ELIXIR

Scientific Programme 2014-2018

Contents

About ELIXIR 1

Executive Summary 2

Europe's Bioinformatics Infrastructure: key challenges 2014-2018 4

ELIXIR's Strategic Objectives, 2014-2018 10

ELIXIR Programmes of Work 14

ELIXIR Pilot Actions 20

ELIXIR Funding Strategy 21

ELIXIR Goverance and Organisation 23

Realising the Benefits of ELIXIR 26

With special thanks to all of those who contributed to the development of this ELIXIR Programme, most notably Heads of Nodes, Technical Coordinators and members of the preparatory working groups.

About ELIXIR

Europe is home to some of the world's leading bioinformatics institutes and resources. ELIXIR represents the collective efforts of its members to coordinate and sustain these vital resources, providing the platform for scientific discovery in the life sciences.

ELIXIR brings together national bioinformatics capacities and those of the European Bioinformatics Institute (EMBL-EBI) to scale up our collective capability to archive, integrate, analyse and exploit the large and heterogeneous data sets produced in modern life science research.

Life science is becoming increasingly collaborative and complex. Scientists use diverse technologies to understand organisms and diseases from the molecular to the systems level. Ever-larger volumes of data must flow freely – at an unprecedented scale – between platforms, laboratories and people. Researchers need to explore heterogeneous data from multiple sources in order to conduct meaningful analyses and extract knowledge. ELIXIR endeavours to meet this challenge by coordinating Europe's national and international capabilities, forming a coherent infrastructure that enables our millions of annual users to seamlessly navigate the ecosystem of life-science data services.

ELIXIR ensures that users – who may be individual scientists, large consortia or indeed other research infrastructures – can easily access data resources that are sustainable, built on strong community standards and safeguarded for the long term. By providing the needed support for discovery, integration and reuse of complex data, ELIXIR makes sure that these resources that can be sustained well into the future form an integral part of European life science projects. Finally, through a comprehensive training programme, ELIXIR will help to build the skills critically needed to extract maximum value from public life-science data.

ELIXIR Nodes

ELIXIR is a distributed infrastructure built around existing bioinformatics centres of excellence: national Nodes and international organisations such as the European Bioinformatics Institute (EMBL-EBI).

Bioinformatics infrastructure in Europe is highly variable. Many countries have long-established national bioinformatics infrastructures that provide nationally coordinated services. Some countries have world-class bioinformatics communities that are fragmented, others have emerging bioinformatics communities and require support from ELIXIR to build capacity. In all cases, ELIXIR brings added value through transnational coordination. One of the lasting benefits achieved by the preparatory and interim phases of ELIXIR is the formation - through the hosting of ELIXIR Nodes - of national bioinformatics communities. Now in the permanent phase, additional capacity building and integration forms a core part of the ELIXIR Programme.

Bioinformatics platforms in the Nodes are tightly linked to research projects within their national life-science community. For example, they might provide biomedical cloud platforms, compute services or participate in collaborative scientific projects in response to calls for proposals from funding agencies. Because of its distributed model, ELIXIR is linked with many user communities throughout Europe and is well placed to provide a platform that builds on national research strengths and priorities to support data-driven life science research.

Legal framework

The legal framework of ELIXIR is based on the ELIXIR Consortium Agreement (ECA), which has been concluded among the Member States and the European Molecular Biology Laboratory (EMBL). Based on the ECA, ELIXIR uses EMBL's legal personality; however, ELIXIR activities are independently funded and governed by the ELIXIR Members through its decision-making body, the ELIXIR Board.

Our users

The resources from ELIXIR's national Nodes and EMBL-EBI are widely used throughout the world. For example, every month 160,000 users worldwide access the Expasy services at the SIB-Swiss Institute of Bioinformatics. ELIXIR's core resources and life-science data archives have long-standing global collaborations with North American and Japanese counterparts: for data infrastructures, the larger the scale of collaboration, the more data sets are available and, accordingly, the more accurate and useful are the services developed. Global collaboration is critical in the development and maintenance of international standards for data exchange and interoperability.

Executive Summary

The ELIXIR Scientific Programme for 2014-2018 sets out the themes addressed by ELIXIR in the first stages of its implementation. It is Europe's strategic response to the data needs of life-science research.

ELIXIR has an unparalleled potential to pool resources and share capabilities across a large user-base in academic and industrial research. This Scientific Programme sets out how the transnational and distributed ELIXIR data infrastructure will drive a step change in European life science research and development.

Sustained and secure data

- ELIXIR will identify core data resources that are essential to the larger international community and develop a robust framework to secure their long-term sustainability;
- ELIXIR Nodes will handle sensitive, personal data through the continued development of secure archives. Research in human subjects requires platforms with authentication services and effective governance processes for secure access and data exchange; ELIXIR will work to provide comprehensive, end-to-end solutions for data privacy that go beyond simple download protection.

Best practice and standards

- ELIXIR will support the development of robust, production-quality resources through best practice established in the Technical Coordinators group and the ELIXIR Training Programme;
- ELIXIR will work with other biological and medical research infrastructures to integrate fundamental data resources, services and standards to ensure that users can effectively make use of data generated with diverse technologies at different sites throughout Europe. This will involve developing and implementing standards for research data management, publication and interoperability;
- An ELIXIR Tools & Service Registry will enable users to find the appropriate resource or service for their work. It will also clarify terms and establish the necessary support levels for developing and maintaining effective workflows.

A highly skilled workforce

- To address Europe's critical gap in highly skilled life-science data management professionals, ELIXIR's Training Programme will train a large and diversifying user base to fully exploit ELIXIR resources;
- ELIXIR will support efforts to ensure Europe has the appropriate skills capacity for data management and the implementation of data standards;
- To facilitate data reuse and innovation, ELIXIR will train and support research communities and other infrastructures to ensure data and their context are described consistently.

Technical services

- The technical services underlying life-science experiments are undergoing fundamental change due to exponential growth in data volumes, complexity and sensitivity. ELIXIR will provide a core set of technical services that interface with and make use of select European and national e-Infrastructures. The aim of these activities is to deliver specific solutions for researchers working in all domains of life science;
- ELIXIR, in close partnership with research infrastructures and user communities, will establish large-scale facilities that provide 'computable storage' platforms and virtual access for diverse research areas.

Expanding partnerships

- ELIXIR will engage with large research consortia and public–private partnerships such as the Innovative Medicines Initiative (IMI) and the Bio-based Industries Consortium (BIC) to ensure that data- and knowledge-management efforts are joined up and sustainable in the long-term. This will be achieved through the development and delivery of shared services arising from bilateral collaborations and consortia;
- Together with other research infrastructures, partners, and large research programmes, ELIXIR will endeavour to establish joint technical strategies and Memoranda of Understanding (MoU) that guide local and national shared efforts in life-science data management and resource provision;
- A portfolio of scientific and technical pilot actions will be established as ELIXIR's primary mechanism for implementation and review of new technologies. These pilot actions, carried out in close consultation with user communities, will drive the adoption of ELIXIR resources and services.

Europe's Bioinformatics Infrastructure: Key challenges 2014-2018

"It is a profoundly erroneous truism, repeated by all copybooks and by eminent people when they are making speeches, that we should cultivate the habit of thinking of what we are doing. The precise opposite is the case. Civilisation advances by extending the number of important operations that we can perform without thinking about them."

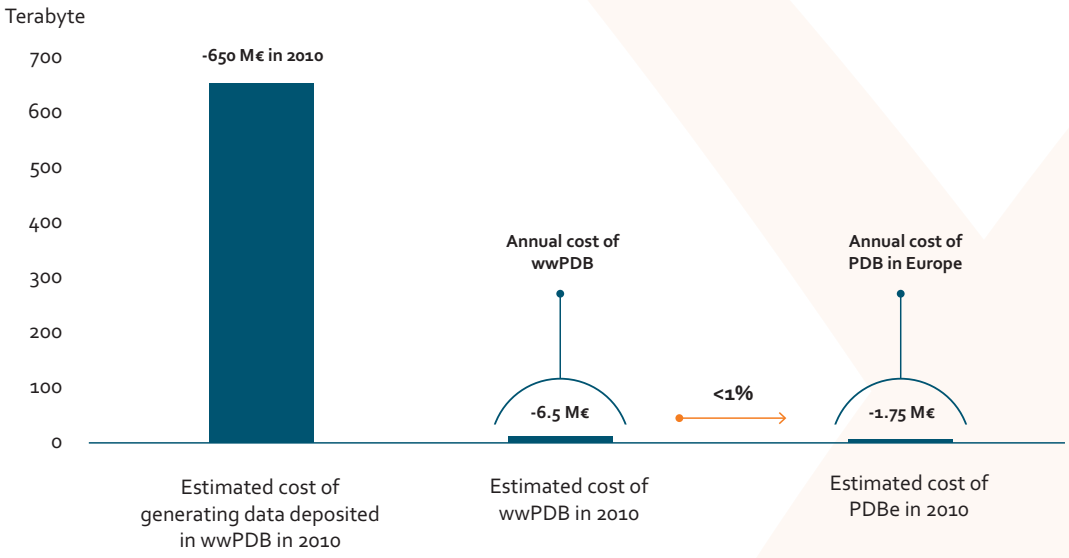
A.N. Whitehead, An Introduction to Mathematics, 1911

Biology has a rich tradition of accurate collection and reuse of data,[1] from the systematic cataloguing that underpins our fundamental understanding of species, tissues, cells and genomes to the massive studies of human variation,[2] functional genomics[3] and integrated cancer atlases.[3]

These data resources serve as critical reference tools for biologists in laboratories: every day more than 10,000 users from across the world access the EMBL-EBI services[4]. Biological data resources have provided the bedrock for many discoveries, both planned and serendipitous, over the decades. A recent example is the identification of novel risk factors for Alzheimer's disease based on a large-scale meta-analysis;[5] such efforts depend critically on prior estimates on human genetic variation calculated from public data sets such as the 1000 Genomes. Meta-analysis of public life science databases, albeit of a very different kind, also underpins the molecular design and docking tools in daily use in drug-discovery laboratories across the world. The development and validation of drug-design tools, many of which are successfully commercialised, have all relied on carefully curated data sets extracted from public archives.

Long-term stewardship

As data has become an essential commodity for biological research, the importance of making both the narrative and the data from publicly funded research openly available is broadly recognised. The importance of long-term stewardship is highlighted by the observation that the odds of retrieving the data from a publication decline by 17% per year.[6] This is in sharp contrast to biomolecular data in a major public resource such as the Protein Data Bank (PDB), which has safeguarded the high-resolution structures of proteins, nucleic acids and complex assemblies since 1971.



Comparison of the annual cost of generating novel data on protein structures worldwide with the cost to Europe of storing it in the PDB. Indeed, the European arm of the PDB (which stores all known protein structures determined worldwide since the 1971) costs much less than 1% of the cost of re-generating one year's data. wwPDB is the worldwide Protein Data Bank.

[1] 'On Not Reinventing the Wheel.', Nature Genetics, 44 (2012), 233.
[2] The ENCODE Project Consortium, 'An Integrated Encyclopedia of DNA Elements in the Human Genome', Nature, 489 (2012), 57.
[3] The Cancer Genome Atlas Research Network et al., 'The Cancer Genome Atlas Pan-Cancer Analysis Project', Nature Genetics, 45 (2013), 1113–1120.
[4] Source: EMBL-EBI web logs
[5] Jean-Charles Lambert et al., 'Meta-Analysis of 74,046 Individuals Identifies 11 New Susceptibility Loci for Alzheimer's Disease', Nature Genetics, 45 (2013), 1452–1458.
[6] Timothy H Vines et al., 'The Availability of Research Data Declines Rapidly with Article Age', Current Biology, 24 (2014), 94–97.

Indeed storing and, importantly, making all structural data available for broad reuse costs less than 1% of regenerating one year's new depositions. Advanced services such as the SWISS-MODEL (with over 280,000 registered users globally),[7] are built on top of these core data resources.

Investment in a sound infrastructure for biological data builds a foundation for all aspects of life science research from biodiversity, agriculture to human health and underpins scientific projects in academia as well as industry. It serves as a solid insurance policy to secure the huge investments in life science research and allows for comparative research and meta-analysis at an unprecedented scale. Nevertheless the data infrastructure in the life sciences needs to meet a number of challenges to ensure that scientists and society can make full use of the opportunity. Based on wide community engagement and expert advice during the ELIXIR Preparatory and Interim Phases, the five most critical issues are outlined in this section. These form the key drivers underpinning ELIXIR's strategic objectives outlined in this Programme.

Growing, distributed, complex and sensitive data

High–throughput sequencing based assays have become a standard tool to understand both genomic variation and the functional aspects of gene regulation, expression and structure. While human health has been the primary driver, life-science disciplines including agriculture and ecology are catching up fast. The need to have meticulously curated metadata and existing knowledge available in computer-readable formats grows in tandem with data generation. Reuse of open research data requires robust annotation to ensure the accurate representation of conditions and experimental context.

Growing

Many life-science data resources now double in size every 6-12 months. For example, the EMBL-EBI data centres are set to manage over 50 petabytes (PB) of storage. With this flood of data, one open issue is the seemingly practical question of how and where to store the research data produced by public research projects. In large, international genome consortia, it is well-established best practice to have a designated data coordination centre that generally involves one or several of the public data repositories. Specific work packages address the data curation, annotation of metadata and data deposition. However, there is an emerging issue of local or regional large-scale data generation efforts (e.g. from clinical cohorts and crop phenotyping efforts) with traditionally weaker organizational links to the major archives. The aggregated resource needs required for deposition and integration of such data into the overall data landscape will pose a significant challenge and require dedicated efforts and resources.

Distributed

As accessibility to the results of publicly funded research becomes the norm – and a requirement of both funding bodies and academic journals – it is critical that the public infrastructures work with the publishers to develop strategies that include both sustained data deposition into the major archives as well as data management efforts to handle the "long tail" of research data that is not suitable for storage in the traditional archives. This will require collaboration across national and organisational borders. However, it is clear that a large fraction of the data generated in Europe will need to be managed in a federated manner. For example, human clinical data can often not leave the network of the home institute without additional ethical review and patient consent. Furthermore, as high-content biology, metabolomics and sequencing-based assays rapidly develop we must incorporate new types of data and support users in the integration and exploitation of these. For example, in the environmental domain, there is a growing need to develop interfaces between multi-layer datasets, such as physical and chemical descriptors of the environment linked to both biomolecular and biodiversity datasets.

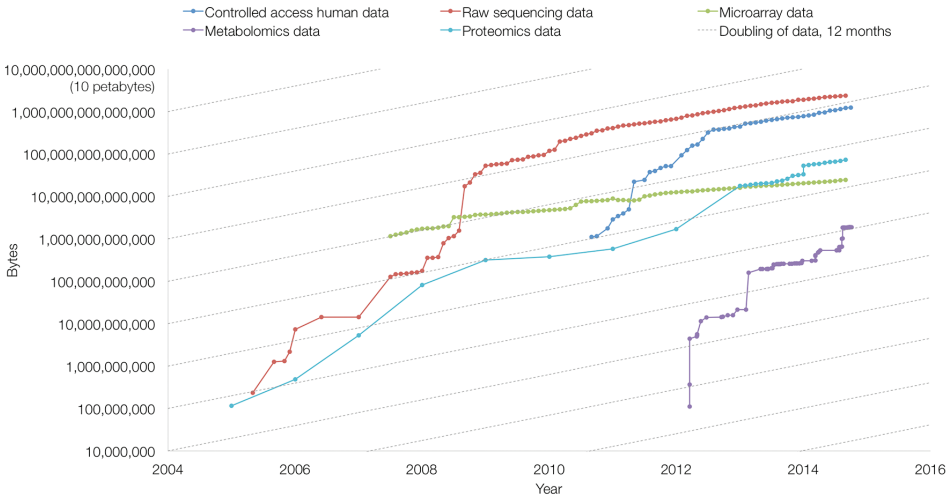
Complex

Despite the life science data surge, the major challenge is not the volume of data but the complexity: biological data is heterogeneous, information dense and has complex assumptions and error models that depend on experimental details. Biological data is also fragmented, with large data sets produced at thousands of sites across Europe, often differing widely in formats, annotations and the standards used. For example, the Wheat Genome Sequencing Consortium involves scientists from over 50 countries. A further challenge is that interpretation often requires extensive integration of very large data sets. In the field of medicine we expect there to be at least 100,000 cancer genomes generated in Europe over the next 10 years from both biomedical research projects and the routine application of sequencing in healthcare. In addition to the massive data storage and exchange requirements from these efforts, they will also necessitate effective access to reference data and tools of critical importance to individual researchers and health-care professionals.

[7] *Source: SWISS-MODEL server logs*

Secure

In the face of the data deluge, limited resources and lack of coordination, the major European data archives will become rapidly unable to meet the full demand. This applies to both raw data archives such as the European Nucleotide Archive as well as the value added resources with extensive annotation such as Human Protein Atlas and UniProt. At the national level, the ELIXIR Nodes have to meet increased support needs for management and publication of research data. It is therefore essential that ELIXIR, through national Nodes as well as coordinated European investments, develops coherent strategies to sustainable life science data management, archiving and services. This requires a coordinated response from researchers, research funders and science policy bodies. A key issue in this context is transparency of the different cost drivers for research data management and archiving.

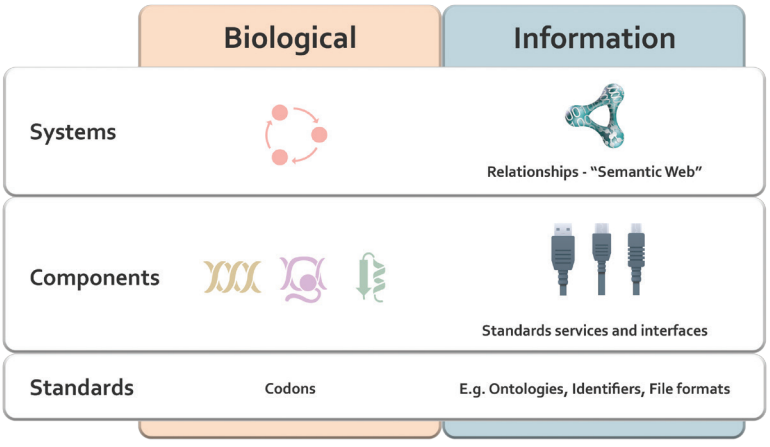


Data growth curves of 5 major EMBL-EBI resources (European Genome-phenome Archive (EGA); European Nucleotide Archive (ENA); Proteomics data repository (PRIDE); Metabolomics resource (MetaboLights); and Functional genomics database (ArrayExpress) over the years 2005-2013. Source: EMBL-EBI.

Community standards

The divergent formats representing biological data create significant barriers to data integration and reuse, and this challenge to data interoperability is a serious bottleneck in the life sciences.[8]

Community standards drive reuse in many areas of the life sciences, and allow new scientific discoveries. A good example is the series of "minimal information" standards that articulate metadata requirements for data depositions. In the functional genomics field, where this was pioneered through the "MIAME" standard, there have been systematic evaluations of data reuse: it is estimated that on average one of every three deposited data sets contributes to new publications.[9] Data reuse underpins and validates additional experimental studies as well as providing a foundation for further,[10] value added resources such as the Connectivity Map that globally integrates human functional genomic data and links this to drugs and diseases to elucidate mechanisms of action for drugs.[11]



Just as biology assembles universal building blocks of life into complex systems, so information assembles compounds with standard services and interfaces to allow analysis and exploration of relationships in the data.

[8] David B Searls, 'Data Integration: Challenges for Drug Discovery', Nature Reviews Drug Discovery, 4 (2005), 45–58.
[9] Heather A Piwowar, Todd J Vision and Michael C Whitlock, 'Data Archiving Is a Good Investment', Nature, 473 (2011), 285–285.
[10] Johan Rung and Alvis Brazma, 'Reuse of Public Genome-Wide Gene Expression Data', Nature Reviews Genetics, 14 (2012), 89–99.
[11] Justin Lamb et al., 'The Connectivity Map: Using Gene-Expression Signatures to Connect Small Molecules, Genes, and Disease.', Science (New York, N.Y.), 313 (2006), 1929–1935.

The value of open-access infrastructure

Researchers in the life sciences take for granted that data on gene structure, expression and function are openly accessible. The rich tradition of well-curated data resources has resulted in services of wide applicability and usage. For example, the UniProt resource of manually aggregated functional data has over 800,000 daily requests.[12]

In agriculture, genomic science enables improved breeding practices and improved crops to meet the challenges to food security posed by growing populations and climate change. 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 is 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.

Using innovative tools to search biological databases, biomedical researchers reuse data to discover insights and solutions. For example, by identifying patterns of genes that are active in different tumours, researchers can predict how aggressive a specific tumour is and decide which drugs to treat it with. Global collaborative projects such as the 1000 Genomes Project sequence the genomes of several thousands of individuals and make the data openly available. This information can then be used in numerous studies to associate minute differences in genome structure with the susceptibility of individuals to common life-threatening conditions such as heart disease and type II diabetes. The importance of access to reference data for validation as well as estimating variance and effects in patient populations was highlighted by US National Cancer Institute in their recent recommendations on the use of omics data in clinical trials.[13]

Industry use of many of Europe's bioinformatics resources is high. For example, the total number of hits to the EMBL-EBI website from 'commercial' IP addresses was around 110 million in 2013.[14] Users from industry range from SMEs to large multinationals and the sectors covered include pharma, biotechnology, healthcare, through to food and agriculture. These industries are major employers globally, generating wealth and supporting transformation to a knowledge-based economy. Bioinformatic services have significant value and application to industrial users, who download public data for integration with proprietary data. Therefore, it is essential that we maintain the accessibility of biological data to industry in order to enhance competitiveness and innovation. Europe invests over $80 billion annually in biological and medical R&D with over 60% of this investment in industrial research.[15] Robust infrastructure and open, collaborative partnerships are essential to remain globally competitive – it is worth noting that while European research investments have been stable since 2007, investments in Asia have grown by 150% to $60 billion over the same period with much of this growth in the industrial sector.

The skills gap in life science data management

A recent analysis from McKinsey global institute[16] on the use and the potential of big data in the healthcare sector identified three major blockers to progress: the lack of standards and agreed norms for care assessment, the challenge of shifting the mind-set from "protect" to "share and protect" data, and, critically, the significant skills gap for the management and exploitation of the information assets.

A network of trained experts is needed to analyse the increasing amount of data generated in Europe.

It is estimated that by 2018 there will be a "talent gap" in the US alone of between 140,000 and 190,000 trained experts to analyse the increasing amount of data being generated.[17] The training needs for European bioinformatics were extensively analysed during the ELIXIR Preparatory Phase with recommendations to closely couple the development of data resources to the provision of training, to support the trainer community with infrastructure components and tightly integrate with other training initiatives, notably the global bioinformatics training network GOBLET but also other European training infrastructures.

[12] *Source: UniProt web logs.*
[13] *Lisa M McShane et al., 'Criteria for the Use of Omics-Based Predictors in Clinical Trials', Nature, 502 (2013), 317–320.*
[14] *Source: EMBL-EBI web logs*
[15] *Justin Chakma et al., 'Asia's Ascent--Global Trends in Biomedical R&D Expenditures.', The New England journal of medicine, 370 (2014), 3–6.*
[16] *McKinsey & Company, 'Open Data: Unlocking Innovation and Performance with Liquid Information', 2013 <http://www.mckinsey.com/insights/business_technology/open_data_unlocking_innovation_and_performance_with_liquid_information>.*
[17] *McKinsey & Company, 'Big Data: the Next Frontier for Innovation, Competition, and Productivity', 2011 <http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation>.*

Bridging data between research infrastructures and scientists

The European large-scale research infrastructures in biomedical sciences, including ELIXIR, bring together national and European investments and ensure broad access to advanced research tools, biological samples and the associated data. For the investments to be effective the infrastructures need to work closely with user communities and research funding networks. In the biomedical domain there are also investments in large public private partnerships, such as the Innovative Medicines Initiative (IMI), that bring public research institutions and academic researchers together with industrial R&D organisations. It is a risk that the large number of stakeholders and programmes drive fragmentation of data management and technology platforms. ELIXIR will have an important role as a bridge between the infrastructures and initiatives through coordination and consolidation of the fundamental data services.

From preparation to implementation

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 pan-European Research Infrastructures that were of key importance for the development of science and innovation in Europe. Among them, ELIXIR is one of very few considered to be of truly global significance.

The Preparatory Phase of ELIXIR began in November 2007, funded by the European Commission (Framework Programme 7, Capacities) and coordinated by EMBL-EBI. The key objectives of the Preparatory Phase were to:

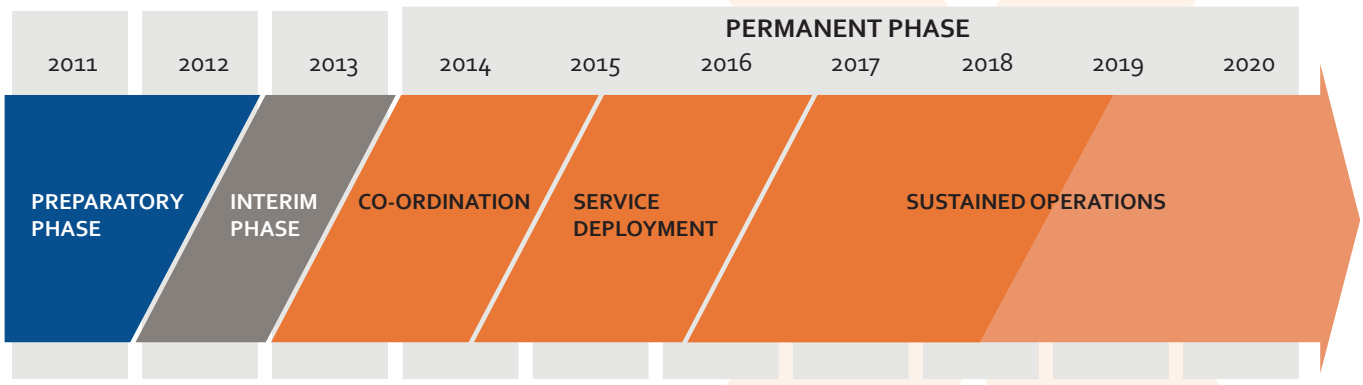
- define the scope of the infrastructure, its role and benefits;
- define an appropriate governance and legal structure;
- define a long-term funding structure to provide a sustainable infrastructure.

The work was carried out by representatives of 32 organisations originating in 13 countries, overseen by the ELIXIR Preparatory Phase Steering Committee, which comprised of many of Europe's most well respected and influential bioinformaticians, users of biological data, and policy specialists.

Following an extensive stakeholder consultation phase, the Preparatory Phase delivered reports with findings and recommendations for the construction of ELIXIR. These were used to draft the Business Case for ELIXIR, detailing the requirements of a European Infrastructure for biological information, and a decision was taken by the ELIXIR Preparatory Phase Steering Committee that a 'light' non-binding Memorandum of Understanding (MoU) for ELIXIR would first be established between interested parties, in order to set up an interim structure for ELIXIR, in advance of a more binding agreement being developed.

Laying the groundwork for an ELIXIR board

In 2011, the Business Case and MoU for ELIXIR were issued to governments and funding agencies for their consideration, with the appointed representatives of those countries signing the MoU to form an Interim ELIXIR Board. The Interim ELIXIR Board was tasked with the negotiation of the ELIXIR Consortium Agreement (ECA) a suitable legal framework for the operational phase of ELIXIR, and a working group for the drafting of the ECA comprising of representatives of the Interim ELIXIR Board was subsequently established. Countries that had signed the MoU were also invited to put forward an application for an ELIXIR Node offering services to ELIXIR, for review by the ELIXIR Scientific Advisory Board.



The first five-year (2014-2018) Construction and Operation Phase of ELIXIR will proceed through three stages.

The ELIXIR Consortium Agreement

The ELIXIR Preparatory Phase officially ended in December 2012 and in April 2013 the ECA was approved by the Interim ELIXIR Board, when the formal ECA ratification process began. ELIXIR's Interim Phase saw the submission and review of ELIXIR Node Applications, the establishment of five technical pilot actions to test the distributed concept and a build up of initial capacity within the ELIXIR Hub. Following the approval of EMBL Council and the ratification of the ECA by five countries, which was completed in December 2013, ELIXIR moved from Interim to Permanent Phase with the ECA's entry into force in January 2014. This will be characterised by the implementation of the ELIXIR's Construction and Operation period, which begins in 2014. In this five-year period, the build of the permanent research infrastructure will proceed through three phases of Co-ordination (2014), Service Deployment (2015) and Sustainable Operations (2016-2018).

In 2012, the EC formed a High Level Assessment Expert Group (AEG)[18] composed of reputed research professionals who were tasked to assess the financial and managerial maturity of ESFRI research infrastructures. In their report published in 2013, the AEG emphasised that, for a research infrastructure to be considered to have reached a mature implementation and operation, there are requirements of governance, operational excellence and management that go beyond the establishment of a legal structure.

Getting ready for implementation

Specifically for ELIXIR, the AEG recommended service level agreements between the Nodes and the Hub, an overarching Ethical Board, project-wide metrics with target values to measure progress and clear plans for the long-term funding strategy of the infrastructure. The AEG stated that, if all their recommendations were followed and completed, ELIXIR should be ready for implementation by 2015.

[18] *http://ec.europa.eu/research/infrastructures/pdf/jd-final-aegreport-23sept13.pdf*

ELIXIR's Strategic Objectives, 2014-2018

ELIXIR has identified eight strategic objectives for the first operational phase, thanks to extensive preparatory work across the emerging Nodes. These strategic objectives drive the development of the ELIXIR infrastructure at national and European levels, and lay the foundation for long-term sustainability and impact of our collective data resources.

The strategic objectives outlined here are delivered through the ELIXIR Programmes of Work, which bring people together across the Nodes to deliver services.

1. Establish a distributed infrastructure that scales with the challenge

At the core of the ELIXIR strategy is the recognition that large-scale data production in the life sciences is not limited to a few sites. A European data infrastructure must be able to cope with the aggregation, annotation and integration of data from thousands of laboratories as well as scaling the data-services to millions of users worldwide (e.g. EMBL-EBI services are accessed by over 2M distinct users every year).[19] As the importance of basic biological methods increase for medical research, the infrastructure should also be capable of handling the different legal, regulatory and ethical requirements throughout Europe.

ELIXIR is a distributed infrastructure in which Nodes in ELIXIR Member States and international organisations such as EMBL-EBI are brought together into a single, coherent infrastructure by an international consortium agreement and a co-ordinating Hub. Distributing resources throughout Europe allows ELIXIR to scale local data production and usage, build on the strengths of European centres of excellence, and give the organisation flexibility to meet national priorities and demand.

Many ELIXIR Nodes, including EMBL-EBI, have existing significant infrastructure that also includes responsibility for national research data management, provision of national computing platforms as well as direct consultancy links to research projects. The Nodes closely collaborate with other distributed research infrastructures in the biological and medical sciences (e.g. biobanks, imaging centres, and screening centres) and service these with a data infrastructure. The Nodes will maintain these key roles within ELIXIR, while the added benefit of the ELIXIR infrastructure is to develop and spread best practices; the ELIXIR Hub will support the Nodes by providing coordinating services and steer European-scale collaboration of key services, technical protocols and training as well as drive joint investments and shared operational costs. Hence it is of critical importance to ensure the long-term sustainability of the infrastructure, including the recruitment of new member states and securing financing for the coming years.

ELIXIR Nodes make the distinction between Node Capabilities, i.e. the expertise and competence that the Node can share within the ELIXIR framework and the Node Capacity, i.e. the sustainably funded services with global impact that are offered through ELIXIR Service Delivery Plans. This distributed infrastructure is core to ELIXIR's strategy as it enables full data integration and maximises the value of the collective and expanding capacity across the continent and across all domains of life science.

The pace of technical and scientific developments in the life sciences is rapid and, as an underpinning infrastructure, ELIXIR needs to be responsive to changing usage patterns and emerging data-needs. Developing an agile infrastructure that is able to rapidly respond to user needs and advances in science requires close alignment and strong links with user communities. During the Interim Phase, ELIXIR ran a series of pilot actions: short, use-case-driven projects that utilise the capabilities of several Nodes (and the Hub) to adapt services to emerging user needs and inform the implementation of novel technologies. This was exceptionally successful: several of the pilot actions have now attracted external funding or moved into a production state for long-term delivery. Hence, running a portfolio of use case driven pilot actions or short implementation "sprints" will be an integral and important component of the ELIXIR strategy during this first Programme. Through pilot actions ELIXIR will work closely with user communities, research infrastructures and other partners to ascertain delivery of relevant services. It is expected that the results of these pilot actions will inform the further development of ELIXIR's strategy and form an important basis for the discussions with the Scientific Advisory Board.

There is also great potential for scientific collaboration between ELIXIR and other world countries on specific

[19] *Source: EMBL-EBI web logs*

subjects and themes. ELIXIR would be keen to establish scientific agreements or Memoranda of Understanding for collaboration with particular countries on specific issues. Areas ripe for exploration include:

- Collaboration with the US on interoperability and data management, working with the NIH Big Data to Knowledge programme to build on the already well-established links that exist;
- International collaboration on plant genome sequencing, building on successful US-led initiatives such as iPlant ;[20]
- Global collaboration on the issue of orphan drugs and rare diseases;
- Developing international collaboration on descriptions of basic studies and samples for long-term tracking and linking of data types;
- Global collaboration on use of genomics data in health.

2. Secure and deliver the core data resources underpinning life-science research

ELIXIR is fundamentally a data infrastructure. During the ELIXIR Preparatory Phase the landscape of European data resources was surveyed and it was noted that, of the over 1000 resources identified, only a small fraction had institutional support and long-term funding commitments. For ELIXIR it is necessary to identify core data resources of wide applicability and usage. Furthermore, these resources must be well-maintained with capacity and processes for professional service delivery, as well as plans for life-cycle management and understanding of service dependencies.

- ELIXIR will identify a set of Core Resources that are globally competitive and of critical importance to the life science community and actively promote their integration and sustainability. The Core Resources will be delivered under their own brands, as services from the Nodes, and will form the backbone of the ELIXIR data infrastructure;
- ELIXIR will identify ELIXIR Named Services from the Nodes that, based on clear eligibility criteria and best practice in service delivery, are visibly branded and provide the bioinformatics user community with a toolbox of stable and well-maintained services;
- ELIXIR will support Emerging Services that do not yet meet the full criteria of ELIXIR Named Services, and support these through best practice, the ELIXIR Training Programme and through the technical coordinator network.

Whilst ELIXIR will only brand stable resources emanating from a professional delivery organisation and with a track record of reliable delivery, in most cases ELIXIR resources and services will not be in a position to offer legally binding service level agreements. Nevertheless, transparency on service availability, terms of use and uptime is critical. For Core Resources and Named Services, ELIXIR aims to provide service level declarations, which are non-binding but visible statements that clarify support levels and historical data on service availability. Coupled with the ELIXIR Tools & Service Registry this approach will provide a transparent and stable set of resources and services to the user community.

3. Provide discoverable tools, services and connectors to drive data access and exploitation

Researchers need straightforward means of knowing what services are available and where to obtain them, along with clear descriptions of availability, licensing, performance, documentation and training materials. Building on the significant community efforts in this area, ELIXIR will develop a comprehensive registry of services and resources allowing for service discovery and for promotion of interoperability. ELIXIR's focus on tools is primarily for data access and interoperability, but there is also a great need for services that integrate different data types and provide value-added annotations and comparisons.

Users have a critical need to integrate tools and services into robust scientific workflows. ELIXIR will facilitate the interoperability of tools through coordinated efforts with Nodes and other research infrastructures. For instance, data access through web services is made interoperable by adopting standard formats and query interfaces. Interoperability can also be enhanced for graphical user interfaces by reusing standard web components and best practices in user interface design. Interoperability is not just one solution, it is a joint set of approaches applicable to a variety of services: data access, submission, curation, analysis, or integration.

To facilitate service interoperability, access modes of each tool will be listed in the Tools & Service Registry. The registry should also highlight dependencies between tools. Thus, it will be possible to develop and maintain operational workflows based on ELIXIR services.

Integration and standard service interfaces are critical for the interoperability between biological and medical research infrastructures. In the BioMedBridges project,[21] there is a concerted effort to work on standards in terms of shared ontologies and identifier mappings between

[20] http://www.iplantcollaborative.org/
[21] http://www.biomedbridges.eu

the infrastructures as well as interfaces and web services. Initially, a set of 14 high priority use cases was identified and currently service interfaces are being delivered based on a common architecture. Long-term support and development of such interfaces will also be a critical ELIXIR activity.

4. Provide robust technical platforms and clouds for secure data access, data exchange and compute

Accessing and computing with the large data volumes from modern biology comes with a specific set of challenges for the technical infrastructure. ELIXIR will develop solutions of widespread utility in the life-science domain; for example, there is a broad interest and active development of cloud-based services within the life sciences. ELIXIR ran a pilot action on private clouds and several Nodes are routinely providing biomedical cloud services to the national, and in some cases international community. Thus, ELIXIR will provide recommendations and solutions for the present and future specialised data service providers in life science and medicine.

Technical services must be sustainable and scalable so that data specialists can rely and build on them. Underpinning the ELIXIR Technical Services will be a range of sustainable national and European e-Infrastructure solutions that have to flexibly adapt as technologies in life sciences and medicine change the data production landscape. For example, translational and biobanking infrastructures at both the European and national level demand effective services to securely access and exchange data. ELIXIR can offer such services, as was demonstrated by the ELIXIR pilot action on federated authentication that enables governance processes and access to personal genome data through institutional logins. The ELIXIR Technical Services Strategy aims for transforming solutions of widespread utility in the life-science domain for accessing and computing with the large data volumes from modern biology.

5. Develop and maintain standards for data management, reuse and integration

In order to derive new insights, data-driven science requires the seamless connection and joint interrogation of newly generated data with interoperable legacy information. However, despite massive investments in research, most research data and tools become inaccessible with time, and even if available, they often cannot easily be reused for new discoveries. Responding to the requirement from funding bodies and society, research data is increasingly open but standards, formats and resources are fragmented and data can be difficult to find.

Data sharing and reuse is tightly coupled to effective research data management and the presence of processes and infrastructure to support data coordination, metadata curation and deposition in suitable archives. At the national level the ELIXIR Nodes are often deeply involved in national Research Data Management efforts including both technical services and policy developments. The data infrastructure developed and maintained by ELIXIR at the European level must enable both long-term data archiving and access but critically also enable accessibility and full data integration to make the best use of Europe's collective and expanding capacity. This strategic objective is tightly coupled to the provision of ELIXIR Core Resources; ELIXIR will close the cycle of data annotation, deposition, provision and integration for reuse.

Data interoperability has both a technical and a social aspect. The ELIXIR focus is on services for improving and smoothing the adoption of standards and vocabularies by the community. This will include technical services such as ensuring programmatic access, data interoperability and identifiers as well as development of, for example, minimum information standards and vocabulary services. ELIXIR will also provide social services such as standards dissemination and reporting, training in standards, and validation of data sets against data formats, ontology usage, minimum information standards, etc. As an infrastructure, key contributions from ELIXIR will be the sustainability of data management activities as well as the career recognition of data creators and curators.

These services require the coordinated development, implementation and deployment across Europe and strongly link with publisher requirements and literature resources such as Europe PubMed Central. Many existing community efforts address sub-challenges and provide sub-solutions related to the broader area of data interoperability. Hence ELIXIR will not dictate a single platform but rather focus on conventions that enable data interoperability, stewardship and compliance with data and metadata standards, policies and practices. The development of standards is an area where global collaboration is imperative and ELIXIR will actively engage with similar efforts globally.

6. Partner with user communities in a sustainable manner to ensure high and lasting impact

It is critical for ELIXIR services to be closely aligned with and respond to the requirements of scientific communities

across the broad field of life sciences and medicine. ELIXIR's portfolio of cross-cutting resources will require additional services to address the demands from specific research domains and it is vital that ELIXIR actively identifies these with the user communities. ELIXIR needs, via a use-case-driven approach, to develop scalable and sustainable partnerships for the work in specific projects and communities. This could be through partnership with large collaborative projects, but also by collaborating closely with other research infrastructures or public-private partnerships such as the IMI projects. Partnerships allow ELIXIR to identify and implement solutions that are applicable across fields, gain from synergies and pave the way to new infrastructure services. A particularly important user group comprises the bioinformatics core facilities across Europe, at which national Nodes have a mandate to ensure that the needs of embedded bioinformaticians are met.

Stable, sustainable resources and services with clear terms of use and management plans are the foundation for such partnerships: when research infrastructures, projects and user communities embed and extend ELIXIR services, long-term relationships are forged that drive the development of an integrated data and knowledge management infrastructure. The continued development of these partnerships requires a high degree of confidence in ELIXIR resources; transparency and life-cycle management are critical for effective management of dependencies. ELIXIR resources and services that form integral parts of other research infrastructures will also require joint sustainability planning and strategic developments.

Thus, ELIXIR will seek to establish framework agreements, shared technical strategies or MoUs with e.g. IMI projects, other ESFRI Research infrastructures and e-Infrastructures. ELIXIR will also support "special interest groups" or user communities, i.e. groups of Nodes working together in collaboration with experts in specific research areas such as clinical genomics, plant or fish genomics or integrated agricultural services.

7. Close the computational biology skills gap through a comprehensive training programme for professionals

The goal of ELIXIR's Training Strategy is to upskill European researchers to enable effective exploitation of the data, tools, standards and compute infrastructure provided by ELIXIR. The comprehensive Training Programme of ELIXIR will support professionals throughout their career. ELIXIR will train research and clinical scientists to take advantage of the data produced using rapidly developing technologies. ELIXIR will also meet the skills gap for infrastructure technologists to ensure a strong cadre of professionals that support scalable, sustainable solutions for managing and interpreting data arising from the recent flood of technological advances.

ELIXIR will also align and collaborate with other bioinformatics and life science training efforts such as the global bioinformatics training community in GOBLET. Additionally, ELIXIR will make use of its involvement in EMTRAIN, a pan-European platform supporting training in the biomedical sciences. This enables ELIXIR to work with the other BMS RIs, industry, professional bodies and other course providers, such as the CASyM (Coordinated Action Systems Medicine Europe) project, in the biomedical sciences to build a catalogue of courses, quality standards, and an emerging framework for continuing professional development (LifeTrain).

Ultimately a cohort of trained individuals would themselves become trainers making the scalability and sustainability of these efforts long lasting. The train-the-trainer ethos will cut across ELIXIR training activities.

8. Support innovation in "big-data biology"

Industry usage of many key bioinformatics resources within Europe is high; users of ELIXIR services range from large multinationals to micro-SMEs and cover areas including pharma, biotech, food and agriculture and blue biotech. Given the number of life science researchers in Europe, and the increasing reliance within commercial R&D upon computational methods, ELIXIR has the potential to support more industry users than perhaps any other ESFRI Research Infrastructure. ELIXIR recently performed an analysis of industry needs. The report entitled, "Developing ELIXIR Interactions with Industry" clearly articulates a number of key value drivers for both SMEs and larger companies. Some of the common themes identified were discoverability and interoperability of tools and data-resources as well as opportunities for companies to collaborate around cloud resources to meet the challenge of rapidly growing life-science data volumes.

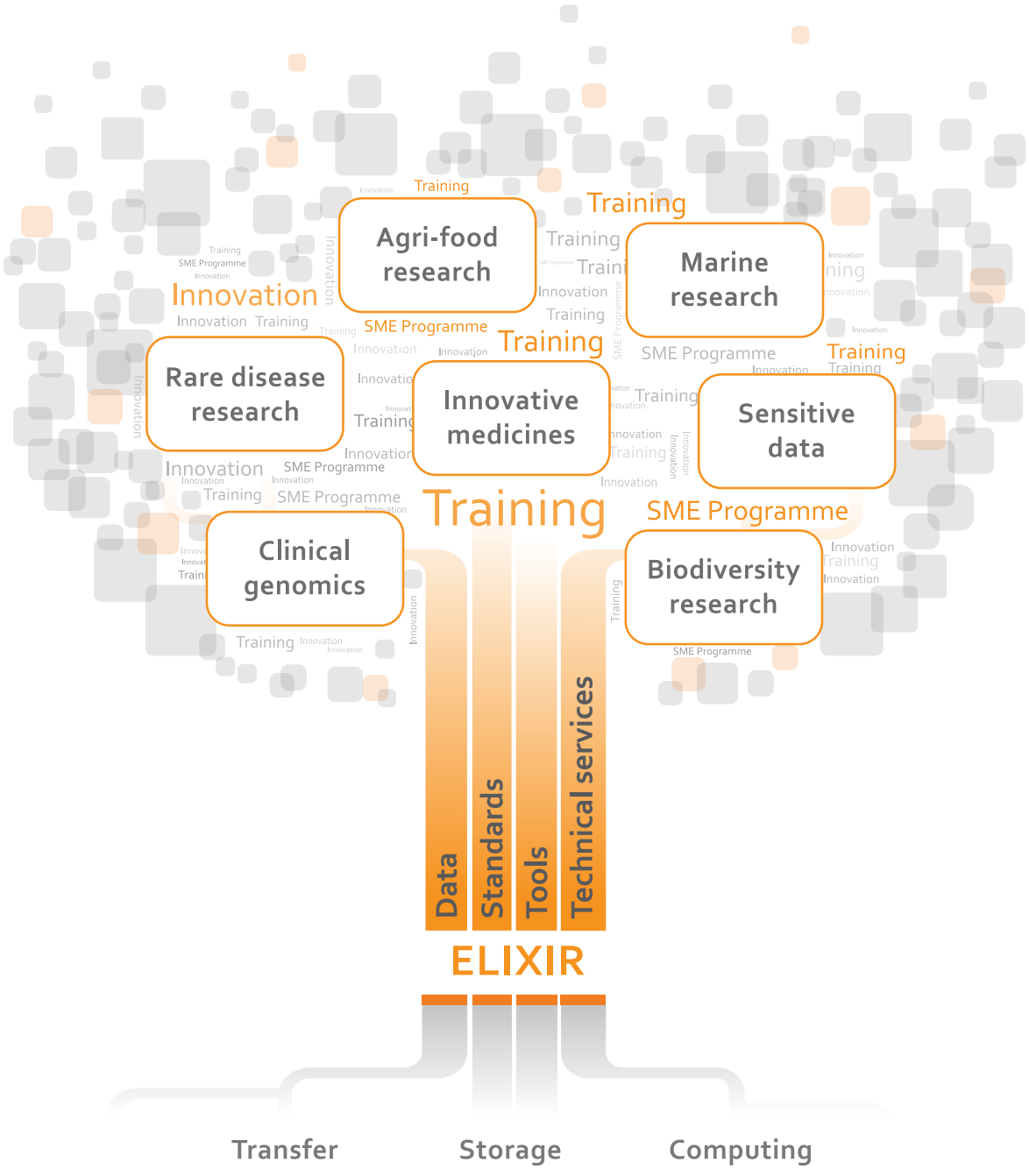
Stimulating innovation and supporting industry is therefore a key objective for ELIXIR. Dedicated posts within the ELIXIR Hub will support various aspects of the ELIXIR Industry Strategy, in particular the Innovation and SME Programme. Additionally, ELIXIR Nodes form a key part of implementing the strategy, not least through hosting ELIXIR Innovation and SME Programme events and through their direct collaboration with industry as well as the local and regional bioclusters.

ELIXIR Programmes of Work

The eight ELIXIR Programmes of Work for infrastructure delivery comprise data resources and services delivered from the ELIXIR Hub and Nodes.

The Programmes bring together experts across Europe to align key ELIXIR resource offerings and to support both the development of best practice and the consolidation of emergent services. They also provide a robust framework for future ELIXIR grant applications and interactions with other research infrastructures and projects.

Each Programme of Work will encompass a core set of mature, stable services as well as a portfolio of ELIXIR pilot actions to ensure ELIXIR's ability to incorporate new technologies and emerging areas of scientific inquiry. Pilot actions will be the basis of ELIXIR's response to the needs of new user communities and novel scientific approaches.



ELIXIR is building a distributed infrastructure that enables our collective integration of biological data, expanding bioinformatics capacity in Europe. Specifically, we aim to achieve:

- Robust access to biological data services in an open, stable and sustainable environment (Sustainable Data Resources & Services);
- A "go-to" environment for users of data, services and tools that provides interoperability, transparency on service levels, sustainability and licensing. (Tools Interoperability and Service Registry);
- Effectively interface with e-Infrastructures and resource providers to ensure robust availability of computable storage, remote data access and transfer, and user identification services that meet the needs of life science research (ELIXIR Technical Services);
- Sustainable standards and services for the effective management, discovery and interoperability of raw and curated life-science research data (Data Interoperability, Vocabulary and Ontology services);
- Close alignment with user communities and other research infrastructures to address the specific needs of different life science domains such as handling of sensitive, identifiable human data in Biomedicine, geo-tagging of metagenomics data or DNA Barcodes for Biodiversity research (Services to Specific User Communities);
- A comprehensive training programme for life science and biomedical professionals that closely aligns and synergises with national training programmes, community led efforts such as GOBLET and EMTRAIN, and other major European initiatives (ELIXIR Training Programme);
- Strong interactions and support to life science SME and industry throughout Europe, particularly through harnessing the strength of Nodes to engage and support local companies (ELIXIR Industry Strategy);
- Effective management and governance within ELIXIR that foster active collaboration and exchange between the Nodes to demonstrate societal benefits and ensure long-term sustainable services (ELIXIR Management and Operations).

Sustainable data resources & services

At its heart ELIXIR is a data infrastructure with the objective to provide transparent access services to biological data in Open Access, stable and sustainable environments and made aware to potential users. Given the rapid growth in biological data and the significant expansion of biological data provision beyond a small number of laboratories and research institutes, the gaps are significant and will require substantial investment.

The Data Resources and Services Programme of Work will deliver:

- A long-term plan to drive the sustainable development of Europe's core data resources;
- Agreed criteria for ELIXIR Core Resources and the process to identify the set of core services, including application of the ELIXIR brand / logo to services;
- Service monitoring, life-cycle management and quality control / peer review of ELIXIR services and data resources. Ensuring effective links with the Programme of Work on Tools interoperability and Service Registry to drive transparency of ELIXIR service offerings;
- A development / enablement path for addition of future core services;
- Robust understanding of benefits and dependencies to drive long-term sustainability of core resources.

The ELIXIR resources and services have to build, strengthen and forge additional international alliances, in order to harmonise efforts, share resources, avoid duplication/ redundancy and maximise effectiveness. Outside of Europe, ELIXIR may collaborate with organisations in North America, the Asia Pacific region or other parts of the world where bioinformatics has a strong foothold. For example, ELIXIR would be a natural partner in resources funded by the US National Institutes of Health or National Science Foundation, such as model organism databases and "Big Data to Knowledge" centres.

Tool interoperability and Service Registry

Discoverability of tools and resources was highlighted as a key challenge in the report on industry needs and expectations. To meet this need, ELIXIR will develop and provide a Service Registry. In addition, biological tools and resources need to meet requirements of increasingly complex analysis pipelines and the interoperability of software, services and underlying resources is often a bottleneck in applied bioinformatics.

Improved interoperability will build on, for example, the ELIXIR pilot project on interfacing the Human Protein Atlas with other proteomics resources.

This Programme of Work will deliver:

- A comprehensive, federated Service Registry providing transparent and effective overview of both ELIXIR and other community services;
- Tools Interoperability within ELIXIR and a solution for interfacing and exchanging workflow-based systems;
- Transparency and benchmarks of the ELIXIR Data Resources and Services through the Service Registry;
- A core set of ELIXIR Named Services;
- The standards processes and tools to drive and increase efficiency in data-resource integration and development (e.g. through efforts such as PSI, DAS, BioJS and IMEx);
- Testing emerging technologies through ELIXIR pilots to ensure that services are relevant.

ELIXIR technical services

Accessing and computing with the large data-volumes from modern biology comes with a specific set of challenges to the technical infrastructure, and the ELIXIR Technical Services Programme of Work will develop solutions of wide-spread utility in the life science domain.

The role of ELIXIR technical services is to provide recommendations and solutions for the present and future specialised life-science data service providers. Services aim to enable storage, transfer, copying, processing and accessing of data. Technical services must be sustainable and scalable so that data specialists can rely and build upon them. ELIXIR technical services will depend on external solutions (e.g., the network provided through GÉANT and its national partners), their proper configuration and specific modifications, the overall architecture and coordination, as well as development of additional services. Architecture, inclusion of services and their modifications and development will be driven by the requirements of the life science community (primarily through ELIXIR partners) to guarantee their uptake and utility.

For instance, there is widespread interest and active development of cloud-based services within the life sciences; in addition to the "Embassy" cloud pilot there is also cloud provisioning from several ELIXIR Nodes. These should be taken as a basis for the ELIXIR data processing services.

Further important aspects of the ELIXIR technical services will be the close collaboration with translational, biobanking and imaging infrastructures at both the European and national level to ascertain that there are effective services to securely access and exchange data. For example, the current ELIXIR pilot on federated authentication enables governance processes and access to personal genome data through institutional logins.

A challenge is to develop sustainable models for resource management and exchange, including transnational models for charging capital and operational costs for resources such as cloud provision.

ELIXIR will collaborate and increase overlap with the European e-infrastructures to address user needs and fully integrate into the European technology landscape. An important part of the Programme of Work will be to define the interface and collaborative models to use technical services from e-Infrastructures and to properly integrate them into the ELIXIR infrastructure.

The ELIXIR technical services Programme of Work will address European data management challenges with the following activities:

- ELIXIR cloud offerings and services;
- Secure data access and exchange;
- Resource management and exchange policies;
- Technology watch and e-infrastructure synergies.

Data Interoperability, Vocabulary and Ontology Services

This Programme of Work encompasses ELIXIR services for semantic interoperability, identifier resolution and mappings and vocabulary and ontology services based on the coordinated development, implementation and deployment across Europe. The Programme of Work encompasses programmatic access, nomenclatures and ontologies as well as the reporting requirements for guiding deposition and facilitating exchange of information (e.g. Minimum Information standards).

[21] *http://www.biomedbridges.eu*

Recognising that the development of standards and practices must be coupled to user-driven demands and strong use cases the ELIXIR Data Interoperability, Vocabulary and Ontology services will focus on defined, time-bound, projects to deliver:

- A community-driven, user-focussed roadmap of current availability and adoption of data stewardship standards and practices, including status, sustainability and usage to support best-practice efforts and provide practical guidance to national infrastructure;
- User focussed data interoperability services for life science research that include major data resources and other critical partners with demonstrated value and impact;
- An identified set of critical standards, rules, vocabularies and ontologies, dictionaries etc. for ELIXIR services and required maintenance resources (e.g. see the BioSharing registry of standards and highly used ontologies);
- A new European architecture for identifier resolution, data citation, provenance, interoperability services and standards, and full use of earlier initiatives.

Services to specific user communities

ELIXIR infrastructure services critically need to be closely aligned and respond to the needs of scientific communities across the broad field of life sciences. In addition, ELIXIR will have the capacity and understanding to develop the standards and systems that projects and other research infrastructures need to adopt to be able ensure their resources are interoperable with those of other projects and infrastructures. To achieve these goals, the set of ELIXIR cross-cutting resources will need additional services to address needs of specific research domains, for example, to protect sensitive human data in biomedicine and health or geo-tagging of environmental data.

This Programme of Work will address major challenges such as:

- The identification of active communities and ways to build positive modes of collaboration with them;
- Collaboration, supporting and/or helping to establish large-scale efforts such as IMI projects and ESFRI BMS research infrastructures;
- Developing an economically feasible and sustainable model for the work of ELIXIR in specific projects and/or with specific communities;
- The identification and implementation of methodologies that can provide solutions applicable across multiple disciplines, maximising synergies between communities and paving the way to stable infrastructures.

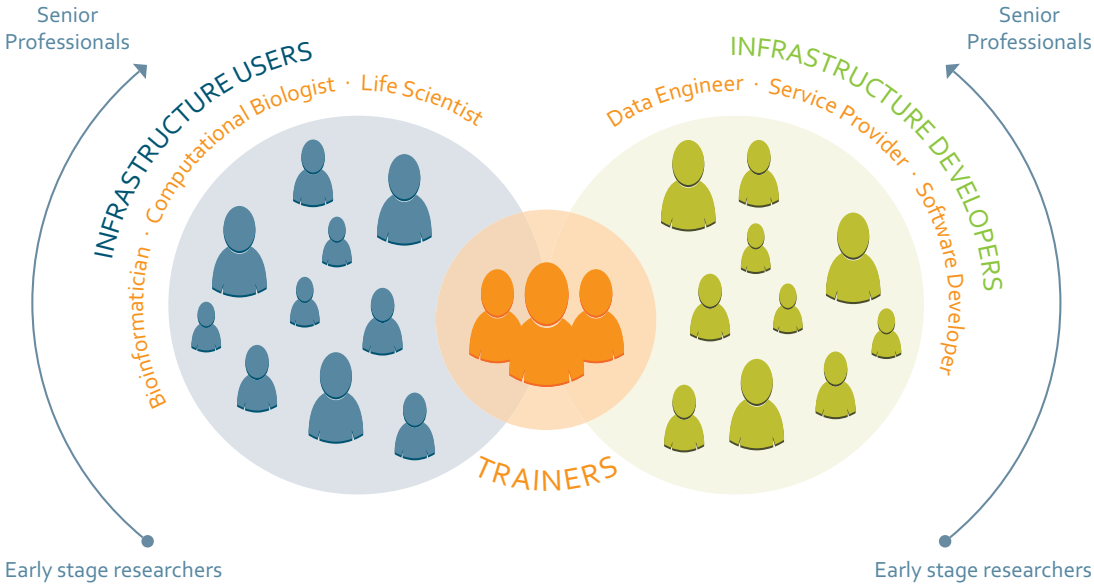
The alignment with core bioinformatics resources and ELIXIR's core methods and systems will be essential. In some cases (e.g. IMI/ESFRI) framework agreements will be established; in others, groups of Nodes working together in collaboration with experts in specific problem areas might be preferable, for example in forestry and marine genomics. This Programme of Work will deliver:

- Strong links with domain-driven initiatives and infrastructures to address interoperability and data services (e.g. IMI and biomedical research infrastructures, Bio-based Industries Consortium(BIC) for industrial biotechnology) by providing standards, ontologies and validation of computational approaches;
- Use-case-driven application and adoption of services for different domains through set of ELIXIR pilot actions;
- Processes to transition successful and critical pilot actions into services that are sustainable over the long term.

ELIXIR training programme

Whilst ELIXIR's broad aim is to facilitate access to biological data for Europe's life science community, the goal of ELIXIR's training strategy is to facilitate accessibility, by up-skilling European researchers who then more effectively exploit the data, tools, standards and compute infrastructure provided by ELIXIR.

ELIXIR's training strategy is tightly focussed on actively supporting and training these users electronically or through face-to-face courses and programmes held throughout Europe. Although this is already a very active area, these efforts lack pan-European coherence. The Training Programme will, in partnership with global efforts such as GOBLET, focus on adequate provision of training to a large and diversifying user base, and the development of a new generation of software engineers, biocurators and other professionals needed to operate ELIXIR. Each of these will need to be underpinned by a European community of trainers who excel in their subject-matter expertise and in their grasp of adult learning and who share a common approach to monitoring and capturing the success and impact of the training.



The dimensions of the ELIXIR Training Programme

The breadth and depth of training experience available across ELIXIR is unparalleled and, if adequately funded, harmonised and coordinated, would provide the expertise required to deliver the following training objectives between 2014 and 2018:

1. Establish a common approach across ELIXIR to benchmark and evaluate training activities;
2. Establish an ELIXIR-wide training community;
3. Disseminate relevant information via a training portal;
4. Define core competencies of users and operators of ELIXIR infrastructure and map these competencies to training activities;
5. Identify training gaps and priorities by engaging with industry and other stakeholders;
6. Expand high-quality training activities–and develop and deliver new ones–in order to fill identified gaps and meet priorities;
7. Synchronise the development of ELIXIR data infrastructures with associated high-quality training materials and activities;
8. Capture and deliver ELIXIR training activities in Open Education Resources (OER), including video and interactive training materials.

ELIXIR industry strategy

One of the clear messages from the ELIXIR industry analysis performed in 2013 was that ELIXIR can play a major role in supporting small and medium-sized enterprise (SME) throughout Europe, particularly by harnessing the strength of Nodes to engage and support local companies
An ELIXIR Innovation and SME Programme would comprise a series of SME-focussed outreach and training events in the Member States, hosted by ELIXIR Nodes. These would feature:

- Attendance of 30 local SMEs at each event, ensuring the hosting ELIXIR Node can build and maintain lasting links with local SME users;
- Resourcing through the ELIXIR Hub budget, with some contribution from ELIXIR Nodes;
- Event programmes that include training or workshops to SMEs, or information on relevant resources and services across the ELIXIR network;
- An agenda developed by ELIXIR Nodes and Hub and tailored for local interests, i.e., pharma, biotech or agri-food.

Provision to roll out the ELIXIR Innovation and SME Programme has been allocated to the proposed budget for 2014. Limited initially to a small number of events in 2014, the Programme could be scaled to take place more regularly once it is operational.

As well as harnessing the strength of the Hub and Nodes structure, the Programme would also seek to ensure

[21] http://www.biomedbridges.eu

complementarity with the industry programmes already being run by ELIXIR Nodes, adding value where possible and ensuring no unhelpful duplication emerges.

ELIXIR management and operations

This Programme of Work addresses the on-going development of ELIXIR strategy and sustainability as well as work streams to develop and improve management and governance within ELIXIR, based on the feedback from SAB and the Assessment Expert Group. In addition, ELIXIR needs to foster the active collaboration and exchange between the National Nodes. This Programme of Work will address the development of an active exchange programme for scientists and engineers involved in the delivery of ELIXIR services.

The ELIXIR Business Case clearly outlines the case for maintaining and strengthening Open Access to biological research data, as charging or restricting access to data would seriously limit the ability of public and private research organisations to exploit and create additional value from the collective research investments. Indeed, a strong argument for Open Access is the example of discoveries getting lost in legal red tape due to the difficulty of integrating data and making it interoperable across a complex web of licenses and contractual limitations.

Intellectual Property

While Open Access to biological research data is the default position for ELIXIR there are complicating nuances: When do ELIXIR resources support embargo of data? How does ELIXIR support value creation, for example through Intellectual Property? In addition, several global data resources have recently introduced user charges to support the long-term sustainability of resources and there are of course, significant costs associated with data deposition and metadata annotation. This was also noted by the ELIXIR SAB: it recommended a work stream to thoroughly analyse ELIXIR's policy and the development of a clear Charter and recommendations for Open Access and Intellectual Property Policy.

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 ELIXIR Services. This is an area partly addressed through the work on Data Security as well as Ethics policy across biomedical research infrastructures within the BioMedBridges project. As joint recommendations are developed, ELIXIR will incorporate these into the ELIXIR Policy on user charges, Open Access and Intellectual Property. The AEG also recommended ELIXIR to establish an Ethics Board for formal governance and advice.

The key objectives for this Programme of Work are to:

- develop ELIXIR operations, performance indicators and impact analysis;
- develop access and Open Data policy, Software Licence policy and Intellectual Property policy;
- define ELIXIR user strategy and benchmark infrastructure reach and effectiveness;
- develop the ELIXIR external funding strategy;
- develop and implement the ethical framework for ELIXIR.

ELIXIR Pilot Actions

ELIXIR pilot actions are short, time-bound projects that address key scientific and technical issues. Pilots play an important role in the implementation of new concepts and form a key part of the lifecycle for ELIXIR services and resources.

The ELIXIR pilots should leverage or join-up established resources or on-going activities to deliver demonstrator or proof-of-concept studies. ELIXIR will rely on pilots as a user-centred mechanism for testing new ideas and allowing the ELIXIR Technical Strategy to adapt to changes in underlying technologies. The pilot actions will additionally enable the adoption of services and standards in different life-science domains. This close connection with the users through ELIXIR Nodes and their connections with the ESFRI BMS infrastructure will ensure that the adapted services and standards are driven by the user needs. The pilot actions also provide for a mechanism to demonstrate ELIXIR services in the absence of formal Collaboration Agreements during the early phases of ELIXIR operations. Finally, if an ELIXIR technical service pilot reaches maturity, it can become an ELIXIR commissioned service in line with strategic priorities and availability of long-term funding.

So far, there have been five ELIXIR pilot actions in the areas of interoperability, data access and exchange and cloud compute on big data. For example, the completed ELIXIR pilot on federated authentication tested how access to services may be established using an external identification system. This pilot was a collaboration between EMBL-EBI and the Finnish ELIXIR Node and it demonstrated the feasibility of accessing the European Genome-phenome Archive (EGA) services at EMBL-EBI using the Finnish Haka identity. The pilot proved to be successful as well as useful and is now adopted in the production version of the EGA services. In addition, this project triggered a new pilot action to evaluate how the level of trust of different identification systems within the eduGAIN federated identity may be assessed. With this new pilot ELIXIR expects to demonstrate how access to content can be simplified and trustworthy exchange of information among ELIXIR services may be established.

Principles and process for starting new pilot actions

ELIXIR pilot actions are agreed by the ELIXIR Director and the Heads of Nodes. This group will accept the pilot actions based on scientific merit and fit into the ELIXIR strategy and portfolio.

The Director will bring forward proposals to the Heads of Nodes via an email procedure to seek an endorsement by the Heads of Nodes Committee by a set date (minimum 5 working days) where a non-response will count as approval. If pilot actions do not require additional resources from the Hub, they will be forwarded directly to Heads of Nodes for a decision based on scientific merit and fit into ELIXIR portfolio. For pilot actions that require resources from the ELIXIR Hub, the Director will prepare a prioritized list based on the following criteria (in order of importance):

1. Portfolio balance and fit to ELIXIR strategy; pilot actions should give a broad cover to the key areas for ELIXIR and respond to emerging user needs;
2. Delivery date, to prioritize pilot actions such that they can support the development of ELIXIR Programme and H2020 applications;
3. Pilots that leverage on-going / delivered pilot actions, i.e. focussed activities that could broaden the utility and/or involve additional Nodes, will be prioritised;
4. Pilots involving new Nodes, Pilot Actions are an important mechanism to develop collaborative links within ELIXIR and hence priority will be given to proposals that involve Nodes that previously have not been involved.

A key criterion for a pilot action is inclusiveness; i.e. the pilot should be open to participation and learning for other Nodes. ELIXIR pilots are not competitive research projects but joint efforts to develop strategies or test-drive technology to inform and shape future ELIXIR services. Thus, while a pilot may be resourced only at a couple of Nodes, workshops and meetings around the pilot to discuss technology or future developments must be open to participation by all Nodes.

ELIXIR funding strategy

The ELIXIR infrastructure rests on member state funding that provides long-term foundations for the ELIXIR Hub and Nodes. ELIXIR members will also apply for additional external funding through EU-level programmes and this will contribute to the operation of services across the infrastructure.

Construction and operation of the ELIXIR Nodes

Through its distributed structure, ELIXIR relies on the stable and high-quality provision of bioinformatics resources by its national Nodes. The funding for ELIXIR Nodes come from the ELIXIR Member States and in EMBL-EBI's case from EMBL; typically Nodes are funded through a range of sources including national and international grant-based funding, direct funding from research councils and funding bodies and in some cases EU Structural Funds. In many instances, ELIXIR Member States have also injected 'new' money into the ELIXIR Node in order to build up capacity and develop ELIXIR-specific services.

Construction and operation of the Hub

The ELIXIR Hub resides in the EMBL-EBI South Building, which along with a lease for off-site data storage at the London Data Centre, has been funded through an award made by the UK's Large Facilities Capital Fund. There are therefore no additional construction costs for the Hub. The on-going operating costs of the ELIXIR Hub, for staff, consumables, technical activities and overheads, are met by the participating Member States based on their relative Net National Income (NNI).

Collective applications for external funding

Collectively, ELIXIR Members will apply for additional external funding from sources such as Horizon 2020 and the Innovative Medicines Initiative (IMI). These grants will help to contribute to implementation of parts of the ELIXIR Programme, and will cover activities such as the integration of services, building the shared e-infrastructure, enhancing user access, training, collaborations with industry and building bridges between the ESFRI Biological and Medical Sciences Research Infrastructures. The opportunity of collectively applying for additional funding is noted as a key benefit for participation in ELIXIR.

Horizon 2020 and the Innovative Medicines Initiative (IMI)

Horizon 2020 presents ELIXIR with significant opportunities to leverage EU funding. Over the course of the seven-year programme, topics of relevance to ELIXIR partners are likely to appear in the following Horizon 2020 Work Programmes:

- Research infrastructures, including e-infrastructures;
- Health and demographic change and well-being;
- Food security, sustainable agriculture, marine and maritime research, and the bio-economy;
- Information and Communication Technology (ICT).

The second IMI will also present topics in which ELIXIR may engage. Data management is a transversal theme that cuts across the EFPIA Strategic Research Agenda, presenting ELIXIR partners with scope to engage in many fields. Participation in IMI grants would also have the advantage of allowing ELIXIR members to build or maintain close links with key industry players.

Over the course of the ELIXIR Scientific Programme, the ELIXIR Hub will explore various other options for suitable external funding ranging from new public–private partnerships such as Bio-based Industries (BIC) Consortium, opportunities for staff exchanges between Nodes through Horizon 2020 Twinning and Teaming schemes and capacity building in Nodes through better use of EU Structural Funds. The ELIXIR Hub will seek to identify suitable topics and regularly inform Heads of Nodes of these opportunities as they appear.

[21] http://www.biomedbridges.eu

ELIXIR's external funding strategy

Given the potential scope for harnessing Horizon 2020 funding, and the value this can bring to ELIXIR members, ELIXIR's external funding strategy is considered a priority aspect of the ELIXIR Programme.

Specific roles within the ELIXIR Hub (Table 1) are required to support the whole project application life cycle, which includes:

- influencing and shaping the funding landscape for sustainable data management;
- identifying forthcoming and current opportunities;
- assisting with the coordination of joint ELIXIR applications and preparation of materials for ELIXIR Nodes;
- management of selected grants .

This is further strengthened by the AEG recommendation to perform a thorough analysis of the funding strategy under external leadership, which will be initiated as part of the ELIXIR Action Plan.

ELIXIR External Funding Strategy cycle

Funding strategy	Role of ELIXIR Hub	Role of ELIXIR Nodes
Positioning	• Engaging key stakeholders in Brussels – see "stakeholder engagement" in Table 2 • Drafting of ELIXIR position papers and consultation responses	• Supporting role
Identifiying	• Collating input from partners and other sources about new opportunities • Communicating these opportunities with Heads of Nodes	• Sharing details of opportunities with Hub
Submission	• Coordinating process of submitting ELIXIR-named applications • Organising logistics around planning meetings • Provision of certain general text and briefing documents for ELIXIR applications • Build and maintain database of ELIXIR submissions	• Agreeing topics that ELIXIR applies for • Coordinating some ELIXIR applications • Leading Work Packages on ELIXIR applications
Management	• Management of some ELIXIR grants • Dissemination of relevant information on grant management to ELIXIR partners	• Management of some ELIXIR grants

Table 1. The various roles and responsibilities of the Hub and Nodes in implementing the ELIXIR funding strategy

ELIXIR Governance and Organisation

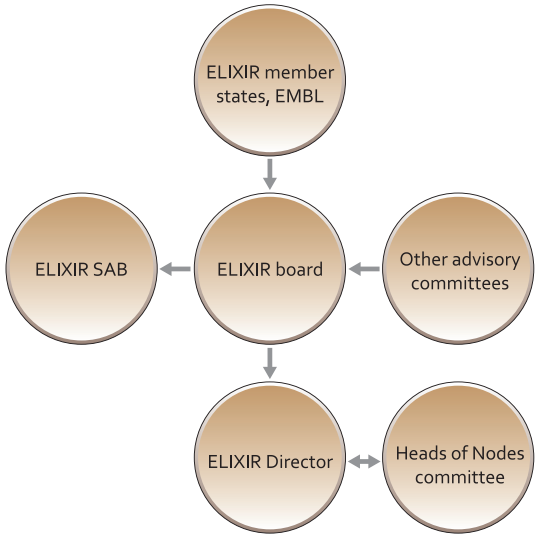
The legal framework of ELIXIR is based on the ELIXIR Consortium Agreement (ECA),[22] which covers ELIXIR's mission, membership, obligations of the Members and the ELIXIR Hub, the relationship between the Hub and Nodes and the governance structure of the Hub itself. Based on the ECA, EMBL carries out such activities that require legal personality.

The ELIXIR Board will monitor the effectiveness of the ELIXIR legal framework and will consider establishing a dedicated working group towards the end of this Programme time-frame to investigate an alternative legal framework based on the experience from the first years of operation.

Key bodies

ELIXIR's governance model takes into consideration ELIXIR's scientific and technical structure. Here, a summary of the governance model is provided; the authoritative source is the ECA.

The ELIXIR Board is ELIXIR's strategic decision-making body and the ELIXIR Member States and EMBL are represented in the ELIXIR Board. The ELIXIR Board appoints and dismisses the ELIXIR Director who is responsible for executing board decisions. The ELIXIR Board also appoints the members of the ELIXIR Scientific Advisory Board (SAB), which advises the ELIXIR Board and the ELIXIR Director in scientific matters. Furthermore, the Heads of Nodes Committee, comprising representatives of the ELIXIR Nodes and EMBL-EBI, gives advice to the ELIXIR Board and Director in relation to ELIXIR activities.



ELIXIR Governance model.

In addition to these committees, the ELIXIR Board may establish additional advisory bodies and form working groups where this is necessary for the proper functioning and achievement of ELIXIR's goals, e.g. an Industry Advisory Committee or an Ethics Advisory Committee.

ELIXIR Board

- As the strategic decision-making body of ELIXIR, the ELIXIR Board provides high-level oversight and approves the ELIXIR strategy and budget (including managing additional bilateral agreements). It oversees management of risks and liabilities of ELIXIR. One of the main tasks of the ELIXIR Board is the approval of the ELIXIR budget and ELIXIR's Programme. Furthermore, the Board establishes and oversees rules and procedures, including the technical rules for the selection of ELIXIR Nodes. In this process, the ELIXIR Board draws on advice from the SAB.
- The ELIXIR Board is composed of the scientific and administrative representatives from each ELIXIR Member State and EMBL, who are authorised to deliberate, negotiate and decide on behalf of the respective Members. Each Member has one vote in the decision-making at ELIXIR Board level. The ELIXIR Board elects a Chair and two vice-chairs. The ELIXIR Board may also decide to admit Observers from prospective Member States, which are willing to contribute to ELIXIR and the Board may invite representatives from charities and industry.
- The Board will meet in person at least once per year, but normally biannually in the spring and during the autumn. The ELIXIR Board meetings will take place at the ELIXIR Hub in Hinxton unless otherwise decided by the Board. The Board will additionally have teleconferences when required.

[22] http://www.elixir-europe.org/system/files/documents/elixir_consortium_agreement.pdf

ELIXIR Director

- The ELIXIR Director is appointed by the ELIXIR Board to manage and administer ELIXIR and the ELIXIR Hub staff.
- The Director is responsible for implementing the decisions of the Board. In addition, the Director prepares the annual budgets for ELIXIR and presents the Board with annual financial and scientific reports. The Director also negotiates and prepares Collaboration Agreements with Nodes.

ELIXIR Scientific Advisory Board (SAB)

- The SAB provides independent consideration and advice to the ELIXIR Board and ELIXIR Director 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 membership of the ELIXIR SAB is made up of distinguished international experts including academics, representatives of other internationally renowned organisations, and representatives of researchers in the commercial sector. This group of independent experts are appointed in their own right by the ELIXIR Board for a period of three years, following the proposal of candidates by the ELIXIR Director after consulting with the ELIXIR Members.
- The SAB elects a Chair and one Vice Chair and meets twice a year. The SAB reports directly to the ELIXIR Board and ELIXIR Director on the results of their scientific reviews of ELIXIR activities.

Heads of Nodes Committee

- The Heads of Nodes Committee is the key body to develop the ELIXIR Scientific and Technical Strategy. In addition to the formal requirement on consultation for the Programme set forth in the ECA, it is expected that this governance body will take the leading role in developing the strategy for ELIXIR services, monitoring of performance as well as identification of service gaps. The Heads of Nodes Committee will also take a key role in developing and preparing grant proposals for ELIXIR.
- The Committee is composed of the Heads of each ELIXIR Node and of EMBL-EBI. The ELIXIR Director leads the Heads of Nodes Committee;
- The Committee will meet biannually such that one meeting will take place in the Hub and the second meeting at the one of the Nodes. In addition the Committee will hold teleconferences as required.

Technical Coordinator Group

- The Technical Coordinator Group is formed as an advisory body to the Heads of Nodes Committee. This group consists of technical experts representing each ELIXIR Node, appointed by the Head of Node. The role of the group to share information and discuss the technical and scientific aspects of ELIXIR as well as lead the identification and implementation of best practice.
- The ELIXIR Chief Technical Officer will act as the Chair of the group and report the findings of the group to the Heads of Nodes Committee.

ELIXIR Ethics Advisory Committee

- The High Level Assessment Expert Group recommends ELIXIR to establish an external Ethics Advisory Committee. During the Co-ordination Phase, the ELIXIR Board will be asked to establish a working group to draft the ELIXIR Ethics Policy and consider the establishment of an ELIXIR Ethics Advisory Committee.

ELIXIR Industry Advisory Committee

- A key finding from the industry user and stakeholder analysis as outlined in the report "Developing ELIXIR Interactions with Industry" was the recommendation to establish a high-level Industry Advisory Committee to advise ELIXIR. During the Co-ordination Phase, the ELIXIR Board will be asked to set up an ELIXIR Industry Advisory Committee.

The ELIXIR infrastructure and services

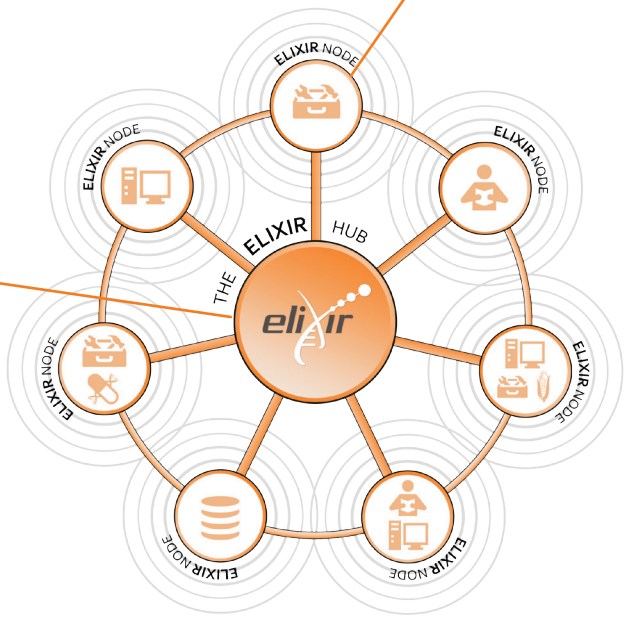
The ELIXIR Hub is responsible for the organisational, technical and coordination interactions with the Nodes and the other ESFRI BMS infrastructures thus ensuring seamless access to biomedical research data, services and tools to life scientists in fundamental as well as applied industrial research. In addition, the Hub is responsible for the day-to-day operational, financial and administrative management of ELIXIR in accordance with the decisions by the ELIXIR Board. This includes organisational support to ELIXIR's governance bodies. The ELIXIR Nodes are located at centres of excellence in the ELIXIR Member States. They will enter into a Collaboration Agreement with the ELIXIR Hub for the provision of technical services with a European or global dimension that have an added value for ELIXIR.

ELIXIR Node

- Research and development of bioinformatics services
- Delivery of services through own "brands"
- Management of core resources
- Develop and deliver training activities
- Participation in international data consortia
- Industry collaboration and support

ELIXIR Hub

- ELIXIR management and secretatiat
- Technical coordination across Nodes
- Drive standards development and implementation
- Policy and outreach
- Lead coordinated infrastructure investments
- Deliver services



The summary of roles of the ELIXIR Hub and Nodes.

The Collaboration Agreements

The Collaboration Agreements form a central part of the ELIXIR governance. In 2013, a Working Group to draft the ELIXIR Collaboration Agreement template was established and the resulting template Collaboration agreement was agreed by the ELIXIR Board. A key priority for the early phase of ELIXIR is to sign Collaboration Agreements with the individual Nodes and develop the accompanying Service Delivery Plan and Commissioned Services Contracts. For this purpose, a number of work streams will be established with an additional task to define Key Performance Indicators (KPIs) to monitor the progress of each of the Service Delivery Plans. The ELIXIR Hub will continue to support the process by providing legal and technical expertise. By the end of 2014 it is expected that the first set of Collaboration Agreements will be concluded.

Collaboration Oversight Group

- Upon the formation of the Collaboration Agreement between an ELIXIR Node and the Hub, a Collaboration Oversight Group is formed. It serves as a forum for a regular exchange of information and joint project coordination, monitoring the tasks, responsibilities and delivery of services of the ELIXIR Node and the respective KPIs as defined in the Collaboration Agreement;
- The Collaboration Oversight Group comprises the ELIXIR Director and the Head of Node and other individuals appointed by them, e.g. representatives of the distributed national network or experts on relevant issues;
- The Collaboration Oversight Group will meet at least once a year and these meetings can also be conducted as teleconferences.

Realising the Benefits of ELIXIR

The biosciences are driving innovation. Biological research addresses fundamental and urgent global challenges such as the ageing population, environmental degradation, and dwindling supplies of food and fuel. As we look to the biosciences to help resolve these challenges, the potential for job creation is enormous. Bioscience industries will contribute significantly to the financial and social well-being of those countries that have invested in these critical, value-adding sectors.

ELIXIR will provide the backbone for new discoveries that address and meet the Grand Challenges and, in doing so, will help:

- 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 sized 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 through a sustained set of core resources.

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.

For a number of countries, participation in ELIXIR will serve to support capacity building in the sphere of bioinformatics. In countries with limited budget for science, or where existing bioinformatics resources have a more limited international use, participation in ELIXIR will enable organisations to improve their expertise in bioinformatics through participation in the development of best practise and access to ELIXIR's professional Training Programme. ELIXIR's independent SAB will also support this process through the provision of valuable advice and suggestions in relation to deploying world-leading bioinformatics services across all Nodes. Staff trained to handle and exploit the biodata are keenly sought in industry and will help to develop local companies in this area of science.

The recommendations from the Assessment Expert Group report emphasise that the added value of the European Research Infrastructure "should result from improved services and quality standards on the science mission, which cannot be reached at national level". Considering the global nature of the bioinformatics data resources, training and tools that ELIXIR will provide, it will also be important to demonstrate the global impact of ELIXIR. Key Performance Indicators (KPIs) will provide ELIXIR with a tool to define the project-wide criteria for success. Through an analysis of the KPIs, it is possible to identify the likelihood to achieve reasonable impact, if the KPIs are credible and include target values where appropriate.

The Magenta Book[23] published by HM Treasury in the UK presents standards of good practice in evaluations of projects, policies, programmes and the delivery of public services. These guidelines are useful for ELIXIR when defining the KPIs that measure the societal benefits of ELIXIR and demonstrate the added value of the research infrastructure. The Magenta Book recommends considering the evaluation criteria as early in the Programme development cycle as possible, to ensure that "the most appropriate type of evaluation can be identified and adopted". Only evaluations of adequate quality will provide reliable results that are useful for demonstrating success. Appropriate evaluations will show if the policies are delivering as planned and if the allocated resources are used effectively.

It is a key role of the ELIXIR Hub to develop, implement and monitor ELIXIR-wide KPIs. This exercise will also have direct benefit for ELIXIR Nodes, helping to show national funding bodies the impact their investments have made.

[23] *https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/220542/magenta_book_combined.pdf*

ELIXIR users

ELIXIR must serve the needs and priorities of a very complex community of users. This was highlighted clearly during the Preparatory Phase, when ELIXIR performed an extensive user survey with 804 respondents representing 318 organisations from 34 different countries.[24] The results from this survey indicate that one of the challenges is the differing demands of the two identified, discrete user classes: the "power users" (with the skills to customise resources and tools) and "non-power users" (or end-users, who rely on graphical user interfaces). The user survey further demonstrated that on top of general requirements for genetic and molecular information, respondents' interests in other biological (or biology-related) data were very specific. Hence, ELIXIR should provide an environment that acknowledges the diversity of users' interests. It is important to realise, however, that this survey had one user group – the private sector users – severely under-represented (89.4% of respondents were from the academic/non-profit sector).

Building from the Preparatory Phase user survey, ELIXIR's users can be divided into a number of different constituencies that all have different requirements. Biology researchers in Europe and beyond, both in academia and industry constitute the largest user group. There is a growing need for services that cater to individual researchers and embedded bioinformaticians in the core facilities with the biggest bottlenecks being service discoverability and interoperability of data to support complex data sets. In addition, there is a demand for an optimal community synergy between resource providers and users, e.g. involvement of future users during the resource development phase and most importantly efficient capture of users' feedback information. The industry users, which are a major user group of ELIXIR, have very similar requirements to academic users.

Users from different life science domains may have very specific requirements on services that they need. For example, scientists working with medical research data need services that are able to handle sensitive data to protect the privacy of patients. The first successful service in this area was the ELIXIR pilot action on Authentication and Authorization Infrastructure to manage access rights to research resources. The pilot was led by ELIXIR Finland in collaboration with the European Genome-phenome Archive (EGA) and with a complex data access committee of the Nordic Control Database (NCDB). ELIXIR's Technical Strategy, together with the domain specific services, will work towards providing solutions for these users.

Training will be integral for users to be able to gain full benefit from the services that ELIXIR provides. This is particularly true in the field of bioinformatics, where the explosion in volume and complexity of the generated data requires a major increase in the number of trained professionals. This includes users of the services which, for ELIXIR, range from clinicians and wet lab scientists to bioinformaticians, as well as more trained bioinformaticians to run services. Therefore, ELIXIR is developing its Training Strategy to ensure collaboration on training programmes throughout the infrastructure.

Catering to the user needs of large project consortia and complex research projects require services from several ESFRI Research Infrastructures, although data will need to be joined up and linked regardless of source. Although the provision of the underlying storage architecture is out of scope for ELIXIR, there is a central role for a European infrastructure in the setting and maintenance of data standards. ELIXIR will be directly interfacing with data coordination centres of e.g. genome variability projects to ensure effective processes for sustainable deposition of data.

Similarly, there is currently significant investment at the European level in a drug-discovery-focussed translational infrastructure through the IMI Knowledge Management projects and although there has been initial discussion across the IMI projects, ELIXIR and other European initiatives (e.g. through the FP7 InBioMedVision project),[25] there is a significant need to establish a long term strategy for sustainable medical knowledge management and translational infrastructure. ELIXIR is expected to play a

ELIXIR provides bioinformatics services to life sciences researchers funded through EU R&D programmes, enabling EU researchers to carry out world-class research. ELIXIR also manages some of the data generated through these programmes.

[24] *http://www.elixir-europe.org/prep/sites/elixir-europe.org.prep/files/documents/reports/elixir_usersurvey_finalreport.pdf*
[25] *http://www.inbiomedvision.eu/*

key role in developing this strategy and will need to engage intensely with e.g. IMI projects as part of the strategy development. Other new initiatives may emerge, such as an infrastructure for rare disease informatics resources, and ELIXIR will want to harness the strength of Nodes to ensure that it is best placed to engage in these emerging initiatives.

The AEG report highlighted the large and growing user community of ELIXIR as one of strengths of the research infrastructure. In order to ensure that the user base keeps expanding and that the users are provided with improved bioinformatics facilities, ELIXIR will update the user survey and closely align with user communities.

ELIXIR stakeholders

The ELIXIR stakeholder group is very heterogeneous and reflects the landscape in which ELIXIR operates. Naturally, for any research infrastructure, the key stakeholder group is the user community of the infrastructure, and these have been separately discussed above. The focus here is on other external stakeholders of ELIXIR, and these can be broadly categorised into five groups:

- Funding bodies
- Policy makers
- Industry
- Other infrastructures, e-infrastructures, and research initiatives
- General public, media and press

In Table 2 (opposite), ELIXIR's external stakeholder groups are listed with a description of each group and the main purpose of engaging them. This is not an exhaustive list and more stakeholders are expected to be identified over time. It is important to emphasise that while the role of the Hub is to ensure good communication with the other research infrastructures, e-infrastructures and global bioinformatics initiatives, the Nodes will have an equally important role to ensure that on the national level the communication with the other research infrastructures is effective.

ELIXIR will play a role underpinning all of these research infrastructures by being able to manage data generated on these infrastructures, if required, and by hosting important reference data.

The Assessment Expert Group recommended establishing a project-wide communication plan. This should include details on the communication strategy involving both external and internal stakeholders. The distributed structure of ELIXIR calls for a well-considered communication strategy, as there is a risk of not delivering the objectives due to lack of communication between internal stakeholders. The ELIXIR Communication Strategy will further define who within the infrastructure is responsible for meeting the needs of each stakeholder.

Stakeholders	Role of ELIXIR Hub	Role of ELIXIR Nodes
ELIXIR Scientific Advisory Board (SAB)	Ensure that the SAB has all the relevant information about ELIXIR to be able to perform its mission.	The ELIXIR Board elects the SAB members following nominations from Nodes.
European Commission (EC)	Ensure that ELIXIR is represented appropriately in discussions with the EC so that we can influence Calls and help secure funding.	Attend meetings representing ELIXIR.
DG CONNECT	Ensure that DG CONNECT sees ELIXIR as an integral part of its own e-Infrastructure ecosystem and that this can lead to funding to help construct ELIXIR's technical activities.	Attend meetings where appropriate, bring issues to the attention of the ELIXIR Hub / Community.
DT RGD	Feed into DG RTD-led policy developments so that ELIXIR and bioinformatics is recognised appropriately in EC policy formulations and documents, and that ELIXIR can respond to Calls for Proposals.	Attend meetings where appropriate, bring issues to the attention of the ELIXIR Hub / Community.
ESFRI	Ensure that ELIXIR is aware of the latest developments within ESFRI Forum and ESFRI Working Groups such that ELIXIR can respond to emerging developments. Ensure that ESFRI Member States are aware of the progress being made by ELIXIR and seek to engage further.	Maintain links with national ESFRI delegates.
ESFRI Biological and Medical Sciences RIs	Ensure that ELIXIR maintains strong political, scientific and technical links with the other ESFRI BMS Research Infrastructures, and ensure that on common issues concerning data, ELIXIR is considered the lead ESFRI RI.	Represent ELIXIR in interactions with national BMS RI Nodes. Keep the Hub informed of the developments.
Member State led Joint Programming Initiatives (JPIs)	Ensure that ELIXIR Member States understand that ELIXIR Nodes can manage data that are generated by JPI-funded initiatives.	Act as a bridge with the national ministries developing JPIs. Keep the Hub informed of the developments.
European Parliament (EP)	Ensure that key MEPs are aware of ELIXIR and our mission, and that ELIXIR can use these links to influence key issues, programmes and policies being discussed within the EP.	Attend relevant meetings.

Table 2. A non-exhaustive list of ELIXIR external stakeholders. The users of the research infrastructure are the most important stakeholder group and discussed separately in the section above.

Stakeholders	Role of ELIXIR Hub	Role of ELIXIR Nodes
Industry, IMI and EFPIA	Engagement with EFPIA will help raise the profile of ELIXIR with European pharma companies and the respective national pharma associations. Ensure that the ELIXIR Industry Strategy is understood by industry and that their needs are taken into account by ELIXIR partners when services are being developed.	Direct involvement in R&D collaborations with industry, host industry events locally, communicate to the Hub on new services and news.
e-Infrastructure community	Ensure that ELIXIR needs are taken into account in Europe's expanding HPC landscape, such as GEANT and the DANTE delivery authority, the European Grid Infrastructure (EGI), EUDAT, PRACE, ETP4HPC and e-IRG.	Active involvement by the Nodes, e.g., the Danish Node leads on the ETP4HPC interactions.
The Research Data Alliance	Ensure that ELIXIR and partners are seen as the partner for life sciences.	Active involvement by the Nodes, i.e. the Swedish Node leads on the RDA interactions.
Global Alliance for Genomics and Health	Ensure that ELIXIR's views on data sharing and security are taken into account in Global Alliance, and ensure that ELIXIR is fully aware of the developments.	With support from Nodes, i.e. EMBL-EBI is also a member of Global Alliance.
Countries interested in joining ELIXIR	Ensure that ELIXIR expands and the number of Member States increases. This includes Member States of the Council of Europe and other states internationally, which can be declared eligible by decision of the ELIXIR Board.	Support emerging regional partners through e.g. Node visits and discussions.
Other policy initiatives, e.g. Science Europe, G8 Group of Senior Officials (G8-GSO), OECD	Ensure these initiatives are aware of and have the relevant information to consider ELIXIR as best practise.	Ensure that the ELIXIR Hub / Community is aware of relevant initiatives and developments of science policy.
Other research initiatives, e.g. Global Biodiversity Information Facility (GBIF), Human Brain Project (HBP)	Ensure these initiatives are aware of and have the relevant information about ELIXIR to adopt ELIXIR as the best practise.	With support from Nodes once ELIXIR's engagement is being defined.

Stakeholders	Role of ELIXIR Hub	Role of ELIXIR Nodes
Technical initiatives, e.g. PSI that work towards defining standards; data service providers outside of the ELIXIR community	Ensure these initiatives are aware of efforts within ELIXIR and that ELIXIR is aware of their work to enable alignment of strategies globally.	Ensure that the ELIXIR Hub / Community is aware of relevant initiatives and developments in these.
General public, press and media	Ensure that these stakeholders are aware of ELIXIR, its mission, goals and achievements.	Align communication strategy with the Hub.

The Research Data Alliance implements the technology, practice, and connections that make data work across barriers. The role, scope and influence of RDA are likely to increase over the coming years and there is a real opportunity for the life sciences data community to engage. ELIXIR is a member of the Global Alliance, a large-scale, international effort to enable the secure sharing of genomic and clinical data. ELIXIR needs to be engaged in such international policy initiatives – both existing and emerging ones – to be able to assess the level of communication that is required. For some of the initiatives and stakeholders, the Nodes are expected to lead the ELIXIR communication. These roles will be defined in the ELIXIR Communication Strategy that will be developed during the Co-Ordination Phase and implemented from the Service Deployment Phase starting in 2015.

Communication with the industry stakeholders will be a key role of both the Hub and the Nodes, while the Hub will coordinate the Industry Strategy. Interactions with IMI and EFPIA are an important role of the Hub to ensure that ELIXIR is best placed to feed into the development of EFPIA's Strategic Research Agenda, and can thus shape the development of topics in future IMI Calls.

ELIXIR Hub

EMBL-EBI South building
Hinxton, Cambridge, UK

***ELIXIR** is building a sustainable European Infrastructure for biological information, supporting life science research and its translation to:*

Medicine

Environment

Bioindustries

Society

Contact

Niklas Blomberg, Director
ELIXIR
Wellcome Trust Genome Campus
Hinxton, Cambridgeshire
CB10 1SD, United Kingdom

+44 (0)1223 492 670
+44 (0)1223 494 468
info@elixir-europe.org
www.elixir-europe.org