European
Life Sciences Infrastructure
for Biological Information

Business Case
European Life Sciences Infrastructure for Biological Information

Prof. Janet Thornton
EBI Director
ELIXIR Coordinator
T +44 (0) 1223 49 2670
elixirpm@ebi.ac.uk
The ELIXIR Preparatory Phase Consortium

EMBL – European Bioinformatics Institute (EMBL-EBI)
Biotechnology and Biological Sciences Research Council (BBSRC)
Federal Ministry of Education & Research (BMBF)
Barcelona Supercomputing Center (BSC)
Spanish National Cancer Research Centre (CNIO)
Council for National Research (CNR)
Center for Advanced Studies, Research & Development in Sardinia (CRS4)
CSC Scientific Computing Ltd.
German Research Foundation (DFG)
Technical University of Denmark (DTU)
Erasmus Medical Centre
Institute of Enzymology (ENZIM)
Genome España
Helmholtz Zentrum München
Institut National de Recherche en Informatique et en Automatique (INRIA)
Linköping University
Ministry Of Science & Technology (MOST)
Medical Research Council (MRC)
Natural Environment Research Council (NERC)
Netherlands Organisation for Scientific Research (NWO)
The Icelandic Centre for Research (RANNIS)
Radboud University Nijmegen
Wellcome Trust Sanger Institute
Sardegna Ricerche
SIB Swiss Institute of Bioinformatics
Syngenta Ltd
Technical University of Braunschweig
University of Bordeaux 2
Swedish Research Council
Wellcome Trust
Institut National de la Recherche Agronomique (INRA)
Institut National de la Santé et de la Recherche Médicale (INSERM)

The European Life Sciences Infrastructure for Biological Information (ELIXIR) project is funded from the European Commission’s Framework 7 Capacities Programme for Research Infrastructures under grant agreement no. 211601.
# Executive Summary

1. **The need for ELIXIR**  
   1.1 Meeting the Grand Challenges  
   1.2 The importance of biological data  
   1.3 Maintaining open access to biological data  
   1.4 The data deluge  
   1.5 The need for a pan-European infrastructure for biological data  
   1.6 Integration with other Biological and Medical Infrastructures  
   1.7 Responding to the needs of industry  
   1.8 Impact on society

# The ELIXIR Model

2.1 The preparatory phase  
2.2 Mission statement and core objectives  
2.3 Topology  
   2.3.1 The ELIXIR Hub  
   2.3.2 The ELIXIR Nodes: criteria and networks  
   2.3.3 Services delivered by ELIXIR Nodes  
2.4 Governance model  
   2.4.1 ELIXIR Board  
   2.4.2 Scientific Advisory Board  
   2.4.3 ELIXIR Executive Management and Secretariat
Executive Summary

Biological information is of vital significance to life sciences and biomedical research, which in turn are critical for tackling the Grand Challenges of healthcare for an ageing population, food security, energy diversification and environmental protection. A robust, effective and sustainable biological information infrastructure is therefore essential for meeting the strategic priorities of the Member States, and for the success of both academic, governmental and industrial research in Europe, including the pharmaceutical and biotechnological sectors.

The challenge that biomolecular data resources face today is two-fold. Firstly, new technologies such as next-generation DNA sequencing are generating massive amounts of data. It has been estimated that between today and 2020 these new technologies will produce data at up to one million times the current rate. Secondly, there is an emerging and pressing need to provide infrastructure that will meaningfully integrate new types of data to be collected in the future. For instance, data associated with images acquired from microscopes will be comparable with the amount of data that is currently being produced by the Large Hadron Collider at CERN.

The collection, curation, storage, archiving, integration and deployment of biomolecular data is an immense challenge that cannot be handled by a single organisation or by one country alone, but requires international coordination. Over recent years European countries have invested heavily in research that produces these data. It is now recognised that there is an urgent need for a pan-European infrastructure that will facilitate the process of extraction of optimum value from current and planned investments in this area.

ELIXIR’s mission is “to construct and operate a sustainable infrastructure for biological information in Europe to support life science research and its translation to medicine and the environment, the bio-industries and society.”
By the end of 2011, ELIXIR will have completed a four-year preparatory phase funded by the EU’s Seventh Framework Programme (FP7) as part of the European Strategy Forum on Research Infrastructures (ESFRI) process. The preparatory phase has included a thorough stakeholder analysis and the development of scientific, technical and legal models for the construction and operation of the infrastructure. The objective has been the development of an infrastructure solution that can deal effectively with the challenges of growth and new types of data, as well as safeguard and make the most of national investments in life science and biomedical research.

After consulting with hundreds of data providers and users from academia and the private sector, ELIXIR’s steering committee concluded that ELIXIR should be built as an infrastructure that is distributed across several sites hosted by centres of excellence distributed throughout Europe (the ‘ELIXIR Nodes’); these sites should be connected to a central hub (the ‘ELIXIR Hub’) located at the European Molecular Biology Laboratory’s European Bioinformatics Institute (EMBL-EBI) at Hinxton, UK.

The ELIXIR Hub will accommodate the ELIXIR Executive Management and Secretariat. It will coordinate services from the European Data Centre run by EMBL-EBI as well as coordinating services run from the ELIXIR Nodes. It will also provide core data (i.e. data essential to virtually all life-science research areas) to users.

ELIXIR Nodes will provide a range of services that are complementary to those provided by the Hub. The integration of the ELIXIR Nodes will provide essential distributed components of the ELIXIR infrastructure at the European level. We envisage the emergence of new core services based in the nodes, which will be essential to biology in the future.

It is anticipated that ELIXIR Nodes will be sited at existing centres of excellence. They will be internationally competitive, relevant at the European level and able to demonstrate sustainability of funding. They must be legal entities that are capable of entering into binding agreements and receiving funding. Some countries may wish to have several ELIXIR Nodes; some may only want to have one whilst others may wish to collaborate in the creation of regional ELIXIR Nodes.

A suitable legal and governance model has been developed. This model takes advantage of EMBL’s legal status to host the infrastructure and to secure sustainable funding from ELIXIR members. The Member States that support ELIXIR and EMBL as the host of the ELIXIR Hub will sign an International Consortium Agreement (ICA). This will clarify how ELIXIR will be established as a “Special Project” within EMBL, allowing ELIXIR funds to be ring-fenced from other EMBL activities. The ICA will also clarify the roles, rights, duties and responsibilities of all members of the ELIXIR Consortium.
Executive Summary

Based on the ICA, ELIXIR will have its own governing body with representatives from ELIXIR Member States and EMBL on the ELIXIR Board. The ELIXIR Executive Management and Secretariat, as part of the ELIXIR Hub, will be primarily responsible for the day-to-day running of ELIXIR. It will coordinate with ELIXIR Nodes, and will coordinate ELIXIR representation in relevant international collaborations and consortia.

The national organisations providing ELIXIR Nodes will be linked to the Hub by separate bilateral service level agreements between ELIXIR and ELIXIR Nodes. The ELIXIR Board, in light of the recommendations of an independent Scientific Advisory Board (SAB), will oversee the appointment of ELIXIR Nodes and their continuation after a defined period of time.

ELIXIR Member States will jointly fund the ELIXIR Hub. ELIXIR Nodes will receive decentralised funding from national funding organisations or consortia of Member-State funding agencies. Ideally, this funding will be provided from long-term infrastructure funds aligned to strategic priorities, for example, through National Infrastructure Roadmaps. Preliminary budgets have been prepared for illustrative purposes (see Section 3.5).

Countries that become ELIXIR Member States will benefit from: enhanced exploitation of the data produced by their investments in biomedical research; optimisation of the value of data by rapid and timely translation into other sectors such as clinical, agricultural and environment research; enhanced competitiveness through European coordination; enhanced bioinformatics capabilities and expertise within their funded communities; and facilitation of coherent strategic planning of their resource deployment.
On the basis of this Business Case, interested states are invited to sign a non-binding Memorandum of Understanding (MoU) in order to enter into negotiations for the construction of ELIXIR. Negotiations will begin when there are six signatories to the MoU.

Following the signing of the MoU the next steps in the construction of ELIXIR will be:

1. **Negotiation of an International Consortium Agreement (ICA);**

2. **Approval from EMBL Council that ELIXIR is included as an EMBL Special Project in the Indicative Scheme 2012–2016; to partly fund the ELIXIR Hub.**

3. **Conclusion of the ICA with the ELIXIR Member States and EMBL;**

4. **Negotiation with countries and national funding organisations for the establishment of ELIXIR Nodes, based on suggestions collected through a recent open call;**

5. **Conclusion of bilateral service level agreements between ELIXIR and the institutions hosting initial ELIXIR Nodes based on scientific and technical criteria. This can only occur once ELIXIR Nodes have obtained funding support for ELIXIR-related activities from national funding organisations;**

6. **Securing the commitment of additional countries and conclusion of agreements with new ELIXIR Nodes as needed in the future.**
1 The need for ELIXIR

1.1 Meeting the Grand Challenges

Europe and its nations are faced with daunting economic, demographic, social and environmental pressures. These Grand Challenges include an ageing population, environmental degradation, the loss of biodiversity, the growing demands for food, water and energy, and the need to respond rapidly to emerging global threats such as pandemics and bioterrorism. The surest way to meet these challenges is to tap into the huge potential of ideas, resources and people across national borders. This necessitates a pan-European collaboration, long-term strategic thinking and economies geared towards innovation that will deliver new solutions and create new wealth.

Fortunately, European governments and European institutions are in agreement that only by persistently investing in research and innovation – particularly in times of austerity – will they be able to plant the seeds of sustainable prosperity for the future. They agree that meeting the Grand Challenges calls for investment in science and innovation.

In October 2006 the European Strategy Forum on Research and Innovation (ESFRI), a body set up by 33 countries at the initiative of the European Council, identified 35 pan-European research infrastructures that are of key importance for the development of science and innovation. Among them, ELIXIR is one of very few considered to be of global significance.

ELIXIR is a major and essential upgrade of Europe’s bioinformatics infrastructure. At the heart of ELIXIR lies the wealth of biological data that researchers in the life sciences produce every day. This includes data about the genes of humans, as well as other species, data about chemical molecules that can help us diagnose and cure disease, and data about proteins, which form the building blocks of every living thing on our planet.

These biological data must be stored, annotated and curated by experts using innovative tools and vast computational and storage resources. These data are the goldmine of the biological revolution. They are not only essential for research in the life sciences, but are in themselves the bedrock of new developments in healthcare, food production and security, environmental protection, energy and a host of other industries that innovate on the infinite potential of life.

Today, Europe’s life scientists face the following challenges: firstly we must manage the current data deluge, secondly we must integrate the data and thereby reduce fragmentation of effort and research; thirdly we must incorporate and exploit new types of data; and lastly it is essential that we maintain open access for biological data to enhance competitiveness and innovation. These challenges are too vast for any single institution or country, and must therefore be managed by joining forces at the European and global level.

The European Member States have realised this and included ELIXIR in the first European roadmap for research infrastructures in 2006.
1.2 The importance of biological data

Data has become an essential commodity for biological research. Ten years ago, if a medical researcher needed to find a gene involved in a certain disease, s/he might need three years of laboratory work. Today, thanks to genomic information stored in large public databases, the same task may take less than 30 minutes.

Using innovative tools to search biological databases, today’s biomedical researchers can arrive at new insights and solutions to the Grand Challenges faster than ever before.

For example, by identifying patterns of genes that are active in different tumours, researchers can predict how aggressive the tumour is and decide which drugs to treat it with. Global, collaborative projects such as the 1000 Genomes Project sequence the genomes of several thousand individuals and make the data openly available. This invaluable information can then be used in numerous studies to associate minute differences with susceptibility to common life-threatening conditions such as heart disease and diabetes mellitus. It is the kind of research that will usher in a new era in diagnosis and therapy by personalising medicine.

Another example is the Barcode of Life initiative. “DNA barcodes” are short pieces of information that identify an organism as belonging to a particular species. This DNA-based library of life will be a global standard and will allow applications ranging from the protection of endangered species and sustaining natural resources through control of agricultural pests and food labelling.

**Figure 1** compares the annual cost of generating novel data on protein structures worldwide with the cost to Europe of storing it in the PDB, one of the oldest biomolecular databases in the world. Indeed, the “European part” of PDB (which stores all known protein structures determined worldwide since the 1960s) costs much less than 1% of the cost of re-generating one year’s data. wwPDB is the worldwide Protein Data Bank.
New research methods generate many different types of data, which need to be pieced together (integrated) to gain valuable knowledge of how living systems function. For example, by looking at the genetic, dietary and metabolic differences of individuals who live a long, healthy life versus those who do not, it may be possible to understand the mechanisms of healthy ageing and use this knowledge to enhance the lives of our ageing population. Such research requires mathematical models, statistical analysis, and the sustainable storage and processing of large sets of biological data. Innovation in the life sciences has moved from the traditional “wet lab” to the computer. Data and computer power combined with human creativity are now key drivers of progress in the life sciences.

Figure 1 compares the annual cost of generating novel data on protein structures worldwide with the cost to Europe of storing it in the PDB, one of the oldest biomolecular databases in the world. Indeed, the “European part” of PDB (which stores all known protein structures determined worldwide since the 1960s) costs much less than 1% of the cost of re-generating one year’s data.

In summary, investing in a sound biological data infrastructure is not only crucial for pushing the envelope of innovation forward, but it is also the best, and cheapest, insurance policy a government, or a continent, can make to secure its huge investments in life science research.

1.3 Maintaining open access to biological data

Considerable moral and ethical arguments support the view that basic life science information constitutes a testament of human and natural evolution and advancement. As such, this wealth of knowledge should be freely available for all to access, study and process.

There are also strong economic and scientific reasons for maintaining open access to biological data. For life science researchers, in academia as well as industry, open access to bioinformatics resources provides a valuable path to discovery, one that in many other areas of research is limited by commercial confidentiality.

Charging for that data, or seeking to restrict access through exercising Intellectual Property (IP) rights, would seriously impede the ability of research and industrial organisations to exploit data and return benefits – indeed many potential discoveries would be lost to legal red tape processes and contractual wrangles. It is therefore vital to maintain open access to biological data. This view is strongly supported by industry, which appreciates the pre-competitive value of accessing valuable data free of charge. (NB Data derived from individual humans is rarely completely open access for reasons of personal security and privacy, but providing secure access to such data is also a priority for future research).

ELIXIR, as a pan-European bioinformatics infrastructure principally funded by public funds, will guarantee that open access to biological data is maintained. Speaking with a single voice will strengthen Europe’s influence in such global discussions.
1.4 The data deluge

One of the biggest recent breakthroughs in the life sciences has been the development of DNA sequencing technologies, to reveal the genetic code of life. Knowledge of DNA sequences has become indispensable for basic biological research as well as in numerous applied fields such as diagnostics, drug development, biotechnology, forensic biology and systems biology.

Over the past decade, innovative technical solutions developed to meet the high demand for low-cost sequencing have resulted in high-throughput sequencing technologies that parallelise the sequencing process, producing billions of bases of nucleotide data per experiment. The wide uptake of this so-called ‘next-generation’ sequencing and other ultra-high throughput technologies by life scientists has led to unprecedented growth in the amount of data generated.

Next-generation DNA sequencing has become so widely used that it is believed to be a ‘disruptive technology’: it is so much better than the technology that it replaces that users have difficulty adapting to it. Modern sequencers are capable of analysing the equivalent of a human genome every 14 minutes at a cost of US $5,000. This rate is 400 times greater than in the year 2000, when the draft human genome was first published. It is expected that further developments will result in sequencers between a thousand times and a million times more productive over the next ten years. The observed expansion of data volume has been supra-exponential.

Figure 2 shows the trend in storage capacity needed to store the volume of biological data at EMBL-EBI (in Terabytes; a terabyte is a million million bytes). This trend is expected not only to persist but to become steeper still, posing a serious challenge to existing bioinformatics infrastructures in Europe.
Data coming from novel gene sequencing technologies are only one of the data challenges faced by today’s life scientists, the other being the integration of different data in meaningful, research-supportive ways. Life science research is becoming increasingly collaborative and complex, using several different technologies to understand organisms and diseases at the systems level.

ELIXIR’s vision for the future is to provide researchers in academia and industry with seamless access to biological information that will revolutionise discovery in the life sciences, by integrating data at different levels of analysis, for example from molecular biology to clinical practice.

The enormity of these tasks necessitates cross-border cooperation on a scale unprecedented to biological and biomedical research. The most obvious solution is the development of a pan-European initiative that will enable European countries to pool their resources in a coordinated manner.

ELIXIR will, for the first time, enable pan-European coordination of all scientific and technical issues related to handling the biomolecular data resources. This is essential to be able to meet the challenges that Europe faces regarding the collection, storage and dissemination of biomedical data.

1.5 The need for a pan-European infrastructure for biological data

Despite significant investments in life sciences in Europe, existing biological data centres lack a centrally coordinated strategy that results in efficient integration of, and synergy between, different countries and programmes. To date the challenge for the core data resources has been met mostly through the EMBL-EBI, with funding from EMBL’s 20 Member States. With a few notable exceptions (e.g. Switzerland), most nation states have been reluctant to invest heavily in providing global data infrastructures from national funds.

The outcome is that beyond the core resources, which are mainly based at EMBL-EBI, the European bioinformatics landscape remains largely fragmented across borders and institutions. Faced with the ongoing data deluge, as well as the scientific challenge of heterogeneous data integration, no single nation or single organisation can effectively address this mounting problem in Europe.

Addressing this issue is of critical importance and, evidently, the solution requires a coordinated, pan-European effort and investment in a pan-European infrastructure for bioinformatics services. In order to be technically effective and economically efficient this pan-European infrastructure should comprise a data infrastructure, a compute infrastructure, a training infrastructure and a tools and standards infrastructure.

Data infrastructure

The data infrastructure should include the data resources plus the infrastructure for the tools that are used to access and interpret the data.

Currently EMBL-EBI, strengthened by with many strong collaborators throughout Europe and beyond, has acted effectively as a central European hub for core data. Many smaller,
specialised data resources have emerged, but these are rarely coordinated with the core data and are often difficult for researchers to find. Indeed, the increasing plethora of niche databases makes the bioinformatics landscape needlessly complex for its users. In the face of the data deluge, limited resources and lack of coordination, EMBL-EBI will rapidly become unable to meet the full European demand.

The future pan-European data infrastructure should therefore:

- Enhance the existing collective data capacity to meet rising demand (the "data deluge");
- Build a distributed infrastructure that enables full data integration so that the collective, expanding capacity across the continent is used optimally;
- Establish universal principles for optimising the use of existing data capacity (such as assessing which data should be stored and made available to users);
- Present a transparent single interface to a distributed infrastructure.

**Compute infrastructure**

To date, life scientists have rarely needed to use European e-infrastructures, which have been mostly developed to provide services for the physical sciences. As a consequence, the existing compute architecture has been optimised for the requirements of physical science communities, which typically address compute-intensive problems. In contrast, biologists require data-rich, massively parallel queries. However, this is now changing because biologists are generating data at previously unprecedented rates, comparable to those generated by the physical sciences. Unless an increase of an order of magnitude in compute infrastructure capacity occurs (including storage and processor power) in a manner suitable to meet demand for life sciences, it is anticipated that the existing European data infrastructure will not be adequate to respond to challenges driven by the data deluge.

The future pan-European compute infrastructure should therefore:

- Enhance the capacity of the existing compute infrastructure to meet demands that evolve from the data deluge;
- Maintain integration despite distribution, with access to central databases such that the collective expanding capacity across the continent is utilised optimally.

**Training infrastructure**

In the past, biological data resources were used by a relatively small community of ‘bioinformatics aware’ researchers, but as access to data becomes increasingly central to biomedical research, this user base is growing and diversifying to include, for example clinicians, experimentalists in the pharmaceutical industry, plant breeders and environmental scientists. In 2010, the EMBL-EBI’s data resources were accessed from 3 million unique IP addresses. It is anticipated that the increase in demand for online bioinformatics services by new users will be accompanied by an increase in demand for training for the use of such services, and this has been borne out by increasing demand for places at training courses organised by EMBL-EBI.

The European training infrastructure is not adequate to accommodate this anticipated increase. Failure to build adequate capacity to meet training needs will create a bottleneck that will impede the optimum use of life science data.
The future pan-European training infrastructure should therefore:

- Empower European researchers to make effective use of the data by making bioinformatics user training accessible to the rapidly growing and diversifying user community;
- Achieve appropriate coordination of the currently fragmented training infrastructures such that the collective expanding capacity across the continent is utilised optimally;
- Integrate training with the development and diversification of data resources and tools to keep pace with changes;
- Enable the development of new training programmes, especially in new accession states, and their integration with existing initiatives.

Tools infrastructure

The data resources that will be provided through ELIXIR will be used in many different ways and in numerous combinations using a huge number of different analytical software tools written by investigators all over the world. Typical analyses can use dozens of tools linked together into data processing pipelines. A key role for ELIXIR is to provide an infrastructure to support maximum efficiency and effectiveness in the deployment and use of these tools.

During the course of the Preparatory Phase we identified four distinct capabilities necessary for the ELIXIR Infrastructure for tool integration. They are:

- Tool discovery: It is essential that potential users of ELIXIR can find the most appropriate tool as quickly as possible and that the tools are available for download and deployment;
- Ease-of-use: Tools should be usable by as wide a range of users as possible without the need for specialised knowledge or IT skills;
- Benchmarking: It is often the case that several tools seem to perform equivalent functions. ELIXIR’s users need access to the results of systematic benchmarking to help them select the most appropriate tool based on objective measures of performance;
- Interoperability: Bioinformatics tools are seldom used in isolation and the means must be provided to link them together in useful ways without the need for specialised knowledge or skills.

Standards infrastructure

Standards are of vital importance to the success of ELIXIR and it is inevitable that ELIXIR will be heavily involved in the development, implementation and deployment of standards. To ensure that this happens in a coordinated manner, ELIXIR will need to create a European-wide infrastructure for standards. During the course of the preparatory phase we identified the following areas where this will be necessary:

- Programmatic access: standardisation of the technology to be used to build connections to databases and tools;
- Nomenclatures: harmonisation of names and symbols of biological entities;
The need for ELIXIR

1.6 **Integration with other Biological and Medical Infrastructures**

Research in the life sciences has changed markedly during the last few years. Traditionally, scientific progress was made by individual research groups, which focused on specific problems (classical reductionist hypothesis-driven research). Nowadays, more holistic approaches (discovery-driven research) that rely on large distributed teams with common access to extensive infrastructure are becoming increasingly important.

Obtaining deeper insight into the role of genes and their products in the context of the complexity of living systems, or understanding the genetic and environmental factors impacting on human health, require interdisciplinary approaches involving a broad spectrum of demanding technologies and resources.

To overcome these challenges, biomedical scientists are now adopting large-scale collaborative approaches, involving scientists from various fields within and outside the biological and medical sciences (BMS) community. Even within the BMS community there is a need to promote greater cross-fertilisation between sub-disciplines. The current fragmentation of scientific communities in Europe is a major obstacle to this, and single institutions or even national networks are often unable to address this need.

Apart from ELIXIR, the ESFRI process has identified 12 other BMS research infrastructures, which are developing in parallel with ELIXIR. These include mouse clinics, structural biology laboratories, biobanks and support for clinical trials. These infrastructures either produce or use data that lie within ELIXIR’s remit. Therefore, close co-operation and coordination with these infrastructures to ensure seamless deposition, access and integration of their data is essential.

ELIXIR should therefore:

- Establish and promote good co-operation between the BMS infrastructures, helping to create an e-infrastructure that allows transfer of data and easy, yet secure access to data for all Europe’s life scientists;
- Provide the proper framework for registration of resources, as well as access to and integration of high-end technologies and expertise, required to address key scientific questions.

1.7 **Responding to the needs of industry**

Many commercial organisations are poorly equipped to deal with the challenges caused by the data deluge using internal resources alone; this is true both for large companies and for small-to-medium enterprises (SMEs).
In addition, the data deluge has arrived at a time when information system budgets are static or shrinking, with reduced internal infrastructure and manpower. Consequently, few industrial organisations can now afford to work alone, and many are re-evaluating the need for pre-competitive IP generated internally. An increasingly collaborative landscape is emerging, with a range of pre-competitive alliances being formed in specialised sectors.

There is a strong business requirement to organise and integrate public and proprietary internal data. Clear and unambiguous data and meta-data standards are necessary and these must be adopted across industry. Thus, there is an unequivocal need for industry to work with the academic community to develop and apply these standards universally. This can only benefit both parties and will be a real boon to the increasingly prevalent public–private research partnerships within Europe.

The benefits to European industrial organisations of such a service-oriented infrastructure are likely to be numerous. The opportunity to reduce duplication of internal effort will make European life science companies more competitive through significant cost reductions. The benefit of integrating external public data with internal data in an efficient way should greatly reduce the probability of overlooking significant facts in the interpretation of internal data. The improved insight into these data will lead to more informed decision making at critical points in the research/development pipeline. It is also considered that open innovation will be supported by these initiatives. The adoption of widely supported standards will reduce barriers between information sharing and help to facilitate cross-company research partnerships. More information on the role of industry in ELIXIR is provided in section 2.8.

1.8 Impact on society

Today, Europe stands at a crossroads. Years of economic and social progress have been undermined by the economic crisis. Public budget constraints, major demographic changes and increasing global competition compound the need to revise old practices and explore new initiatives for economic prosperity.

Fortunately, there is an abundance of potential in Europe. Our countries enjoy a long tradition of world-class standards in higher education. Some of the world’s top universities and research institutions reside on our continent. Europe’s knowledge capital is enormous. However, new models must be pursued to tap into this capital and spur innovation.

Without doubt biosciences are at the core of current and future innovation. Biological research aims to address the most fundamental and urgent global challenges, which include an ageing population, environmental degradation, and dwindling supplies of food and fuel. As we look towards the biosciences to help us address these challenges, the potential for job creation in the biosciences is enormous. Bioscience-related industries will contribute significantly to the financial and social wellbeing of those countries that have invested in these critical, value-adding sectors.

A pan-European bioinformatics infrastructure will deliver the backbone for new discoveries that address and meet the Grand Challenges and, in doing so, will:

- Spur economic development and innovation, and thus create new, knowledge-intensive, highly skilled jobs;
• Generate opportunities to increase Europe’s knowledge-based industry and competitiveness, by supporting the success of innovative small-to-medium biotech enterprises;
• Attract world-leading scientists to Europe (increase “brain gain”) and retain key expertise (stem “brain drain”);
• Increase the application of new innovations in the biotechnology and pharmaceutical industries, as well as in agriculture and environmental protection;
• Safeguard the investment that nations have already made in funding biological research by ensuring that data are safely kept, and openly accessed by everyone.

Access to biological information is the major driver of biological research. Neglecting to maintain and continuously upgrade the infrastructure that provides this basic service would be detrimental to European biological research and bioscience-related business, impacting directly and negatively on the prosperity of European society. It is now time to turn this risk into an opportunity.
2.1 The preparatory phase

The proposed model for ELIXIR described in this business case has been derived as the result of multiple consultations and project work during ELIXIR’s preparatory phase (2008–2011). This phase was funded by the European Commission (Framework Programme 7 under “Capacities”) and was carried out by representatives from 32 participating organisations originating from 13 countries as well as EMBL-EBI. The work was undertaken in work packages whose reports are available on the ELIXIR website (www.elixir-europe.org).

The preparatory phase looked into the needs of bioinformatics users, the best strategy for providing data resources, issues of data integration and interoperability with other ESFRI BMS research infrastructures, training strategy and infrastructure for tools integration. It produced feasibility studies for archiving and aggregating images associated with cell-based assays, supercomputing facilities for genomic data, and European resources for systems biology. It investigated interdisciplinary interactions between biological information and medical/health and nutritional information, as well as with chemical, plant, agricultural and environmental research. The legal and organisational model, as well as the most suitable funding strategy for ELIXIR, were also investigated and defined.

The proposed model has taken into account stakeholder recommendations elicited and analysed during project work. These have been set in the international context of bioinformatics research and services, and in the context of the scientific and technological challenges that face European bioinformatics.

Recommendations from user communities were of vital importance in configuring the proposed model for ELIXIR, because the ultimate indicator for ELIXIR’s success will be its acceptance and use by researchers in academia and industry.

Users throughout Europe, including industry (which accounts for approximately 20% of public access to the EMBL-EBI’s data resources), clearly expressed a substantial need for an extended infrastructure to provide and maintain databases, especially those with primary data originating from Europe.

Data resources are the foundation of that infrastructure. A survey identified about 500 existing biological databases in Europe and has collected detailed information from over 200 databases hosted by about 100 institutions. The palette of available databases ranges from large projects for major core data sets, with upwards of 20 staff members, to specialist collections that are the part-time effort of individual researchers. The total European effort currently involves at least 350 staff Europe-wide with reported annual direct costs of about €30 million. However, even conservative estimates would put the total expenditure (including indirect costs) at more than €50 million per annum. The EMBL-EBI’s user community alone accessed its website from approximately 3 million IP addresses in 2010, and is generating some 150 million web hits a month.
The preparatory phase demonstrated an urgent need for a pan-European infrastructure for biological data, and clearly indicated the organisational and operational model for such an infrastructure: ELIXIR should comprise a central hub, which would act as the coordination point for several nodes distributed throughout Europe.

2.2 **Mission statement and core objectives**

ELIXIR’s mission is *“to construct and operate a sustainable infrastructure for biological information in Europe to support life science research and its translation to medicine and the environment, the bio-industries and society.”* 

In order to achieve its mission ELIXIR will construct, operate and enhance the following components of its infrastructure in accordance with the requirements of the community and the direction of the ELIXIR Board.

- The ELIXIR Hub and connected ELIXIR Nodes infrastructure
- The ELIXIR data infrastructure
- The ELIXIR compute infrastructure
- The ELIXIR tools and standards infrastructure
- The ELIXIR infrastructure to support the other ESFRI Biomolecular science infrastructures
- The ELIXIR infrastructure for training
- The ELIXIR international collaboration infrastructure
2.3 Topology

ELIXIR will be organised as a distributed topology comprising a single hub in a new building envisaged to be constructed on the Wellcome Trust Genome Campus at Hinxton, UK. ELIXIR Nodes will be located at centres of excellence in the ELIXIR Member States.

2.3.1 The ELIXIR Hub

The ELIXIR Hub will be responsible for the following:

Hosting the ELIXIR Executive Management and Secretariat

The ELIXIR Hub will host the ELIXIR Executive Management and Secretariat, who will report to the ELIXIR Board for delivery of ELIXIR services (see section 2.4 for more information on ELIXIR’s governance structure). The Executive Management will initially comprise the Chief Executive Officer supported by a small secretariat with administrative, financial and legal expertise. They will be responsible for: coordinating and leading the infrastructure; ensuring that its service provision is of the highest quality; liaising with ELIXIR Member States; implementing bilateral service-level agreements with the ELIXIR Nodes; arranging meetings of the ELIXIR Board and the Scientific Advisory Board; planning and securing ELIXIR’s financial sustainability, including applications for pan-European funding; and representing Europe in global bioinformatics initiatives.
Data delivery from the European Data Centre

The ELIXIR Hub will include components of EMBL-EBI services that will be funded by ELIXIR. The most significant of these is the delivery of the ELIXIR core data resources to European life scientists with very high levels of availability and robustness through the European Data Centre in London. EMBL-EBI will manage the European Data Centre, ensuring maximum accessibility to data for the benefit of ELIXIR. This includes 24/7 delivery with archiving and replication to ensure no loss of data and continuity of service during failures. EMBL-EBI will continue with its current activities in data collection, curation and database production. ELIXIR will complement these activities in the European Data centre by focusing on data delivery and also coordination of data nodes and integration around Europe.

Technical coordination activities

The ELIXIR Hub will provide oversight of biological data throughout Europe, by creating data registries and coordinating data standards. Both of these require involvement of relevant domain specialists, often in a global context, to ensure adoption by the scientific community. The Hub will help to coordinate these activities across Europe. The major focus at EMBL-EBI is currently on biomolecular data and this will remain the core activity for the Hub, but as new data types emerge, ELIXIR will need to consider its role and evolve according to community needs.

In the ELIXIR model, a distributed network of data resources is envisaged, with ELIXIR Nodes acting as data centres linked together by the ELIXIR Hub. This ambitious goal will rely on state-of-the-art technical developments and alignment of standards and annotations. The ELIXIR Hub will also provide a coordination function across ELIXIR Nodes, to minimise duplication and waste and to ensure the provision of a complete range of top quality services.

Establishing the Compute infrastructure

The ELIXIR Hub will seek to enable the best compute infrastructure for life scientists throughout Europe. As data sets grow in size, the need to co-locate data and compute will increase. Therefore, it is inevitable that the large data Nodes will also become large compute Nodes and vice versa. ELIXIR will seek to establish the best solutions and ensure good interconnectivity, building where possible on the e-infrastructures already developed in Europe.

Building bridges with other BMS Infrastructures

The ELIXIR Hub will spearhead interactions with the other ESFRI BMS infrastructures, enabling ELIXIR members to work with these infrastructures and helping to coordinate the e-infrastructure for data that they will require. ELIXIR already provides such an infrastructure for some aspects of BMS data, such as for structural data through the PDBe (INSTRUCT) or omics related data for translational research (EATRIS) and biobanks (BBMRI). Building bridges between the different types of data, such as clinical trials and biobank data, will be essential for future research and for addressing the Grand Challenges. ELIXIR, together with the other 12 BMS Infrastructures, is currently exploring possibilities for interoperability and abstraction of data between the infrastructures.
Coordination of user training

The ELIXIR Hub will coordinate the provision of training activities for ELIXIR users across Europe. The need for training to ensure optimal usage of the data and associated tools is paramount. Experts based at the Hub and Nodes, working closely with the data and tool developers, will provide training. This will be delivered either electronically or through face-to-face courses held throughout Europe. The ELIXIR Hub will help to ensure the provision of appropriate training both by providing support for and coordination of suitable trainers in the Nodes, and by supplying an infrastructure for the management of training materials, so that they can be used and re-used by all of ELIXIR’s trainers.

2.3.2 The ELIXIR Nodes: criteria and networks

The ELIXIR Hub will connect with ELIXIR Nodes at existing centres of excellence in the ELIXIR Member States, which will retain their independent legal personalities. Each Node will take responsibility for providing a defined set of scientific and technical activities that are in line with the strategic priorities of the host country but are accessible to researchers throughout Europe. Contributions could, for example, comprise biomolecular data resources, bio-compute facilities, software tools, standards or bioinformatics training. In due course, we envisage that this will lead to the emergence of nodes that offer core services that are an essential part of the ELIXIR European-wide infrastructure.

There are a number of criteria that are likely to apply to all ELIXIR Nodes, including the following:

- Nodes will be located in a Member State of ELIXIR;
- Nodes will be, or will be represented by, a legal entity that is able to enter into formal agreement with the ELIXIR Hub regarding such matters as service level agreements;
- Nodes will need to have the capability to provide one or more, internationally competitive services at the European level. An institute that only provides services at a National level would not normally be considered for inclusion in ELIXIR. Bearing in mind the stringency of this criterion, it is anticipated that most Node applications will be from pre-existing institutes with a significant history of robust service provision. Nevertheless, ELIXIR will welcome applications from new and innovative Nodes whose management can demonstrate that they can meet this requirement;
- Nodes will need to be able to demonstrate sustainability of funding. This is a very important criterion because of the importance of ELIXIR services to the European research community.

ELIXIR Nodes that perform similar functions will work closely together and form ELIXIR service networks. For example, Nodes that provide data resources will collaborate to form a data network; Nodes that provide training will collaborate and form a training network, etc. (see Figure 4).

A number of countries have indicated a desire to form National Coordination Nodes (NCNs), suggesting that there will be one contact point in the country for all ELIXIR activities. There seem to be two main motivations for this; the first occurs in countries where bioinformatics is very well developed and where there are many players. The idea is that an NCN will simplify interactions with ELIXIR and allow local funders to exercise a degree of internal coordination as a means to achieving national strategic priorities. The second motivation occurs in countries where bioinformatics activities are still at an early stage.
of development and where there is a desire to develop quickly and in a well-coordinated manner. NCNs will have to fulfill the coordination activities whilst providing a service at the European level.

2.3.3 Services delivered by ELIXIR Nodes

ELIXIR Nodes will provide some or all of the following services:

**Data resources**

Nodes may provide data services similar to those provided by the Hub. These may be core data resources that are better located near a facility that has high levels of appropriate domain expertise, or they may be specialist data resources that are only needed by a small part of the community but are nonetheless of exceptional scientific importance. In the future, as the amount of data grows, some of the heavily used services might need to be duplicated for performance or **load-balancing** reasons.
**Compute provision**

Nodes based at major computer centres will provide large-scale computational resources for the more demanding services.

**Training**

Nodes that have the appropriate training expertise and facilities will provide: training throughout Europe on data resources provided by the node; application expertise; and use cases based on the node’s specialty; and/or eLearning expertise and infrastructure.

It is likely that many Nodes will be involved in providing training, specialising in their particular area of expertise and combining forces to deliver a comprehensive suite of training tools and courses.

**Tools Infrastructure**

A robust infrastructure that supports best practice in tool deployment and use is a key component of ELIXIR. Tool Infrastructure Nodes will need to provide capability for users that supports tool discovery and benchmarking, and/or capability for developers that promotes ease-of-use and interoperability in software development. There is some overlap between this and the kinds of capability that will be provided by Standards Nodes. It is thus likely that there will be at some combined Standards and Tools infrastructure Nodes.

**Standards Infrastructure**

A Node offering standards infrastructure should provide capability to address standardisation activities in one or more of the following: programmatic access to databases and tools; biological and medical nomenclature; controlled vocabularies and ontologies; and/or reporting requirements for data deposition and exchange.

### 2.4 Governance model

The governance model depicted in Figure 5 takes into consideration ELIXIR’s scientific and technical structure. The ELIXIR Board will oversee ELIXIR’s activities. The ELIXIR Member States and EMBL will appoint the ELIXIR Board.

The ELIXIR Board will appoint the staff at the ELIXIR Executive Management and Secretariat, which will be located at the ELIXIR Hub. The Scientific Advisory Board (SAB) will advise the ELIXIR Board and Executive Management in scientific matters, including the selection of ELIXIR Nodes. The scientific and technical aspects of the ELIXIR Hub and Nodes will be regularly reviewed by the ELIXIR SAB.

#### 2.4.1 ELIXIR Board

The ELIXIR Board will be composed of two representatives from each ELIXIR Member State, preferably a scientific and an administrative delegate. Each country will have one vote. Members will have due authority to deliberate, negotiate and decide issues relating to budget and to ELIXIR’s strategic programme, as well as establishing and monitoring rules and procedures, including those for the selection of ELIXIR Nodes. They will be expected to report back to their own Member States, and will draw on advice from the SAB.
The ELIXIR Board may decide to admit observers, for example from charities and industry.

**Roles and responsibilities**

As the most senior body of ELIXIR, the ELIXIR Board is expected to give high-level oversight, and approval of devolved ELIXIR strategy and budget (including managing additional bilateral agreements). It will oversee management of risks and liabilities of ELIXIR, and resolve disputes and disagreements. The Board will provide Member States with voting rights in order to:

- Agree and monitor the strategic plan;
- Adopt budget;
- Commission and receive operational reports;
- Approve agreements;
- Appoint (and dismiss) the Executive Management;

![ELIXIR's governance structure](image)
• Agree new Nodes and terminate nodes that are no longer serving ELIXIR’s purpose;
• Appoint members of the Scientific Advisory Board;

A more detailed list of tasks and responsibilities will be defined in the International Consortium Agreement and the ELIXIR Statutes.

2.4.2 Scientific Advisory Board

For ELIXIR to understand and assess its position and role within the wider international body of research activity, it must be able to draw coherently on the independent views and knowledge of the wider research community. Membership of the SAB will be made up of distinguished international experts including academics not associated with ELIXIR Nodes, representatives of other internationally renowned organisations operating outside of Europe, and representatives of researchers in the commercial sector. This group of no more than 20 members will be appointed by the ELIXIR Board, following the proposal of candidates by the ELIXIR Executive Management.

Roles and responsibilities
The SAB will be expected to provide independent consideration and advice to the ELIXIR Board and ELIXIR Executive Management on scientific issues in order to:
• Review applications for new Nodes and make recommendations to the ELIXIR Board;
• Ensure scientific and technical excellence and relevance (including independent quality assurance);
• Identify and recommend emerging challenges and opportunities, both within and beyond ELIXIR activities (including specific periodic reviews);
• Carry out periodic scientific reviews of elements of both the ELIXIR Hub and the Nodes, and functional activities (i.e. training, data storage etc.).

The SAB will meet regularly and report directly to the ELIXIR Board, and will act independently of EMBL.

2.4.3 ELIXIR Executive Management and Secretariat

The ELIXIR Board will appoint the ELIXIR Executive Management for a set period with scope for renewal. The Secretariat will be established to assist and facilitate the activities of the ELIXIR Executive Management. Being part of the ELIXIR Hub, the Executive Management and Secretariat will be located at Hinxton, UK; its staff will be employed by EMBL-EBI.

Roles and responsibilities
The Executive Management, assisted by the Secretariat Staff, will be responsible for:
Day-to-day operational (management, financial and administrative) activities such as:

- Operational responsibility for delivery of strategy and services;
- Coordination of ELIXIR Nodes and Hub
- Supporting ELIXIR standing committees;
- Maintaining accounts, auditing etc.;
- Coordination of ELIXIR publicity and communications.

- Management and administration of ELIXIR in accordance with the decisions of the ELIXIR Board;
- Convening meetings of technical representatives from the Nodes as required;
- Presenting the annual report to the ELIXIR Board;
- Preparing the strategic programme;
- Submitting the annual budget estimates.

Appointment and termination of employment would be in accordance with EMBL staff rules and regulations.

2.5 Legal model

During the preparatory phase project, a large number of possible legal models were evaluated for ELIXIR using expert legal advice from internationally operating law firms. One of the main questions addressed was whether ELIXIR should be set up as a new organisation or by using an existing organisation. The recommendation was to establish ELIXIR, at least initially, under the umbrella of EMBL with the option to set up an independent new legal entity, possibly as a European Research Infrastructure Consortium (ERIC), later.

The goal is to establish a functional distributed infrastructure with the ELIXIR Hub at EMBL-EBI and the ELIXIR Nodes at existing research organisations in the ELIXIR Member States, which will retain their independent legal personalities. The main difference between ELIXIR and other distributed research infrastructures is the large size of the ELIXIR Hub. This is necessary because, in addition to having a coordinating role, the ELIXIR Hub will continue to provide core data resources and house the European Data Centre. This will ensure the long-term stability to ELIXIR.

ELIXIR will initially be established using the mechanism of an "EMBL Special Project". ELIXIR will thereby be able to profit from EMBL's established privileges and legal personality. The details will be defined in an International Consortium agreement (ICA), which will be agreed by the ELIXIR Member States and EMBL. All ESFRI Member States are eligible to become members of ELIXIR. Because the negotiations of the ICA are likely to take some time, first a non-binding MoU will be agreed with the countries that are ready to enter into negotiations in early 2011. It will also be possible for additional bilateral agreements, for example with industry and charitable funding bodies, to be established by ELIXIR.

The ICA will clarify the details of the "Special Project" and in particular how ELIXIR will be established. The ICA will also clarify the role of EMBL-EBI within ELIXIR, as well as the rights and duties arising from ELIXIR’s role to support other ELIXIR Member States. This also includes EMBL-EBI’s rights and obligations with regards to the ELIXIR Hub.
Under the mechanism of an “EMBL Special Project”, ELIXIR’s budget and activities will be ring-fenced from EMBL with separate accounts, auditing, clear reporting lines and stand-alone communication activities to ensure visibility. EMBL Council will oversee EMBL’s participation in ELIXIR but will not oversee ELIXIR itself. ELIXIR will have its own governing body, the ELIXIR Board.

The ELIXIR Hub and Nodes will all have separate legal identities. Bilateral service level agreements between the Hub and each Node will deal with the relationship between EMBL, representing the ELIXIR Hub, and the ELIXIR Nodes. For legal reasons the agreements will have to be concluded between EMBL representing the ELIXIR hub and the organizations that are hosting the ELIXIR nodes.

These agreements will define their mutual responsibilities and obligations within ELIXIR. These could, for example, include the goals; duration; administrative issues; monitoring and reviews; representation and reporting structures; and mechanisms for termination.

The ELIXIR Hub will not have a service level agreement with EMBL-EBI because legally it is part of EMBL-EBI and it is impossible to have an agreement with oneself; EMBL-EBI’s rights and obligations – including the tasks of the ELIXIR Hub – will be defined in the ICA, which is the only agreement that deals with the ELIXIR members’ rights and obligations.

The ELIXIR Hub will be responsible for the overall technical coordination of ELIXIR activities. The ELIXIR Nodes will be responsible for ensuring that they comply with applicable local, national and international laws, regulations and guidelines (including health and safety, data protection, good scientific and clinical practice).

2.6 Interim steps

Initially, ELIXIR will go through an interim phase for the construction of the research infrastructure before entering into a permanent phase. A MoU will be signed by countries willing to participate in the interim phase. The ELIXIR Statutes, which will be part of the ICA, will describe the procedures within ELIXIR in detail once the permanent phase has been reached. The following is a brief overview of the most important steps in the interim phase, before signing the ICA.

Construction

The construction phase will start once five countries and EMBL have signed the MoU, which is limited in duration to 24 months. During this time, any other country interested in participating can join by signing the MoU. The goal of this interim phase is to negotiate the ELIXIR ICA and statutes that will be signed by the ELIXIR Member States and EMBL as soon as it is ready. An interim ELIXIR Board will be established with Member-State representatives. These representatives will appoint a Scientific Advisory Board to provide scientific advice and help with the selection of the ELIXIR Nodes.

Evaluation and initial selection of ELIXIR Nodes

A SAB will be established by the interim ELIXIR Board. This will be similar in size and composition to the SAB for the permanent phase of ELIXIR and will be involved in the initial selection of ELIXIR Nodes.
New Member States

The admission of new Member States to ELIXIR will be a rolling process that allows any country to join if the ELIXIR Member States unanimously agree.

The procedures for the ELIXIR Permanent phase will be negotiated and decided during the ELIXIR Construction phase and will be included in the ICA and ELIXIR Statutes.

2.7 Feasibility of the proposed solution

The proposed legal and organisational solution is feasible for this infrastructure for the following reasons:

Experience in international collaborative projects. National centres of excellence as well as EMBL-EBI are organisations with many years of experience in coordinating and participating in international scientific collaborative projects. This experience at the scientific level, as well as the administrative and operational levels, will be applied in successfully setting up and efficiently operating ELIXIR by making use of existing resources at the Hub and the Nodes and spreading investments at international and national levels.

Service continuity. ELIXIR will be seamlessly built on existing data resources and services. Its construction will not affect the level of current service provision but will deal effectively with the challenges that a distributed infrastructure requires.

Timeliness. Both the ELIXIR Hub and candidate national Nodes have the means to collaborate effectively on the technical and scientific coordination issues that ELIXIR will require. This will have a positive effect on the timely delivery of the ELIXIR construction phase, with a direct and beneficial impact on restraining budgets and mitigating associated risks.

Scalability. The proposed model for ELIXIR is scalable in many dimensions. New ELIXIR Nodes can be added as the infrastructure grows into the future, as well as additional services, data resources and tools. New technologies will be better evaluated and assimilated by the participating parties, thus guaranteeing cost-effective implementation and migration solutions across the infrastructure.

2.8 The role of industry

The private sector currently represents approximately 20% of public usage of data resources (based on EBI statistics), and will of course continue to have open access to all the data where appropriate. Industry has participated in ELIXIR and in developing this Business Case. However, closer collaboration would be of benefit to all. Currently, the bio-industries are facing exactly the same data challenges as academia. They see some major risks in their current way of interacting with publicly funded data providers. These include budget and manpower challenges, duplication of effort, difficulty integrating private and public data and lack of necessary standards. They also see the added value of ELIXIR in terms of reduced costs and duplication of effort, higher scalability, access to common data and interface standards and much better public–private data integration.
Their key requirements from ELIXIR are a service-oriented delivery organisation, unrestricted resource distribution, agreed levels of service, secure access to protect IP and common standards.

There are several possible models, which would allow close collaboration and some cost sharing. At one extreme there could be an Industry Node – funded by a consortium of interested companies, which seeks to obtain maximum benefit for industry from ELIXIR. This could include consideration of how services might be made more useful for commercial users; how security can be addressed; and support or development of specific resources or standards to address priorities in industry. At the other extreme, the private sector could interact through targeted bilateral agreements with the Hub or with individual Nodes. Other modes of interaction will also be considered. These interactions should involve not only large companies in the life science sector (including pharmaceutical, consumer goods, healthcare, agriculture and biotechnology companies) but also SMEs, which will require a different model of engagement.

Industry-specific services may include:

- Industry-sponsored customised training
- Industry-sponsored customised development of data resources or tools
- Pre-competitive projects
- Industry-sponsored development of relevant standards
- Private industry servers deployed in and supported by the ELIXIR Data Centre
- Industrial provision of cloud-based server access to copies of public data.
3 Funding ELIXIR

3.1 Investing in a research infrastructure for biological information

The current high standards of living in Europe have been realised through public and private investment in the creation of infrastructures: rail, road, shipping and air transport are among the most prominent. It is a well-documented fact that investment in infrastructures spurs economic growth. More often than not, infrastructures act by their own virtue as catalysts for new ideas, new products and new services, hitherto unimagined. They are the safest way to invest in the medium and long-term future of a society.

A new type of research infrastructure is now needed if the riches held in research communities (data, access to patient populations, knowledge and skills or other wealth) are to be exploited to meet the Grand Challenges and for the benefit of Europe and the world.

ELIXIR aims to create a sustainable infrastructure for biological information in Europe. This will be for scientists working in industry and academia and will support life-sciences research in medicine, agriculture and the environment. It will be the bedrock of other life science research infrastructures including biobanks, clinical trials and marine biology.

ELIXIR offers essential services to the scientific community in the life sciences. These need to be expanded and maintained to make best use of previous and future large investments into research. These services cost only a fraction of the initial investment in generating the data, some of which is unique and could never be regenerated.

European countries have struggled to develop a mechanism that would allow the main biological data resources to be provided in a sustainable manner. EMBL-EBI has been financed by the 20 EMBL Member States, which have been very supportive in increasing the EMBL budget to enable the EMBL-EBI to respond to the data deluge and maintain Europe’s competitive position in this area. It has also been necessary to raise additional funds for EMBL-EBI to manage the data, with major contributions from the European Union, the US National Institutes of Health, the Wellcome Trust and the UK Research Councils.

ELIXIR provides the opportunity to develop a longer-term plan. It will enable the load to be distributed across several sites in Europe by setting up a distributed research infrastructure linking EMBL-EBI with other centres of excellence in biological data provision in the ESFRI Member States. This has specific implications for the funding of ELIXIR: funding will have to come from a number of sources at European, international and national levels. This is positive because ELIXIR’s distributed nature provides funding opportunities that a single-site operation may not have access to, but it is also a major challenge, for example due to the lack of synchronisation between the different sources of income, and uncertainty created by heterogeneous funding sources.

---

ELIXIR’s Member States will benefit from:

- Enhanced exploitation of data being produced from their investments in biomedical research, especially genomics and other high-throughput technologies;
- Optimisation of the value of data by rapid and timely translation into other sectors such as clinical, agricultural and environment research;
- Enhanced competitiveness through European coordination;
- Enhanced bioinformatics capabilities and expertise within their funded communities;
- Access to expertise and advice beyond national boundaries to structure their own national infrastructure for biological information.

Towards the end of the ELIXIR preparatory phase, several countries have already made funds available: the UK has provided 10 million GBP towards the ELIXIR Data Centre, which is already operational. Finland, Sweden, Denmark and Spain have made funding available to potential ELIXIR Nodes and funding has been applied for in several other countries.

3.2 Funding model

ELIXIR will be implemented using the mechanism of an EMBL “Special Project”. Figure 6 shows the streams of funding that go to the Hub and the Nodes and the potential to leverage this funding to jointly apply for additional funds from other sources, including European Framework programme funds.

The ELIXIR Hub will be located at EMBL-EBI and will make use of EMBL’s legal personality to make ELIXIR operational at the earliest opportunity. The ELIXIR budget will be ring-fenced from other EMBL activities, and the ELIXIR Executive Management will report to the ELIXIR Board on the use of the budget.

The ELIXIR budget will consist of two different categories. The ELIXIR international consortium will provide funding for the ELIXIR Hub, whereas national funding organisations and others will provide funding for the ELIXIR Nodes.

The ELIXIR international consortium will consist of those EMBL Member States that have agreed to participate in the EMBL Special Project, other ESFRI Member States, and EMBL.

Funding for ELIXIR can be split into funding the construction and funding the operations of the infrastructure.

3.3 Funding the construction

- Construction of Hub

Because the ELIXIR Hub is located in the UK, funding for the construction of a building for the ELIXIR Hub, to renew and extend the ELIXIR Data Centre as the need for more storage grows, and to purchase some of the equipment needed to cope with the data deluge has been requested from the UK Large Facilities Capital Fund. This
is in the order of 75 million GBP. Funds have been earmarked for ELIXIR, pending the UK Government’s approval of the business plan for the Technical Hub. The Wellcome Trust, which has provided significant financial support to the EMBL-EBI in the past, has already signalled its support for the current phase of development by providing the land needed for the ELIXIR Hub.

- **Construction of the Nodes**
  The provision of pan-European services may require expanding existing facilities in the ELIXIR Member States, and funding may have to be raised from local and other sources, including EU Structural Funds. Several countries have already committed funds to prepare for the construction of ELIXIR nodes, including Finland (€6.85 M), Sweden (€1.7 M), Denmark (€5 M) and Spain (€1.7 M).

### 3.4 Funding the operation

- **Funding the operation of the ELIXIR Hub**
  The ELIXIR Member States will fund the ELIXIR Hub. States that are also members of EMBL will be asked to fund ELIXIR as part of the EMBL Indicative Scheme 2012–2016,
as an EMBL Special Project. A decision is expected from EMBL Council in November 2011, and a detailed budget is in preparation. ELIXIR Member States that are not members of EMBL will also be expected to contribute to the budget of the ELIXIR Hub.

• **Funding for the operation of ELIXIR Nodes**

ELIXIR Nodes will be funded at the national level by government agencies, charities and other funders. ELIXIR Nodes are expected to apply for their own funding in coordination with ELIXIR, and proof of sustainable funding will be required for a Node to become part of ELIXIR.

• **Funding from the European Union Framework Programme**

Funding from the European Union Framework Programme will be sought to support some aspects of the operation of ELIXIR in addition to its core activities. ELIXIR will jointly apply to the European Union, and other international funders, for additional funding for projects that involve building e-infrastructure, enhancing user access, training, collaborations with industry (for example, through the Innovative Medicines Initiative) and building bridges between ELIXIR and the BMS research infrastructures.

### 3.5 Estimated budget

ELIXIR will be constructed gradually starting with the Hub and a few Nodes and adding more Nodes over time. Currently, we anticipate that the hub will be fully operational after five years. For the purposes of providing an indicative budget, we estimate that the member states will wish to create approximately 20 ELIXIR Nodes over the same time period. Over the following five-year period, some of these Nodes will expand, in line with expected data growth, and more will be added to meet the demands for handling new types of biological and biomedical data in Europe.

ELIXIR will require a total investment for construction of several hundred million Euros and is expected to have over 70 Mio € operating costs per year at the end of the first five-year period and more than 350 staff employed at the ELIXIR nodes in total.

Capital construction costs will include the costs of staff accommodation (mainly incurred at the start of construction), plus the larger and arguably more important computing component. The costs for the storage and compute for ELIXIR are large and it is also important to note that IT equipment needs to be renewed every three years, requiring an annual capital budget. In addition to this, the IT budget will have to grow as the data increases and demand expands. The infrastructure will therefore need an initial upfront capital investment for construction and IT costs, plus a yearly computing budget to allow regular replacement and upgrading.

Operational costs for the infrastructure include costs for staff, networking and also the cost of electricity for computing. By the end of the first five-year period, we plan that the ELIXIR hub will be fully operational with approximately 100 staff. By this stage we expect 20 ELIXIR Nodes will have been established in the member states with more than 200 staff in total. The staff will be responsible for maintaining and running the infrastructure at each centre. The Nodes will vary in size from a few scientists to 25 or more staff.
3.5.1 ELIXIR Hub

Construction of the ELIXIR Hub

The capital construction costs for the hub include the cost of the European Data Centre, based in London, the IT equipment within it, and the technical hub at Hinxton. The Technical Hub will provide facilities to house up to 100 ELIXIR staff, a purpose-built training suite and an Industry and Innovation suite.

The UK government has already made available 10 million GBP in 2009 towards the construction of the ELIXIR Hub. This has been used to purchase a five-year lease for the European Data Centre and to provide initial computing equipment for it. In addition, significant funding (74 Mio GBP) has been requested by EMBL-EBI from the UK Large Facilities Capital Fund for the construction of the ELIXIR Technical Hub and for extension of the European Data Centre lease, as well as for enhancing and upgrading the storage and compute equipment over the next ten years. This ELIXIR equipment will contribute to the sustained production and delivery of the core data resources from EMBL-EBI. These funds have been earmarked subject to approval of the business plan by the UK Government.

Operation of the hub

The operational costs at the hub include staff costs, IT networking, electricity costs and consumables. The staff at the hub comprises all ELIXIR staff needed to operate the hub, as described in Section 2.3.1. This includes the ELIXIR Executive Management and Secretariat, who will coordinate the operation of the whole infrastructure. In addition, the provision of data to the European user community will require expert technical staff to operate the new European Data Centre and to manage the delivery of world-class data services. These services will include delivery of all the core data resources currently hosted by the EMBL-EBI and emerging services related to literature, medical informatics, agriculture, biodiversity, chemistry and imaging. Critically the hub will handle the interactions with the Nodes and the other BMS infrastructures, ensuring a coordinated infrastructure with seamless access to data and tools around Europe. This ambitious objective will require coordination of standards and tools with the Nodes and other infrastructures to improve the integration and access to data. In addition, the hub will coordinate training, outreach and industry relations, all in close cooperation with the ELIXIR Nodes. Total staff would increase gradually from ten in 2012 to approximately 100 in 2016 under this scenario.

Thus the operational staff budget of the ELIXIR Hub will rise to 12 million Euros per annum by 2016, with a total cost from 2012 to 2016 of 28 million Euros.

3.5.2 ELIXIR Nodes

The Nodes will have very diverse functions and their size will largely depend on what they do. Section 2.3.3 describes some of the services that could be delivered by ELIXIR Nodes in more detail; typical tasks include the provision of data resources, the provision of compute and storage resources, an infrastructure
for tool integration and the development of standards. However, there will also be other more specialised contributions of individual Nodes, related to specific areas of biological expertise. It is expected that most Nodes will provide or host training courses to transfer know-how and ensure maximal and optimal usage of the infrastructure.

**Construction of ELIXIR Nodes**
The cost of the construction of a Node will depend on many factors and local circumstances, and it is impossible here to make precise estimates for the cost of these Nodes. The organisations that have submitted expressions of interest to establish ELIXIR Nodes will work out the details of their local budget requirements. The actual costs will also depend on what type of infrastructure is already available at the host organisation and what would have to be added to establish a Node. It is very difficult at this point in time to estimate the investment required but it could range from a few million for a small, specialised Node to tens of millions of Euros, for example for a large data centre.

**Operation of ELIXIR Nodes**
The costs of the operation will also depend on the Nodes’ function and size and will range from several hundred thousand Euros per year up to 10 million Euros per year for a very large Node with more than 30 staff and high IT costs, for example for a compute-intensive Node. ELIXIR will only function if coordination between the Hub and the Nodes is ensured, which will require staff both at both ends.
3.5.3 Summary of projected ELIXIR budget

**ELIXIR Hub projected budget**

<table>
<thead>
<tr>
<th>Funder</th>
<th>Technical Hub</th>
<th>European Data Centre</th>
<th>IT equipment</th>
<th>Total Capital</th>
</tr>
</thead>
<tbody>
<tr>
<td>UK LFCF</td>
<td>23,000</td>
<td>10,000</td>
<td>41,000</td>
<td>74,000</td>
</tr>
</tbody>
</table>

*(all figures in 1000’s GBP)*

<table>
<thead>
<tr>
<th>Funder</th>
<th>Operating Costs</th>
<th>Total Operating costs 2012-2016</th>
</tr>
</thead>
<tbody>
<tr>
<td>ELIXIR Hub</td>
<td></td>
<td></td>
</tr>
<tr>
<td>and others</td>
<td>1.000</td>
<td>1.900</td>
</tr>
<tr>
<td></td>
<td>4,900</td>
<td>8.400</td>
</tr>
<tr>
<td></td>
<td>12,300</td>
<td>28,500</td>
</tr>
</tbody>
</table>

*(all figures in 1000’s Euro)*

<table>
<thead>
<tr>
<th>ELIXIR Hub staff</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
<th>2015</th>
<th>2016</th>
</tr>
</thead>
<tbody>
<tr>
<td>ELIXIR MS and others</td>
<td>9</td>
<td>17</td>
<td>41</td>
<td>70</td>
<td>100</td>
</tr>
</tbody>
</table>

*(FTE)*

**ELIXIR Nodes projected budget**

Resources required per node

<table>
<thead>
<tr>
<th>Staff (FTE)</th>
<th>Facilities (sqm)</th>
<th>Storage (Terabyte)</th>
<th>Compute (Cores)</th>
</tr>
</thead>
<tbody>
<tr>
<td>small node</td>
<td>10</td>
<td>150</td>
<td>10</td>
</tr>
<tr>
<td>medium node</td>
<td>20</td>
<td>500</td>
<td>100</td>
</tr>
<tr>
<td>large node</td>
<td>30</td>
<td>1200</td>
<td>10000</td>
</tr>
</tbody>
</table>

Construction of nodes

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>small nodes</td>
<td>3</td>
<td>5</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>medium nodes</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>large nodes</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Total number of nodes</td>
<td>5</td>
<td>8</td>
<td>14</td>
<td>17</td>
</tr>
</tbody>
</table>

*(No of nodes)*

The exact costs of the ELIXIR Nodes will have to be determined by the Node coordinators, their host institutions and national and other funding organisations. Here we can only provide an indication of the resources required for the construction of the ELIXIR Nodes. Construction costs will depend on whether existing facilities can be used or whether new ones have to be constructed and on local prices. IT equipment will have to be upgraded every 3-5 years. This represents a recurring capital investment and the amount depends on the size of the compute facilities and local prices. Operating costs will depend on local salaries and local prices of electricity.
EMBL Council has taken note of the draft ELIXIR MoU, and ELIXIR’s Steering Committee has approved it along with this Business Case.

Both of these documents are currently being communicated to governments in ESFRI Member States. The ELIXIR preparatory phase will continue until the end of December 2011. During this time frame, the following actions will take place:

<table>
<thead>
<tr>
<th>Target</th>
<th>Objective / Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summer 2011</td>
<td>• MoU becomes effective once 5 countries and EMBL have signed.</td>
</tr>
<tr>
<td>Autumn 2011</td>
<td>• Interim ELIXIR Board constitutional meeting; • Procedure for setting up a Scientific Advisory Board approved; • Negotiations of International Consortium Agreement (ICA) begin.</td>
</tr>
<tr>
<td>November 2011</td>
<td>• Decision on EMBL Indicative Scheme 2012–2016 expected.</td>
</tr>
<tr>
<td>December 2011</td>
<td>• End of ELIXIR’s EU-funded FP7 preparatory phase project.</td>
</tr>
<tr>
<td>January 2012</td>
<td>• Scientific Advisory Board established; review of ELIXIR Nodes begins</td>
</tr>
<tr>
<td>2012/2013</td>
<td>• Draft ICA finalised and approved by signatories; • Negotiation of bilateral service-level agreements with the institutions operating initial Nodes. This will be based on scientific and technical criteria after obtaining funding support for ELIXIR-related activities from national funding organisations (once ELIXIR has been formally established); • Additional countries will sign the ICA and establish Nodes as needed in the future.</td>
</tr>
</tbody>
</table>
This Business Case will be submitted to the ESFRI Member States together with a non-binding MoU. The latter will set up an interim structure for ELIXIR, including an interim ELIXIR Board that will take part in the negotiation of the ICA. The MoU will also set up a SAB to initiate the review of ELIXIR Nodes.

Already a considerable number of European research institutes, from 23 different European countries, have expressed an interest in becoming ELIXIR Nodes (see Appendix I).

This interim phase (see also section 2.6) will end with the signing of the ICA that will establish ELIXIR’s permanent structure.

During 2011/12, ongoing consultation with potential ELIXIR Member States and with centres of excellence that have expressed an interest in becoming ELIXIR Nodes will be crucial to shape the initial direction of ELIXIR, and to ensure a level of coordination that will both enhance European infrastructure for biological information and add value to the research priorities of those Member States that wish to be involved.

In the course of the next 18–24 months it is anticipated that the initial ICA will be finalised and agreed by those Member States wishing to become involved at the outset. In addition, principles for associated SLAs will be developed, prior to the agreement of specific SLAs with those first Nodes wishing to found ELIXIR.
In addition, it is important to continue consultation with these organisations, and with other potential ELIXIR Nodes, their associated Member States (including relevant funding agencies) and the European Commission to continue to map out how ELIXIR will develop to continue to meet current and emerging needs, and to make the most of new opportunities.

During this time it is expected that those interested in establishing ELIXIR Nodes will continue to discuss with their relevant funders how to resource activities at their Node, for central operations (at the Hub), and finally to apply for resources through appropriate national funding mechanisms.

Figure 7 shows the timelines for establishing ELIXIR.
<table>
<thead>
<tr>
<th>Year</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011</td>
<td>EMBL Council approves EMBL Indicative Scheme 2012-16 and ELIXIR Special Project</td>
</tr>
<tr>
<td></td>
<td>ICA to be signed. ELIXIR starts operation</td>
</tr>
<tr>
<td>2012</td>
<td>ELIXIR Node proposals to be endorsed by ELIXIR SAB</td>
</tr>
<tr>
<td></td>
<td>Funding for ELIXIR Nodes to be approved by national funders</td>
</tr>
<tr>
<td></td>
<td>Nodes to join ELIXIR through bilateral agreement</td>
</tr>
</tbody>
</table>

Figure 7: ELIXIR Timeline chart.
## Appendix

List of Expressions of Interest for ELIXIR Nodes (as of October 2010)

<table>
<thead>
<tr>
<th>PI</th>
<th>Lead Institute</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dr Joachim Jacob</td>
<td>Flanders Institute for Biotechnology (VIB)</td>
<td>Belgium</td>
</tr>
<tr>
<td>Prof. Kristian Vlahovicek</td>
<td>Croatian Computational Biology Initiative</td>
<td>Croatia</td>
</tr>
<tr>
<td>Dr Jiri Vondrak</td>
<td>Academy of Sciences of the Czech Republic (IOCB)</td>
<td>Czech Republic</td>
</tr>
<tr>
<td>Prof. Søren Brunak</td>
<td>Technical University of Demark (DTU)</td>
<td>Denmark</td>
</tr>
<tr>
<td>Prof. Jaak Vilo</td>
<td>University of Tartu</td>
<td>Estonia</td>
</tr>
<tr>
<td>Dr (Adj.Prof) Tommi Nyrönen</td>
<td>CSC – IT Centre for Science</td>
<td>Finland</td>
</tr>
<tr>
<td>Prof. Mauno Vihinen</td>
<td>University of Tampere</td>
<td>Finland</td>
</tr>
<tr>
<td>Prof. Mark Lathrop</td>
<td>Centre National de Génotypage, Commissariat à l'énergie atomique et aux énergies alternatives (CEA) and Fondation Jean Dausset – CEPH</td>
<td>France</td>
</tr>
<tr>
<td>Dr Claudine Medigue</td>
<td>Centre National de la Recherche Scientifique (CNRS)</td>
<td>France</td>
</tr>
<tr>
<td>Jean-Louis Romette</td>
<td>University de la Méditerranée</td>
<td>France</td>
</tr>
<tr>
<td>Prof. Frank Glöckner</td>
<td>Max Planck Institute for Marine Biology, Bremen</td>
<td>Germany</td>
</tr>
<tr>
<td>Prof. Peter Lichter</td>
<td>Deutsches Krebsforschungszentrum (The German Cancer Research Center, DKFZ)</td>
<td>Germany</td>
</tr>
<tr>
<td>Dr Steffen Möller</td>
<td>Debian society</td>
<td>Germany</td>
</tr>
<tr>
<td>Dr Wolfgang Müller</td>
<td>Heidelberg Institute for Theoretical Studies (HITS)</td>
<td>Germany</td>
</tr>
<tr>
<td>Prof. Dietmar Schomburg</td>
<td>Technical university of Braunschweig (TU-BS)</td>
<td>Germany</td>
</tr>
<tr>
<td>Dr Christian Stephan</td>
<td>Medizinisches Proteom-Center, Ruhr-Universität-Bochum</td>
<td>Germany</td>
</tr>
<tr>
<td>Dr Vassilis Aidinis</td>
<td>B.S.R.C. Alexander Fleming, Greece</td>
<td>Greece</td>
</tr>
<tr>
<td>Prof Artemis Hatzigeorgiou</td>
<td>B.S.R.C. Alexander Fleming, Institute of Molecular Oncology</td>
<td>Greece</td>
</tr>
<tr>
<td>Mr Zoltan Kovacs</td>
<td>Astrid Research Inc.</td>
<td>Hungary</td>
</tr>
<tr>
<td>Dr László Patthy</td>
<td>ENZIM</td>
<td>Hungary</td>
</tr>
<tr>
<td>PI</td>
<td>Lead Institute</td>
<td>Country</td>
</tr>
<tr>
<td>----</td>
<td>-------------------------------------------------------------------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Prof. Des Higgins</td>
<td>Conway Institute, University of Dublin</td>
<td>Ireland</td>
</tr>
<tr>
<td>Prof. Michal Linial</td>
<td>Life Science Institute, The Hebrew University</td>
<td>Israel</td>
</tr>
<tr>
<td>Prof. Gianni Cesareni</td>
<td>Department of Biology, University of Rome</td>
<td>Italy</td>
</tr>
<tr>
<td>Prof. Graziano Pesole</td>
<td>National Research Council</td>
<td>Italy</td>
</tr>
<tr>
<td>Prof. Anna Tramontano</td>
<td>University of Rome “La Sapienza”</td>
<td>Italy</td>
</tr>
<tr>
<td>Prof. Alessio Valentin</td>
<td>Consorzio Interuniversitario per le Applicazioni di Supercalcolo per l’Università e la Ricerca (CASPUR)</td>
<td>Italy</td>
</tr>
<tr>
<td>Prof. Giorgio Valle</td>
<td>University of Padova</td>
<td>Italy</td>
</tr>
<tr>
<td>Dr Gianluigi Zanetti</td>
<td>CRS4</td>
<td>Italy</td>
</tr>
<tr>
<td>Dr Juris Viksna</td>
<td>Institute of Mathematics and Computer Science and Latvian Biomedical Research and study centre (BMC)</td>
<td>Latvia</td>
</tr>
<tr>
<td>Dr Ruben Kok</td>
<td>Netherlands Bioinformatics Centre (NBIC)</td>
<td>Netherlands</td>
</tr>
<tr>
<td>Prof. Inge Jonassen</td>
<td>University of Bergen</td>
<td>Norway</td>
</tr>
<tr>
<td>Prof. Janusz Bujnicki</td>
<td>International Institute of Molecular and Cell Biology, Warsaw</td>
<td>Poland</td>
</tr>
<tr>
<td>Dr Leszek Rychlewski</td>
<td>Bioinfobank Institute</td>
<td>Poland</td>
</tr>
<tr>
<td>Prof. Piotr Zielenkiewicz</td>
<td>Polish academy of sciences</td>
<td>Poland</td>
</tr>
<tr>
<td>Dr José Pereira Leal</td>
<td>Laboratorio Associado de Oerias (LAO)</td>
<td>Portugal</td>
</tr>
<tr>
<td>Prof. Damjana Rozman</td>
<td>Institute of Biochemistry, University of Ljubljana</td>
<td>Slovenia</td>
</tr>
<tr>
<td>Prof. Alfonso Valencia</td>
<td>Spanish National Bioinformatics Institute (INB)</td>
<td>Spain</td>
</tr>
<tr>
<td>Prof. Bengt Persson</td>
<td>Bioinformatics Infrastructure for Life Sciences (BILS), Linköping University and Karolinska Institute</td>
<td>Sweden</td>
</tr>
<tr>
<td>Prof. Ron Appel</td>
<td>SIB Swiss Institute of Bioinformatics</td>
<td>Switzerland</td>
</tr>
<tr>
<td>Prof. Richard Baldock</td>
<td>Medical Research Council (MRC) Human Genetics Unit</td>
<td>UK</td>
</tr>
<tr>
<td>Prof. Geoffrey John Barton</td>
<td>University of Dundee</td>
<td>UK</td>
</tr>
<tr>
<td>PI</td>
<td>Lead Institute</td>
<td>Country</td>
</tr>
<tr>
<td>--------------------------</td>
<td>--------------------------------------------------------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>Prof. Antony Brookes</td>
<td>University of Leicester</td>
<td>UK</td>
</tr>
<tr>
<td>Dr Dawn Field</td>
<td>Natural Environment Research Council (NERC) Centre for Ecology and Hydrology</td>
<td>UK</td>
</tr>
<tr>
<td>Dr Stephen Gilmore</td>
<td>School of Informatics and Centre for Systems Biology, University of Edinburgh</td>
<td>UK</td>
</tr>
<tr>
<td>Prof. Anthony Harmar</td>
<td>Centre for Cardiovascular Science, University of Edinburgh</td>
<td>UK</td>
</tr>
<tr>
<td>Dr Andrew Law</td>
<td>The Roslin institute, University of Edinburgh</td>
<td>UK</td>
</tr>
<tr>
<td>Dr Johanna McEntyre</td>
<td>UK PubMed Central</td>
<td>UK</td>
</tr>
<tr>
<td>Prof. Christine Orengo</td>
<td>University College London</td>
<td>UK</td>
</tr>
<tr>
<td>Prof. Chris Rawlings</td>
<td>BBSRC, Rothamsted Research</td>
<td>UK</td>
</tr>
<tr>
<td>Dr Jane Rogers</td>
<td>BBSRC, The Genome Analysis Centre</td>
<td>UK</td>
</tr>
<tr>
<td>Prof. Christopher Taylor</td>
<td>University of Manchester</td>
<td>UK</td>
</tr>
<tr>
<td>Mr Stéphane Nauche</td>
<td>European Patent Office</td>
<td>Europe</td>
</tr>
<tr>
<td>Prof. Christian Ohmann</td>
<td>European Clinical Research Infrastructures Network (ECRIN)</td>
<td>Europe</td>
</tr>
</tbody>
</table>
Abbreviations

BBSRC • Biotechnology and Biological Sciences Research Council
BMS • Biological and Medical Sciences
ELIXIR • European Life Science Infrastructure for Biological Information
EMBL • European Molecular Biology Laboratory
EMBL-EBI • European Molecular Biology Laboratory’s European Bioinformatics Institute
ERIC • European Research Infrastructure Consortium
ESFRI • European Strategy Forum on Research Infrastructures
ICA • International Consortium Agreement
IP • Intellectual Property
MoU • Memorandum of Understanding
NCN • National Coordination Node
PDB • Protein Data Bank
SAB • Scientific Advisory Board
SLA • Service Level Agreement
SME • Small-to-Medium Enterprise
UK LFCF • United Kingdom Large Facilities Capital Fund
Annotations Explanatory or additional contextual information added to database records. For example, a DNA sequence may be annotated with information about the species it came from and the geographical location from which it was collected.

Benchmarking A set of techniques for assessing the performance of a software tool

Biomolecular data Information that helps researchers to understand biology at a molecular level, such as DNA sequences and protein structures

Biobank Information that helps researchers to understand biology at a molecular level, such as DNA sequences and protein structures

Bioinformatics The science of collecting, organising and analysing large amounts of biological information, particularly that relating to the molecules of life

Cloud-based computing Location-independent computing, in which the user accesses software and data through the internet

Core data resources Data collections that are of fundamental importance for life-sciences research, and to which other data collections need to refer. They often include primary data.

Curation The quality control, maintenance and organisation of a collection or database

e-Infrastructure Computers, data storage and networking

ERIC The European Research Infrastructure Consortium (ERIC) is a new legal form designed to facilitate the joint establishment and operation of research facilities of European interest.

Load balancing A means for sharing computational work among computers

Mouse clinic A lab that specialises in finding out which characteristics are associated with genetic variations in mice. These can then be correlated with equivalent variations in humans.

Omics A generic name for high-throughput technologies in biology such as genomics, transcriptomics, proteomics and metabolomics
Ontology

In computer science and information science, an ontology is a rigorous and Exhaustive system for representing entities and the relationships between them.

PDB

The Protein Data Bank (PDB) contains information about experimentally determined structures of proteins, nucleic acids, and complex assemblies.

Primary database

A database consisting of data that were derived experimentally, such as DNA and RNA sequences or three-dimensional structures. The data in primary databases are considered to belong to the person who submitted them, and cannot be modified by a third party.

Registries

Databases for storing information about computer systems.

Secondary database

A database containing information derived from primary data, for example, translations of DNA sequences into protein sequences.

Structural Biology laboratory

A lab that specialises in determining the three-dimensional shapes (structures) of proteins and other large molecules. This often provides important clues as to the function of the molecule.

Server

A computer that is dedicated to providing a specific service.

Specialised data collections

Data collections that are used by a specialist community of researchers (for example, researchers studying a particular organism or disease), rather than to the research community at large. Specialised data collections often refer back to core data resources.

Terabyte

1,000,000,000,000 bytes.