Management: *“From the technical point of view, the requirements can be met using a distributed/federated, multi-layer, service oriented, and ontology driven architecture.”*

This is indeed very close to our understanding of a knowledge grid. Moreover, the Knowledge Management work package identified the following actions (p13 of report [5]) in order to address the scientific requirements:

1. Develop a strategy to identify the areas of interest to all stakeholders,
2. Provide mechanisms for data federation across heterogeneous data sources,
3. Provide a flexible and secure collaborative environment serving all stakeholders,
4. Provide standards and mechanisms for consistent data integration and data sharing,
5. Provide standards and mechanisms for consistent integration of complex scientific tools and computational models,
6. Insure interoperability of computing services across organizations,
7. Develop broad and generic research projects for bridging gaps in current technologies.

Most of these actions are explicitly identified in the technical roadmap described in deliverable D3.3 although using different phrasing:

- Action 2 can be achieved by the provision of reliable and robust grid services for the federation of heterogeneous databases and their deployment on an operational data grid (milestone MD2 of deliverable D3.3).
- Action 3 corresponds to the deployment of a knowledge grid (milestone MD3).
- Actions 4 and 5 are directly related to the standardization actions recommended in D3.3 (milestones MS1 and MS2). However, the scope here is extended beyond medical imaging and electronic health records to molecular biology
- Action 6 is enabled by the deployment of a computing grid and the availability of a robust and easy to install grid middleware (milestone MD1)
- Action 7 is similar in scope to one of the three R&D activities proposed in D3.3 chapter 5, namely to deploy biomedical applications to validate and guide the technology evolution.

Action 1 is not a technical recommendation, rather it is an organizational one for the community, although, as in all socio-technical systems, a requirements engineering stage would be included. This, however, is not unique to grid computing.

In the rest of this chapter, we are going to further explore the research topics, the deployment actions and the timeline proposed in the HealthGrid technical roadmap and in the Innovative Medicines Initiative Strategic Research Agenda (IMISRA report, [5]).

8.1. RESEARCH TOPICS

The scientific requirements identified by the IMISRA report can be summarized as follows:

- Capacity to search, query, extract, integrate and share data in a scientifically and semantically consistent manner across heterogeneous sources (public and proprietary) ranging from chemical structures and “omics” to clinical trial data,
- Capacity to integrate and share scientific tools (e.g., modelling, simulation) as modules in a generic framework and apply them to relevant dynamic data sets,
- Expressive data representation and exchange standards,
- Dynamic and customizable configuration of applications,
- Encapsulation of validated physiological models, when applicable,
- Flexible, secure (covering all aspects of data protection encountered in a biomedical context), and scalable IT infrastructure.

Health grids can become the infrastructure for biopharmaceutical research and development provided the technology matures to support a distributed/federated, service oriented, and ontology driven architecture which provides a collaboration medium, facilitates effective computation and is capable of generating, organizing and managing knowledge

Moreover, in complement to the presently proposed roadmap described in D3.3, a number of gaps are to be addressed notably in the area of data representation and exchange standards, ontology development, data protection, and text mining. As a consequence, several research topics identified in IMISRA report must be included in the Health grid roadmap:

1. Develop enhanced standards for data protection in a web services environment,
2. Develop standards and models for exposing web services (semantics), scientific services, and the properties of data sources, data sets, scientific objects, and data elements,
3. Develop enhanced knowledge representation models and data exchange standards for complex systems, presently largely inconsistent or incomplete, looking for synergies with other initiatives,
4. Develop new, domain-specific ontologies, built on established theoretical foundations and taking into account current initiatives, existing standard data representation models, and reference ontologies,
5. Develop advanced text mining tools for capturing implicit information about complex objects, relationships and processes, as described in patents and literature, beyond and above simple pair-wise relationships between entities,
6. Build a core reference database of validated experimental and clinical research data extracted from the literature,
7. Design standards for and build an expert tool (ontology/schema/rules negotiator) for exposing the properties of local sources in a federated environment,

8. Design standards for and expert tool (services/data negotiator) to guide users through the complexities of the data, data models, simulation and modelling tools, etc.
9. Analyse the costs and benefits associated with developing and running health grids, for society as well as for individual stakeholders.

8.2. DEPLOYMENT ACTIONS

The technical infrastructure recommended by the IMISRA report involves the following independent layers:

- Infrastructure: defines the whole of hardware and software components that support basic operations and provide such functionality as availability (Quality of Service), data integrity, etc. (i.e., firewalls, redundant systems, backup infrastructure, computer clusters, etc.)
- The backbone: comprising a set of services providing basic functionality and interoperability (e.g., messaging, brokering). The backbone is also in charge of managing services and data access (security),
- Data access to heterogeneous resources could be provided through two sub-layers:
- Data Virtualization layer: to decouple data from their local schema and make data access platform- and schema-independent,
- Data Abstraction layer: to provide a common view of all accessible data via a set of ontology / rule-mapping mechanisms,
- Services layer, making services (core, administrative or scientific services) accessible over the backbone and connecting to data resources,
- Connections layer: providing a secure access point to all authorized users and processes,
- Organizations: describing users and allowing them to share data, share services, and collect information.

The grids described in deliverable D3.3 are definitely able to provide the Infrastructure, Backbone, Services, Connections and Organizations layers, but their interest will be limited for researchers as long as they don't provide data access to heterogeneous resources. The recommended approach to achieve data access in the IMISRA report is through the Data abstraction and Data Virtualization layers. Another approach developed by the EMBRACE Network of Excellence is to develop web service interfaces to all the data resources in molecular biology. In the case of Embrace, data integration is achieved through the definition of standardized interfaces to the resources. This approach allows a progressive integration of the resources but depends on the development of new standards and ontologies for data exchange and representation as stressed in the previous section.

These different strategies for heterogeneous data sources integration on grid environments must be properly evaluated through deployment actions. The IMISRA report recommends building a core reference database of validated experimental biomedical data extracted from the literature and creating disease-specific European Imaging Networks for establishment of



standards, validation of imaging biomarkers and development of regional centres of excellence. It would be highly desirable to explore the feasibility of a grid approach to these projects following the example of BIRN in the United States. .

8.3. TIME PLANNING

The IMISRA report recommends 440 millions euros of investment for a period of 7 years corresponding to the FP7 program 2007-2013 in order to revitalize the biopharmaceutical R&D environment in Europe. Out of it, about 49 millions per year would be devoted for the development and implementation of the knowledge management part of the Strategic Research Agenda.

This compares to the estimated 10 years in deliverable D3.3 to deploy a knowledge grid. The recommended 21 million euros investment in the IMISRA report to develop knowledge management could directly contribute to building the knowledge grid provided the developments coming out this research are made publicly available to the research community at large.

9. CONCLUSION

In conclusion, this document attempted to analyse the roadmaps proposed for health grid from technical and ELSE perspectives in view of the Strategic Research Agenda of the Innovative Medicines Initiative and more specifically a specific use case relevant to drug discovery.

The first outlines of the technical and ELSE parts of the Health grid roadmap developed in deliverable D3.3 and D4.2 identify useful and necessary steps toward building the flexible, secure and scalable IT infrastructure required for biopharmaceutical research and development.

The *in silico* drug discovery use case shows how grids can start to be used today for biopharmaceutical research and development but proper management of the knowledge input to the process as well as generated by it is needed to really bring added value to the drug development process.

However, further, specific steps are needed to improve knowledge management. In the biopharmaceutical R&D process, an enormous quantity and diversity of data is created. A critical factor is the ability to turn this mass of information into actionable knowledge. The need to develop enhanced knowledge representation models and data exchange standards for complex systems, as well as to design standards for the federation of data bases must be properly addressed in the roadmap.

The roadmap generally stresses the importance of deploying applications on the existing grid infrastructures. More specifically, adoption of the grid technology in the biopharmaceutical sector would gain from using it to build a core reference database of validated experimental data extracted from the literature or to create disease-specific European Imaging Networks.