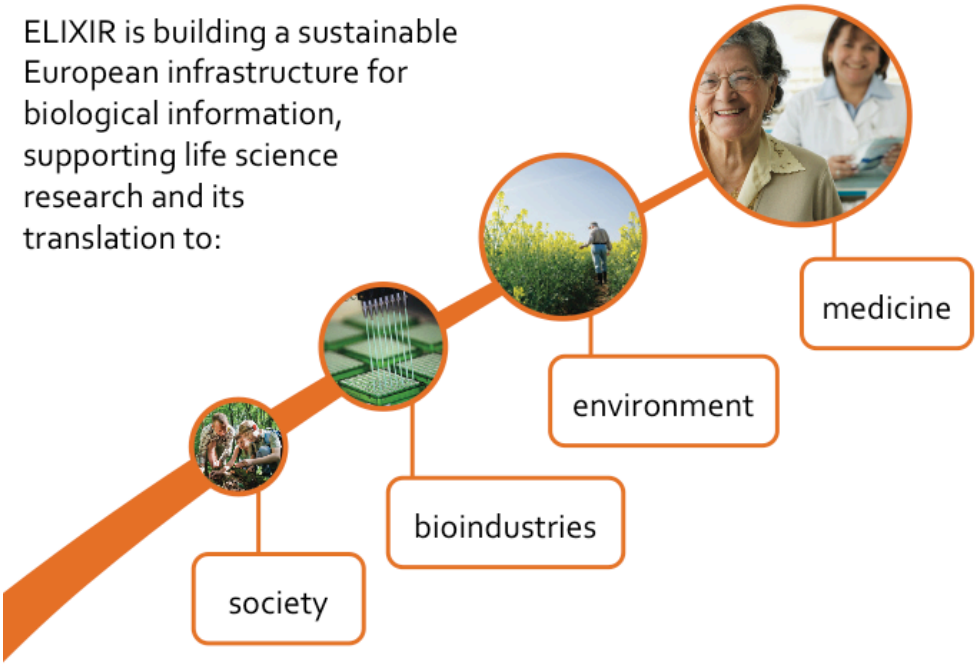
ELIXIR/2013/14
Date: 10 October 2013

ELIXIR 2014-2018 Programme



Purpose of the paper

To present to the ELIXIR Interim Board a draft proposal for the five-year ELIXIR Programme as briefing document and background for the discussion at the November 2013 Interim Board meeting.

This document is also intended as a briefing document for the ELIXIR Heads-of-Nodes meeting on 21-22 October 2013 where the ELIXIR Programme will be further developed.

The ELIXIR Programme will be presented to the full Board for approval at its first Board meeting scheduled for April 2014.

The ELIXIR 2014-2018 Programme is supported by the ELIXIR 2014-2018 Financial Plan (Board paper ELIXIR/2013/22) that outlines the indicative budget and resource requirements for the ELIXIR Hub.

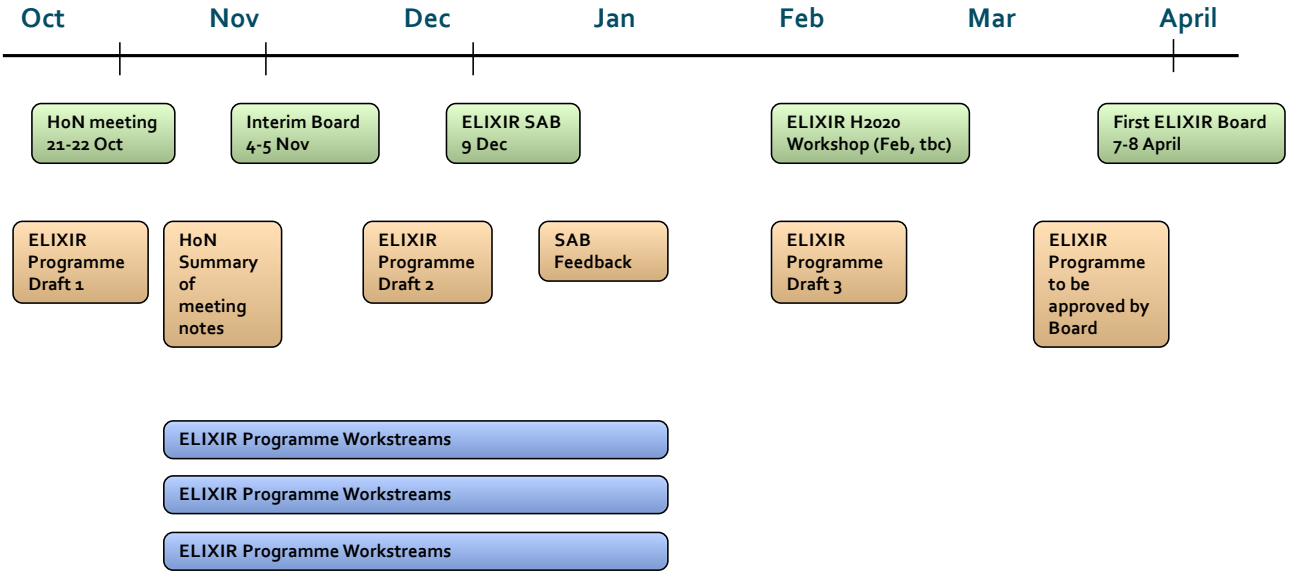
Action required

The draft is presented to the ELIXIR Interim Board for information and discussion.

Table of Contents

1 Developing the ELIXIR Programme

This Board paper marks the start of an extensive process to develop and consult on the ELIXIR Programme up to 2018; this process will include a workshop with the Heads of Nodes committee, a first discussion with the ELIXIR Interim Board as well as a review by the ELIXIR SAB. In addition there will be a set of focussed Work streams for the detailed development of specific areas of the Programme initiated at the HoN meeting. These Work streams will be led by Heads of Nodes and will draw on other expertise as required.



The development of the ELIXIR Programme is also informed by, as well as informs, the development of the ELIXIR Collaboration Agreements for Node services. In addition, as Member States continue to develop and refine national roadmaps, and as new countries join ELIXIR and the ESFRI prioritisation process continues over the coming months, there is likely to be additional input required during the first quarter of 2014. The full ELIXIR Programme will be presented to the ELIXIR Board for approval at its first full Board meeting in April 2014.

1.1 Terms and Abbreviations

Programme of Work	The scientific and technical infrastructure service delivery within ELIXIR, and various other areas of strategic importance, will be organised into a set of broad ELIXIR Programmes of Work. These broad programmes will address a set of specific challenges and user requirements, and thus deliver the value of the infrastructure. Each Programme of Work may contain several Work streams.
Work stream	A time-bound activity to address a specific piece of policy, strategy, or technical aspect. Work stream is led by an appointed individual and includes other members as appropriate.
AEG	High Level Assessment Expert Group established by the European Commission and ESFRI
BMS RI	Biological and Medical Sciences Research Infrastructures on the ESFRI roadmap
ECA	The ELIXIR Consortium Agreement
ESFRI	European Strategy Forum on Research and Innovation
KPI	Key Performance Indicator

2 Introduction

The ELIXIR Programme outlines the key objectives and expected deliverables for ELIXIR during the period 2014-2018. The ELIXIR Programme is accompanied by the ELIXIR 2014-2018 Financial Plan presenting the associated indicative budget for the period. Upon approval of the ELIXIR Programme, detailed objectives and budget for each coming year will be presented annually to the ELIXIR Board for approval at the fall Board meeting. These will build on the overall framework of the ELIXIR Consortium Agreement (ECA), ELIXIR Programme and Financial Plan.

In October 2006 the European Strategy Forum on Research and Innovation (ESFRI), a body set up by 33 countries at the initiative of the European Council, identified 35 pan-European Research Infrastructures that are of key importance for the development of science and innovation in Europe. Among them, ELIXIR is one of very few considered to be of global significance and this Programme will, in addition to the European capability build, outline the internationalization efforts for ELIXIR.

As ELIXIR moves beyond its Preparatory and Interim Phases into permanent existence there is an increased requirement for ELIXIR to coordinate Node actions, ensure effective outreach and engagement with other ESFRI Biological and Medical Sciences Research Infrastructures (BMS RIs) and support technical/scientific interoperability. There is also a key requirement for the Hub to support the development of Node Collaboration Agreements, engage with future Member States and coordinate and drive applications for sustainable ELIXIR funding.

A first draft of the ELIXIR 2014-2018 Financial plan was presented to the ELIXIR Interim Board for discussion at its meeting in April 2013 (Board paper ELIXIR/2013/9) and the Interim Board endorsed the build of the permanent ELIXIR infrastructure to proceed through its construction and operation in three phases: *Co-ordination, Service deployment*, and *Sustained operations*, with a gradual growth of Hub activities and resources matching the on-going construction of ELIXIR Nodes. The first draft of the Financial Plan also laid out a relatively modest resource increase in the first years, with an acceleration taking place post-2015, with the clear rationale of only expanding ELIXIR activities once the Hub / Node network is operational.

The ELIXIR Programme is also in-line with the ESFRI objective that " by 2015, Member States together with the Commission should have completed or launched the construction of 60% of the priority European Research Infrastructures currently identified by the European Strategy Forum for Research Infrastructures (ESFRI)[1]" as well as the recognition by the EC High Level Assessment Expert Group (AEG) that for a Research Infrastructure to be considered as to have reached a mature implementation and operation there are a set of requirements on governance, operational excellence and management that goes beyond the establishment of a legal structure.

This first draft of the ELIXIR Programme builds on the extensive documentation and reports from ELIXIR's Preparatory Phase, culminating in the *ELIXIR Business case*, as well as the discussions with ELIXIR stakeholders throughout the current Interim Phase, including meetings with Node representatives held during visits to Nodes, stakeholder workshops, and interactions with other ESFRI infrastructures. Substantial input to this Programme has also come from the ELIXIR Nodes workshop (March 2013 Noordwijk, Netherlands), the ELIXIR technical coordinator meeting (May 2013, Hinxton, UK), and an analysis of industry needs and expectations delivered through an external agency (ConnectedDiscovery). The ELIXIR Scientific Advisory Board (SAB) reviewed the overall ELIXIR progress as well as the Node applications from ELIXIR Member States and the recommendations from the SAB are an important foundation for this Programme. Furthermore, the ELIXIR Programme and in particular the ELIXIR

[1] COM(2010) 546. SEC(2010) 1161. http://ec.europa.eu/research/innovation-union/pdf/innovation-union-communication_en.pdf

External Funding Strategy also reflects the themes and early intelligence emerging from the Horizon 2020 programme.

The shaping of the ELIXIR Programme will be the focus of the ELIXIR Heads-of-Nodes meeting on 21-22 October (Hinxton, UK) and it is expected that this workshop will substantially reshape the current draft. In particular the workshop will discuss key areas such as ELIXIR Core Services, the ELIXIR kite-mark as well as resource sharing policies.

2.1 The Need for ELIXIR

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

Life-science is moving into an era of intense compute with large, heterogeneous data, often with the extra complication of managing the ethical and personal integrity challenges. Recognizing the challenges in archiving, integrating and analysing the large, often heterogeneous, datasets resulting from modern life-science research, planning for a pan-European research infrastructure for biological data started in 2006 and after a 5-year Preparatory Phase ELIXIR is now being established as an independent organization under the EMBL (European Molecular Biology Laboratory). ELIXIR is constructed as a distributed e-Infrastructure built around existing centres of excellence throughout the European Member States, represented through National Coordination Nodes and the EMBL-EBI. Thus ELIXIR has catalysed the formation of national organisations with explicit support from the research funding agencies / research ministries that often have the explicit task to coordinate national efforts in handling of life-science data.

Data resources now double every 6-8 months with for instance the EBI managing over 30 petabytes (PB) of storage. Within ELIXIR it is however clear that a large fraction of the data will need to be managed in a federated manner, human clinical data can often not leave the network of the home institute without additional ethical review and patient consent. Hence, a bioinformatics infrastructure must manage the current data deluge as well as integrate the data and thereby reduce fragmentation of effort and research. Furthermore, as high-content biology, metabolomics and sequencing based assays rapidly develop we must ascertain the incorporation and support the exploitation of new types of data. Lastly it is essential that we maintain the accessibility of biological data to enhance competitiveness and innovation. ELIXIR have a strong engagement from industry, both as direct partners in ELIXIR Nodes as well as an extensive Industry Work stream during the Preparatory Phase. In addition, ELIXIR recently performed an updated analysis of industry needs (see section 6) that clearly articulated a number of key value drivers for both SMEs and larger companies; some of the common themes identified was 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.

In the face of the data deluge, limited resources, and lack of coordination, the major European data archives will rapidly become unable to meet the full demand and it is therefore essential that ELIXIR, through National Nodes as well as coordinated European investments develop coherent strategies to sustainable scale life-science data management, archiving and services.

2.2 The ELIXIR Strategy

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; high-content biology, and in particular sequence-based biological assays are becoming routine at essentially every major bio-centre in Europe. Thus a European data infrastructure

must be able to cope with the aggregation, annotation and integration of data from hundreds of laboratories as well as scaling the data-services to millions of users worldwide (e.g. the EMBL-EBI services are accessed by over 2M distinct users every year). 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 across Europe.

Thus the ELIXIR infrastructure (Figure 1) is being built as a distributed organisation with National Nodes within each Member States and a co-ordinating hub based at the EMBL-EBI (Hinxton, UK). Distributing resources across Europe allows ELIXIR to scale towards the local data production and usage, leverage the individual strengths of European centres of excellence as well as the flexibility required in investments to meet national priorities and demand. The Nodes also have a key role in the close collaboration with other BMS RI (e.g. biobanks, imaging centres) to service with data infrastructure and ascertain project data management costs (including long term accessibility) are contained and manageable. The ELIXIR Hub will provide coordination and steer to drive European-scale collaboration of key services, technical protocols and training as well as drive joint investments and shared operational costs. For a summary of the roles of the Hub and the Nodes, please refer to Figure 2.

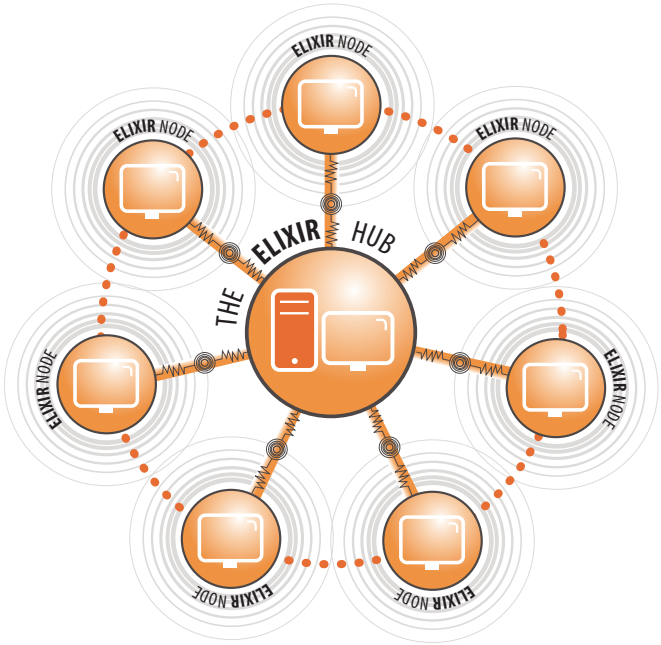


Figure 1. The distributed infrastructure of ELIXIR.

As ELIXIR Nodes have a significant national role, often with significant investments in national research infrastructures for bioinformatics and biological data, as well as responsibility for service delivery at a European and global scale, the distinction of *Node Capabilities* i.e. expertise and competence that can be shared within the ELIXIR framework versus the *Node Capacity* i.e. sustainably funded services offered through *ELIXIR Collaboration Agreements.* Thus, the core strategy of ELIXIR is to build a distributed infrastructure that enables full data integration so that the collective, expanding capacity across the continent is optimally used.

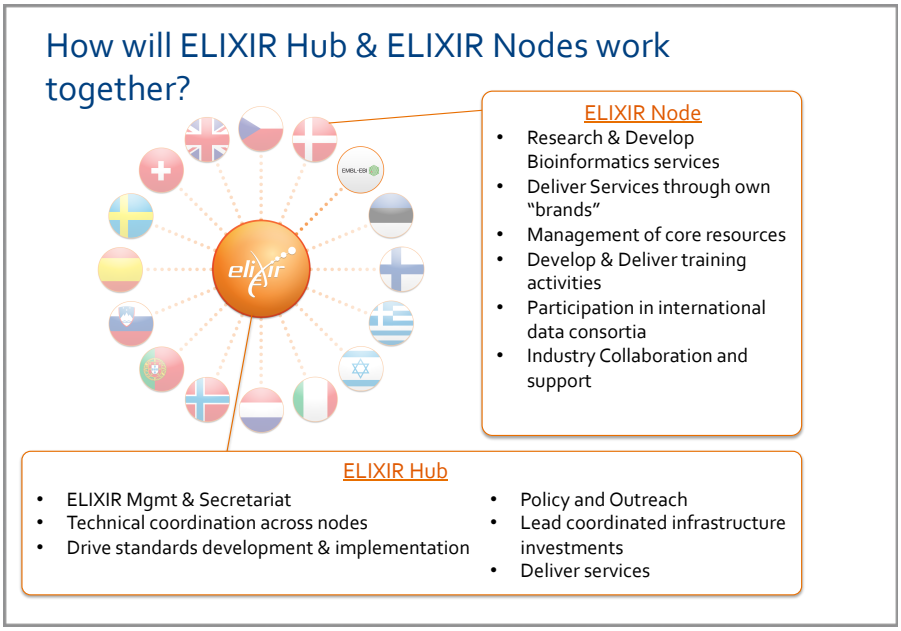


Figure 2. Description of the roles of the Hub and the Nodes.

2.3 ELIXIR 2014-2018 Objectives: Deploy and Operate Sustainable services

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, the negotiation of the ELIXIR Consortium Agreement (ECA) and a build up of initial capacity within the ELIXIR Hub. Following the ratification of the ECA by five countries, which is foreseen for late 2013, ELIXIR will move from Interim to Permanent Phase. This will be met by the implementation of the ELIXIR's Construction and Operation period, which will begin in 2014. In this five-year period, the build of the permanent research infrastructure will proceed through the three phases of *Co-ordination*, *Service Deployment* and *Sustainable Operation (*Figure 3*)*.

During the *Co-ordination* Phase it is envisioned that a number of the Memorandum of Understanding (MoU) signatories will join ELIXIR following the national ratification of the ELIXIR Consortium Agreement and hence there are a number of transitional measures in place to bridge from the Preparatory and Interim Phases to the Permanent Phase. While it is envisioned that additional countries will also be joining ELIXIR after 2014, the process laid down by the ECA will then supersede the transition agreements. Hence the Service Deployment Phase of ELIXIR includes the delivery of a number of key governance Work streams. From 2016 and onwards the ELIXIR infrastructure and its core services will be fully operational and additional services such as those delivered by BioMedBridges will commence and gradually scale.

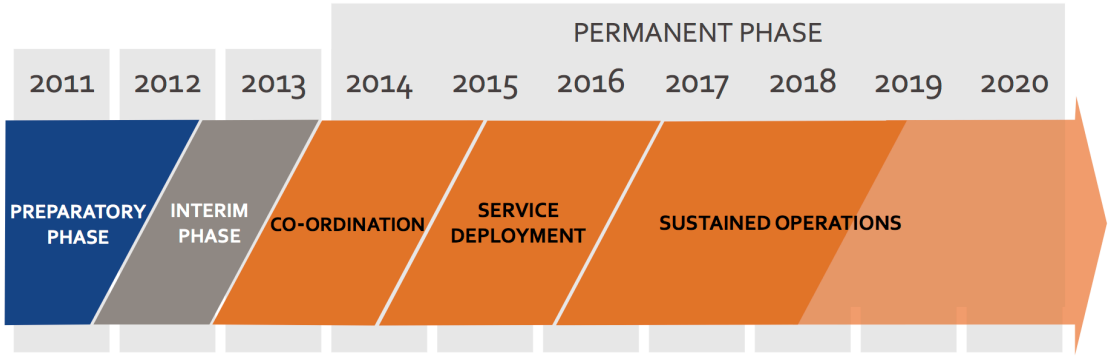


Figure 3. The first five-year Construction and Operation Phase of ELIXIR will proceed through three stages.

The key objectives for each phase are outlined below:

Co-ordination (2014)

- Complete the build of effective secretariat for coordination, outreach and engagement with future Member States.
- Promote ELIXIR in Europe and beyond to support the integration of current MoU signatories as well as application of new Member States and Nodes.
- Interface with other ESFRI Research Infrastructures to ensure effective coordination and alignment resulting in identification of joint services at the European as well as National Node level.
- Drive the establishment of strong links with industry and roll out the ELIXIR Innovation and SME Programme.
- Deliver the agreed ELIXIR technical pilots and support the technical build-up and integration of the Node services through start of a second wave of pilot actions.
- Complete Collaboration Agreements with National Nodes.
- Identify and coordinate the submission of Horizon 2020 infrastructure grants.
- Engage beyond Europe to ensure impact and collaboration at a global scale.

Service Deployment (2015)

- Promote ELIXIR in Europe and beyond to support the integration of current MoU signatories as well as application of new Member States and Nodes.
- As routine delivery of ELIXIR services begin, drive delivery of interoperable data, services and software through execution of ELIXIR interoperability infrastructure including technical support to Node services.
- Establish services and data for ESFRI BMS infrastructure interoperability through technical advice and operational support with data-handling.
- Establish processes and resources at the secretariat to ensure timely identification and coordinated submission of infrastructure supporting grants.

Sustained Operations (2016-2018)

- Consolidation of first and second wave Node applications into a coordinated and interoperable research infrastructure based on open and established standards across a pan-European infrastructure of life-science computing capabilities.
- Begin production phase of the first wave of "new ELIXIR services" on EU and global scale.
- Continued support to BMS Infrastructures in terms of technical advice, operational data provision and standards development.

2.4 Interface with BMS RI and other initiatives

Research infrastructures, by definition, support research projects, and as many scientists and research projects in the life-sciences will draw on services across the BMS RI landscape, the effective collaboration and "*jointness*" of infrastructures is critical to meet user demands. Hence the close collaboration and support to the data management needs with the other ESFRI BMS Research Infrastructures is core to the ELIXIR strategy. ELIXIR is coordinating the BioMedBridges project to build data-bridges between the infrastructures and at a national level many of the Nodes are tightly linked to other infrastructures,

sometimes even through joint consortia or institutes. As the national priorities and participation in different research infrastructures varies, the distributed nature of ELIXIR provides a natural mechanism to leverage national strengths and meet the priorities at the national level. Nevertheless there is a challenge to establish priorities for service delivery across the infrastructures at a European level; ELIXIR will meet this challenge through establishment of *Domain specific services* (section 9.7) where it is expected that a subset of Nodes together with the Hub will coordinate the service offerings. As the BioMedBridges services mature these will be scaled into the ELIXIR portfolio to ascertain long-term sustainability of these critical data-bridges.

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 - http://www.inbiomedvision.eu/), 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 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 will 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.

2.5 ELIXIR Policy on user-charges, open access and intellectual policy

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

Nevertheless there are complicating nuances – when does ELIXIR resources support embargo of data? How does ELIXIR support value creation, e.g. 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 meta-data 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 (see section 9.8.2).

It is important to note that 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 and 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 (see section 9.8.5).

3 ELIXIR Benefits realisation

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

ELIXIR, as the pan-European bioinformatics infrastructure, will deliver 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 biotech enterprises;
- Attract world-leading scientists to Europe (increase "brain gain") and retain key expertise (stem "brain drain");
- Increase the application of new innovations in the biotechnology and pharmaceutical industries, as well as in agriculture and environmental protection;
- Safeguard the investment that nations have already made in funding biological research by ensuring that data are safely kept, and openly accessed by everyone.

Access to biological information is the major driver of biological research. Neglecting to maintain and continuously upgrade the infrastructure that provides this basic service would be detrimental to European biological research and bioscience-related business, impacting directly and negatively on the prosperity of European society.

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 the adoption of best practice techniques and the application of the latest methodologies in use within the discipline. ELIXIR's independent Scientific Advisory Board 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 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[2] set forth by HM Treasury in the UK presents standards of good practice in evaluations of projects, policies, programmes and the delivery of services. These guidelines may be 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

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

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.

Hence a function will be required within 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.

3.1 ELIXIR stakeholders

The ELIXIR stakeholder group is very heterogeneous and reflects the landscape where ELIXIR operates. To effectively communicate and engage with its stakeholders, ELIXIR will develop a Stakeholder Engagement Strategy and define who within the infrastructure is responsible to meeting the needs of each stakeholder. Naturally, for any research infrastructure, the key stakeholder group is the user community of the infrastructure. Due to the importance of this community, users of ELIXIR are covered separately in section3.1.1, and focus here is on other external stakeholders of ELIXIR. These external stakeholders can be broadly categorised into four groups:

- Funding bodies
- Policy makers
- Industry
- Other infrastructures, e-infrastructures, and research initiatives

In Table 1, 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 as the over time. It is important to emphasise that while the role of the Hub is to ensure good communication 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 store 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-thought communication strategy, as there is a risk of not delivering the objectives due to lack of communication between internal stakeholders. The ELIXIR Communication Plan will be developed out of the ELIXIR External Stakeholder Engagement Strategy.

Table 1. A non-exhaustive list of ELIXIR external stakeholders.

Stakeholder	Role of the Hub	Role of the 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 nominates the SAB members.
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.
DG RTD	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 store 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.
Industry, IMI and EFPIA	Engagement with EFPIA will help raising 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. More details in section 6, Table 3.	Direct involvement in R&D collaborations with industry, more details in section 6, Table 3.
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, i.e., 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.	With support from Nodes once ELIXIR's engagement is being defined.

Global Alliance	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)	To 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.
User Community	The users of the research infrastructure are the most important stakeholder group. The user communication plan will be defined separately in ELIXIR User Strategy (see section 3.1.1).	

The European Commission is a stakeholder in ELIXIR in two ways. As a funder of collaborative R&D through FP7, the Innovative Medicines Initiative and, in future, Horizon 2020, large amounts of data will be generated. ELIXIR will be the infrastructure that can store, annotate and add value to, and make available these data to others. Secondly, researchers working on EU-funded grants will also need to access ELIXIR services to be able to carry out their research in the life sciences. Within the European Commission, DC CONNECT is responsible for managing the digital agenda of European Union, including the development of the ICT and e-infrastructure landscape, while DG RTD is responsible for the development and implementation of the European research and innovation policy. Both are key stakeholders.

The Research Data Alliance implements the technology, practice, and connections that make Data Work across barriers. The role, scope and influence of RDA is 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.

Communication with the industry stakeholders will be a key role of both the Hub and the Nodes, while the Hub will coordinate the Industry Advisory Board. 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 (See section 4.4 in ELIXIR Funding Strategy).

For some of the initiatives and stakeholders, the Nodes are expected to lead the ELIXIR communication. For example, the Danish Node will take a lead on interactions with the ETP4HPC community. These roles will be defined in the Stakeholder Engagement Strategy, which will be established by the latter half of

the Service Deployment Phase and fully implemented from the Sustained Operations Phase starting in 2016.

3.1.1 The ELIXIR Users are our key stakeholders

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[3] with 804 respondents representing 318 organisations from 34 different countries. The results from this survey indicate that one of the challenges is the very distinct demand of the two identified, discrete user classes: the "power users" and "non-power users (or end-users)". 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 a biological information environment that acknowledges the diversity of users' interests. This requirement of a multi-disciplinary data environment becomes even more critical for researchers from integrative disciplines, e.g. systems biology, metagenomics, and drug discovery.

The survey also highlighted that to become proficient, ELIXIR would need to overcome several bottlenecks in bioinformatics resources exploitation: lack of resources interoperability, programmatic access, input/output format standardisation and user-friendly web interfaces. Besides these efforts in resources development, there is a need for an optimal community synergy between resources providers/developers and users: involvement of future users during resources development phase, efficient capture of users' feedback information, in addition to development of resources documentation and tutorials. The final conclusion from the survey is the need for parallel developments in bioinformatics training that will be fundamental to maximise the benefits of the research infrastructure.

It is important to realise, however, that this survey had one user group - the private sector users - severely under-represented (89.4% of answering respondents were from the academic/non-profit sector). 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. Hence ELIXIR have commissioned an in-depth report on Industry usage, needs and expectations which now forms the basis of the ELIXIR Industry Strategy (See section 6 for ELIXIR Industry Strategy and Annex 3 for the full report).

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 needs to develop and implement a User Strategy.

3.2 ELIXIR KPIs

The Assessment Expert Group emphasised that an extensive list of parameters alone is not enough to monitor the impact of ELIXIR, but target values of such parameters have to be identified as well. It is critical to define complete and complimentary project-wide KPIs, but also more specific KPIs, for example, to measure the impact and quality of the services each Node will provide. The definition of ELIXIR metrics requires a coordinated approach by the Hub involving the Nodes to enable comparison and to be meaningful across the research infrastructure and also for the national stakeholders of the Nodes.

At the ELIXIR Scientific Workshop in Noordwijk, the ELIXIR metrics were discussed during a breakout session. The report from this session outlines both tangible and intangible metrics to be the most relevant methods for demonstrating the criteria of success of ELIXIR. Such metrics will allow ELIXIR to:

[3] http://www.elixir-europe.org/prep/sites/elixir-europe.org.prep/files/documents/reports/elixir_usersurvey_finalreport.pdf

- Measure ELIXIR overall progress;
- Drive the ELIXIR policy;
- Demonstrate the added value of ELIXIR to funders and other stakeholders.

The workshop report further outlines the types of metrics that should be included:

- Capability building metrics that measure progress against construction plans, e.g.
    - number of ELIXIR Member States and observers;
    - number of Nodes with Collaboration Agreements;
    - number of established / kite-marked services;
    - how much outside funding has been attracted / number of projects coordinated / participated by ELIXIR.
- Numerical KPIs on service deliverables across the RI
    - including training events, usage stats for services, etc.
- Narrative and portfolio metrics that are examples of how ELIXIR
    - contributes to grand challenges, major research programmes and other RI;
    - contributes positively to industry;
    - impact and benefit on setting research policy agenda and data stewardship deposition practises.

Based on the suggestions in the Magenta Book, the narrative and portfolio metrics could be categorised into three main types of evaluation:

- *Process evaluations that assess whether a policy is being implemented as intended and what, in practice, is felt to be working more or less well, and why.*
- *Impact evaluations attempt to provide an objective test of what changes have occurred, and the extent to which these can be attributed to the policy.*
- *Economic evaluations, in simple terms, compare the benefits of the policy with its costs.*

For the ELIXIR KPIs to be appropriate and enable measuring progress against target values, it is important to distinguish between the three different phases in the ELIXIR five-year Programme: Co-ordination, Service Deployment and Sustained Operations. During the Co-ordination Phase, the baseline values of ELIXIR services will be established.

In section 3.3, the proposed ELIXIR scorecard for the Co-ordination Phase is presented, with a suggestion for KPIs. These will further be developed at the HoN Workshop.

The ELIXIR scorecards with meaningful KPIs for Service Deployment and Sustained Operations Phases will be developed later in the project, in Q3-4 of 2014 and in 2016, respectively. It is further suggested that a user survey to be set up for the Service Deployment and Sustained Operations Phases, in order to help judge the size of communities against their needs and ensure representative metrics.

3.3 ELIXIR Scorecard for the Co-ordination Phase (2013 - early 2015)

User community / Customer perspective	**Improved Life Science Research**
	Objectives
	By the Service Deployment Phase starting in 2015, develop a plan to ensure that current and future bioinformatics resources developed in ELIXIR Member States will become more widely used internationally and, through better integration with other ELIXIR services, will ultimately become more powerful resources, than if they were developed and operated in isolation.
	Establish services to expand industry usage as part of the portfolio.
	By the Service Deployment Phase starting in 2015, act as a driver for the technological evolution of existing e-Infrastructure capacities and investments nationally and in Europe.
	KPIs
	Narratives on how ELIXIR contributes to major research programmes and other infrastructures to demonstrate the value of ELIXIR
	Narratives on ELIXIR's positive societal impact, how ELIXIR contributes to the Grand Challenges
	Narratives on how ELIXIR positively impacts industrial research
Financial accountability perspective	**Financial**
	Objectives
	Maintain positive cash balance.
	Establish a Work stream and start to develop the ELIXIR Engagement Strategy for funders to ensure sustainability with Member States and other stakeholders.
	KPIs
	Manage accounts at < 5 % of agreed half-year targets
Internal process perspective	**Capacity building in relation to Service Deployment Phase plans**
	Objectives
	Establish the ELIXIR services in the Hub and in Nodes, including a world-class training infrastructure.
	Establish, evaluate and coordinate cloud computing for ELIXIR
	Ensure that the ELIXIR family grows to include all Member States that signed the MoU and where possible, additional, new Member States.
	Develop and apply for a portfolio of grants from the H2020 program.

	Establish a Work stream to develop ELIXIR User Strategy to expand the user base. Collect baseline values of the user community.
	KPIs / **Baseline values**
	Number of obtained grants
	Success rate of grants (Awards/submissions)
	Number of projects coordinated by ELIXIR
	Number of projects with ELIXIR as a partner
	Number of new ELIXIR Member States
	Number of new observer states
	Number of new Member States, outside of current MoU
	Number of Nodes with signed Collaboration Agreements
	Number of ELIXIR core services
	Number of ELIXIR Kite mark services
	Number of established ELIXIR services
	Number of training events
	Number of participants in training events
	Number of Industry events
	Number of partners in industry
	Number of users
	Geographic distribution of users, number of countries
Innovation, learning and growth perspective	**Enablers of ERA**
	Objectives
	Facilitate communication within ELIXIR to enable staff within Nodes to learn new skills and adopt the state of the art quicker through collaborating closely with other ELIXIR Nodes and the Hub.
	Reduce fragmentation nationally by bringing together several resources and service providers into one dedicated ELIXIR resource, provide a long-term collective framework to ensure better cooperation both within and between countries.
	Ensure interactions with research policy makers to have an impact in the future policies regarding Open Access and the development of data management policies and systems.
	KPIs
	People-centred objectives, Node-to-Node visits and joint workshops
	Defined research exchange policies, Node-to-Node service agreements
	Narrative on the impact and benefit on setting research policy agenda and data stewardship deposition practise

4 ELIXIR Funding Strategy

ELIXIR is built on stable Member State funding, which will provide secure 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, both new and stable.

4.1 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; 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.

4.2 Construction and operation of the Hub

The construction costs of the ELIXIR Hub have been met through a substantial award made by the UK's Large Facilities Capital Fund (LFCF), which has been used to construct the building housing the ELIXIR Hub and off-site storage at the London Data Centre.

The on-going operating costs of the ELIXIR Hub are met by the participating Member States based on their relative Net National Income (NNI). In the Interim Phase the operating budget of the ELIXIR Hub is approved on an annual basis and within the framework of a three-year plan approved by the Interim Board (2012-2014). Once the ECA comes into effect, Member States will be asked to approve a five-year Financial Plan for the Hub.

4.3 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 operations of ELIXIR and cover activities including such as the integration of services, building the shared e-infrastructure, enhancing user access, training, collaborations with industry and building bridges between the other ESFRI Biological and Medical Sciences Research Infrastructures.

4.4 Opportunities within Horizon 2020 and IMI

Whilst the full details of Horizon 2020 are not yet known, it is clear that the annual Calls will present ELIXIR with significant opportunities to leverage EU funding. Over the course of the seven-year programme, topics of relevance to ELIXIR are likely to appear in the following Horizon 2020 Work Programmes:

- Research Infrastructures;
- Health and Demographic Change and Wellbeing;
- Food security, sustainable agriculture, marine and maritime research, and the bio-economy; and
- ICT.

Similarly, the second IMI will also present topics that ELIXIR will wish to apply to. Data management is a transversal theme that cuts across the EFPIA Strategic Research Agenda, presenting ELIXIR partners with scope to engage in many fields. In addition to the financial aspect, participation in IMI grants would also allow ELIXIR members to build close link with key industry players.

Over the course of the ELIXIR Programme, the ELIXIR Hub will explore various other options for suitable funding ranging from schemes such as COST, new public private partnerships like BRIDGE2020 for bio-based industries, and supporting Nodes through EU Structural Funds to ERA Chair calls and twinning schemes. The ELIXIR Hub will seek to identify suitable topics and regularly inform Heads of Nodes of these opportunities.

4.5 ELIXIR's External Funding Strategy

Given the potential scope there is 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 2) are required to support the whole project application life cycle, which includes: helping to influence and shape the funding landscape to suit ELIXIR; identifying forthcoming and current opportunities; assisting with the coordination of joint ELIXIR applications; and management of selected grants (Figure 4). This is further strengthened by the AEG recommendation to perform a thorough analysis of the funding strategy under external leadership, this will be initiated as part of the ELIXIR Action Plan (Annex 1, see also section 9.8.4).

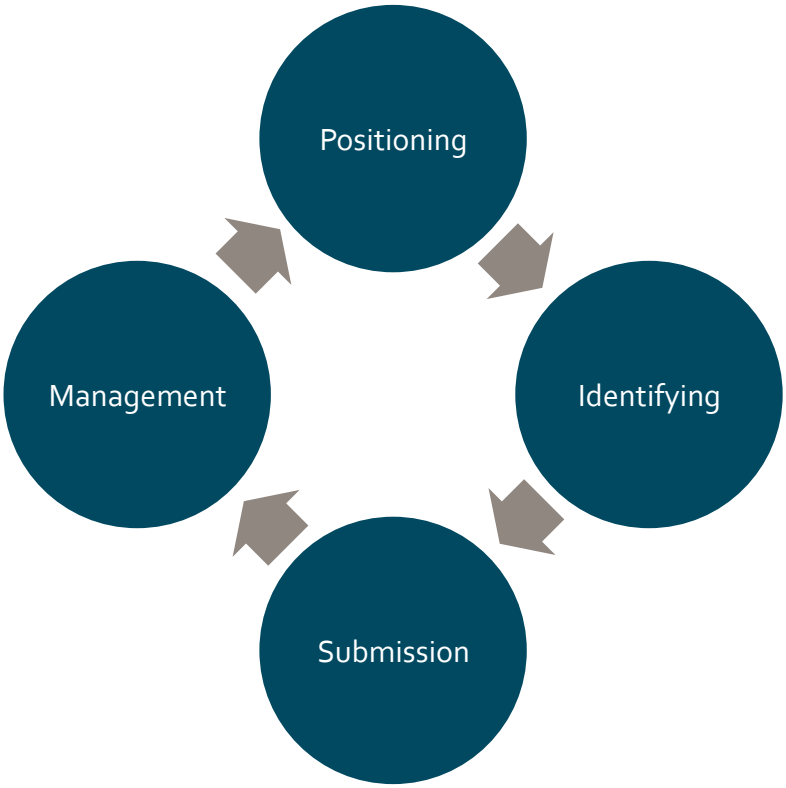


Figure 4. ELIXIR External Funding Strategy cycle

Table 2. The table sets out the various roles and responsibilities of the Hub and Nodes in implementing the ELIXIR Funding Strategy.

Funding Strategy	Role of ELIXIR Hub	Role of ELIXIR Nodes
Positioning	• Engaging key stakeholders in Brussels – see Stakeholder engagement in Table 1 • Drafting of ELIXIR position papers and consultation responses	• Supporting role
Identifying	• 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 • Coordinating some ELIXIR applications • Organising logistics around planning meetings • Provision of certain general text 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

4.6 Principles to best ensure success

- It is clear that there will not be one single grant for ELIXIR from Horizon 2020; rather, ELIXIR will have to build up a portfolio of grants from Horizon 2020 and other sources.
- H2020 evaluation will be excellence-based, with peer reviewers looking closely at the appropriateness of each partner for each activity. ELIXIR needs the best partners from within the ELIXIR network for each grant. This position was endorsed by ELIXIR's SAB. Not all ELIXIR partners will take part in all grants.
- Only ELIXIR Member States can apply for funding using the name ELIXIR with the exception that during the transition phase 2014-15 MoU countries in process of ratifying the ECA and that have paid membership contribution will also qualify (see Board paper ELIXIR-2013-5).
- ELIXIR Director and the Heads of Nodes should agree on what applications are submitted under the name of ELIXIR, avoiding competing applications being submitted.
- Focus should be on topics that can fund bioinformatics services provision with relevance to data, tools, compute, training, standards or industry. Grants that are focused more on pure research are out of scope for ELIXIR, unless they are research grants that have large data requirements and ELIXIR can add value. As a research infrastructure, ELIXIR's natural focus should be on research infrastructure grants – grants that build and provide services and should try and steer clear of pure research projects – where there is discussion on what is in scope, ELIXIR should consider whether there is a strong data component necessary and irreconcilable.
- Applications to funding will be agenda items at future Heads of Nodes meetings and ad hoc meetings of Heads of Nodes to discuss/plan/submit applications will be planned.

- Meetings in Brussels will be used for coordination and development of grant applications, to minimize travel time for all and venue costs.
- The ELIXIR Hub will support, where possible, Nodes applying for national funding, through for example writing letters of support and material on ELIXIR.

5 ELIXIR Governance and Organisation

The ELIXIR infrastructure is being built as a distributed organisation (Figure 1) with National Nodes within each Member State and a co-ordinating Hub based at the EMBL-EBI (Hinxton, UK). Distributing resources across Europe allows ELIXIR to scale towards the local data production and usage, leverage the individual strengths of European centres of excellence as well as the flexibility required in investments to meet national priorities and demand. The distributed character of ELIXIR poses specific challenges for the organisation and governance of the project, which were also highlighted by the Assessment Expert Group. They state a distributed infrastructure, such as ELIXIR, should pursue *adequate mechanisms and procedures for managing the relationships between the centralised governing bodies and the national authorities*. *Service level agreements should be negotiated to establish the process flow.* The Group further emphasises the important role of the Hub to *ensure standardisation and to avoid overlap across the infrastructure*.

The Assessment Expert Group report demonstrates with further guidelines how an appropriate governance and legal structure is important for the research infrastructure *to organise itself as a project capable of attracting resources*. For example, it is necessary to ensure that *adequate decision-making bodies with an appropriate degree of authority and autonomy are appointed with sufficient power to take decisions.* The Assessment Expert Group also emphasises the importance of transparent procedures when nominating and electing all Boards and highly recommends establishing an external Ethical Board to ensure that social responsibility issues and reputational risks are managed and mitigated. Finally, the Group reminds that the research infrastructures need to coordinate with other world-wide initiatives in the same research area.

5.1 Key bodies

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

The ELIXIR Board will oversee ELIXIR's activities. The ELIXIR Member States and EMBL will appoint the ELIXIR Board, which in turn will appoint the ELIXIR Director (Figure 5 displays a graphical representation of the ELIXIR's Governance Model).

In addition to the Board, there are external and internal governance bodies. The ELIXIR Scientific Advisory Board and Industry Board, appointed by the ELIXIR Board will advise the ELIXIR Board and Executive Management in scientific matters and relations to industry, respectively.

The internal governance bodies include the Heads of Nodes Committee, which gives advice to the ELIXIR Board and Director in relation to ELIXIR activities, and the Technical Coordinator Group, which advises the Heads of Nodes Committee. The Node Steering Groups monitor the tasks defined in the Collaboration Agreements between each Node and the Hub. These committees will be established following the entering into force of Collaboration Agreements between each Node and the Hub.

In addition to these committees, the ELIXIR Board may form additional advisory bodies and working groups where necessary, e.g. the Collaboration Agreement Working Group or Node Selection Working Group.

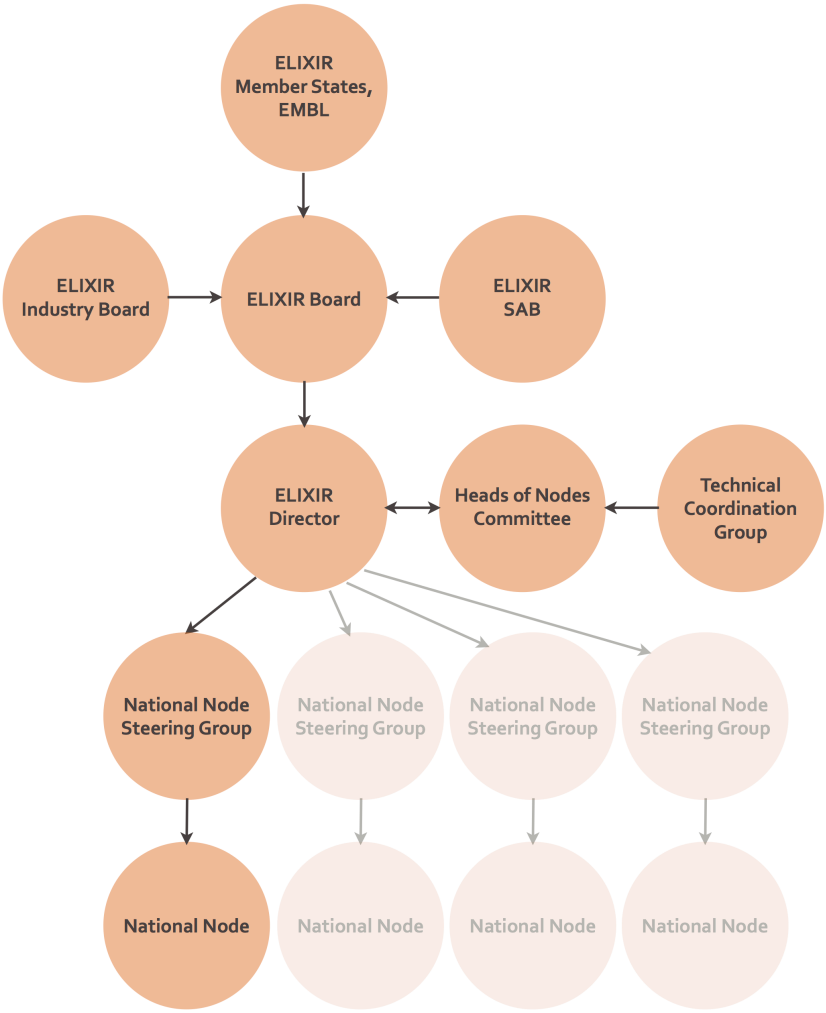


Figure 5. ELIXIR Governance Model

5.1.1 The ELIXIR Board

- As the most senior decision-making body of ELIXIR, the ELIXIR Board will provide high-level oversight and approve the ELIXIR strategy and budget (including managing additional bilateral agreements). It will oversee management of risks and liabilities of ELIXIR, and resolve disputes and disagreements. Each country will have one vote and members will have due authority to deliberate, negotiate and decide issues relating to budget and to ELIXIR's strategic Programme, as well as establishing and monitoring rules and procedures, including those for the selection of ELIXIR Nodes. They will be expected to report back to their own Member States, and draw on advice from the SAB.
- The ELIXIR Board is composed of the scientific and administrative representatives from each ELIXIR Member State. The ELIXIR Board elects a Chair and two vice-chairs. The ELIXIR Board may also decide to admit observers, from prospective Member States and from charities and industry.
- The Board will meet in person at least one per year, but preferably biannually in the spring and during the fall. One of these meeting will take place in the Hub, while the second will be hosted in the Nodes. The Board will additionally have teleconferences when required.

5.1.2 The ELIXIR SAB

- The SAB will provide 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.).
- Membership of the ELIXIR SAB will be made up of distinguished international experts including academics not associated with ELIXIR Nodes, representatives of other internationally renowned organisations operating outside of Europe, and representatives of researchers in the commercial sector. This group will be appointed by the ELIXIR Board, following the proposal of candidates by the ELIXIR Director.
- The SAB will elect a Chair and one Vice Chair and will meet twice a year. The SAB reports directly to the ELIXIR Board and ELIXIR Director, and will act independently of EMBL.

5.1.3 ELIXIR Ethical Board

- The Assessment Expert Group recommends ELIXIR to establish an external Ethical Board.

5.1.4 ELIXIR Industry Board

- A key finding from the industry user and stakeholder analysis (section 6) was the recommendation to establish a high-level Industry Board to advise ELIXIR.

5.1.5 Heads of Nodes Committee

- The Heads of Nodes Committee is the key body to develop and agree the ELIXIR Scientific and Technical strategy. In addition to the formal requirement on consultation for the Programme set forth in the ELIXIR Consortium Agreement it is expected that this group 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 the ELIXIR Funding Strategy.
- The Committee is composed of the Heads of each ELIXIR Node and a representative of EMBL-EBI. The ELIXIR Director leads the Heads of Nodes committee, and the committee will nominate a vice chair.
- The Committee will meet biannually such that one meeting will take place in the Hub and the second meeting in the Nodes. In additional the Committee will hold teleconferences as required.

5.1.6 Technical Coordinator Group

- The role of the ELIXIR Technical Coordinator Group is to share information and discuss the technical and scientific aspects of the project. In addition, the ELIXIR Board may task the Technical Coordinator Group to address and report back on specific technical issues.
- The group is composed of technical experts representing each ELIXIR Node, appointed by the Head of the Node. 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.
- The Technical Coordinator Group will meet biannually at technical workshops that are organised in appropriate locations considering the workshop topics. In addition, there will be teleconferences when needed and the Group will also communicate via purposefully created mailing lists.

5.1.7 Node Steering Group

- Based on the current draft of Collaboration Agreements upon the formation of a Collaboration Agreement between an ELIXIR Node and the Hub, a Node Steering Group is formed. The Group is responsible to ensure that a regular exchange of information and joint project coordination is established. The Group monitors the tasks, responsibilities and delivery of services of the ELIXIR Node, and the services of the ELIXIR Hub that the Node is relying on in accordance with the terms of the Collaboration Agreement. Most importantly, the Steering Group will monitor the KPIs established for the Node and the Hub as defined in the Collaboration Agreement and is responsible to act and mitigate if the services fail to meet their targets.
- The Node Steering Group comprises of the ELIXIR Director, the Head of the Node and other individuals, e.g. representatives of the distributed national network or experts on relevant issues, appointed by the Head of the Node.
- The Steering Group will meet at least once a year and will hold teleconferences when needed.

5.2 The ELIXIR Hub

The ELIXIR Hub will host the ELIXIR Executive Management and Secretariat, who will report to the ELIXIR Board for delivery of ELIXIR. The Hub ensures effective coordination of ELIXIR activities and maintains effective interactions with the other ESFRI BMS infrastructures at the European level, enabling ELIXIR members to work with these infrastructures and helping to coordinate the e-infrastructure for data that they will require. The role and requirements of the ELIXIR Hub are further detailed in section 7.

5.3 The Collaboration Agreements

The Collaboration Agreements form a very central part of the ELIXIR Governance structure and the importance of these service levels agreements was also highlighted in the Assessment Expert Group report. In 2013, a Working Group to draft the first Collaboration Agreement template was established and this draft template is expected to be available by the first HoN meeting. In 2014, a major priority is to finalise the templates and the accompanying Annex of Service Delivery Plan, and for this purpose, a number of Work streams will be established. In addition, KPIs to monitor the progress of each of the Service Delivery Plans will be defined. The ELIXIR Hub will continue to support the process by providing legal expertise.

5.4 Key actions and deliverables for the Programme

The ELIXIR Action Plan, presented as Annex 1, displays ELIXIR's responses to the Assessment Expert Group recommendations. In addition, the Action Plan outlines the steps that need to be taken to ensure that ELIXIR can achieve the objectives of the three Phases: Co-ordination, Service Deployment and Sustained Operations. It is important to note that the Action Plan is not a static document but will be periodically reviewed and updated.

The ELIXIR 5-year Programme comprises several distinct Programmes of Work and different Work streams within these Programmes. Definition and redefinition of strategies, including Funding and Policy Engagement Strategy and the coordination of the Programmes and Work streams will be an important role of the ELIXIR Hub.

The Assessment Expert Group criticised most of the infrastructures for the lack of a risk register and mitigation plans. The first draft of the ELIXIR risk register that takes into account the Expert Group recommendations is presented in Annex 2. Going forward a Work stream will be established to prepare a project-wide risk register for ELIXIR. The Hub will ensure that the register is periodically reviewed and risks mitigated.

6 ELIXIR Industry Strategy

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 as 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. Stimulating innovation and supporting industry is therefore a key objective for ELIXIR.

A stable, robust and reliable bioinformatics infrastructure - fit for industry's need - will also ensure that ELIXIR Member States and funding bodies derive maximum value from their investments in ELIXIR. ELIXIR's Industry Strategy should complement and enhance existing interactions between industry and ELIXIR Nodes, which in many cases are already fruitful and well established.

In order to develop an effective, holistic Industry Strategy, ELIXIR commissioned ConnectedDiscovery to carry out an extensive scoping exercise, involving meetings and conference calls with over 40 industry users. The in-depth analysis of industry needs, expectations and key findings delivered by ConnectedDiscovery is attached as Annex 3 and will form the basis of the ELIXIR Industry Strategy. The main recommendations, suggestions for implementation and guidance on the role of the ELIXIR Hub and Nodes in this process are described below.

6.1 Key findings and recommendations

- Current understanding of scope and activities of ELIXIR amongst industry is currently limited and confused; key workshops/events to present the role of ELIXIR services to industry are required.
- Industry would value the development of the following distinct activity lines: the formation of an ELIXIR-wide network of news and events relevant to industry; development of standard legal agreements for service development and collaboration building; a focussed SME outreach programme; the development of SME Embassy Cloud sandbox environments; the formation of high-level special interest groups from industry users, suppliers and publishers; support in understanding the infrastructure landscape; avenues to collaborate with partners on EU-level funding programmes such as IMI and Horizon 2020.
- Industry urges ELIXIR to develop and implement kite-marks to brand services and would value clarification both on ELIXIR's position with respect to data operability and ELIXIR's vision as regards other relevant global initiatives.

6.2 Implementing the ELIXIR Industry Strategy

Whilst ELIXIR will look to secure funding through Horizon 2020 where it is appropriate to contribute towards the costs of the Industry Strategy, it is clear that the core costs of this must be met by the operating budget of the ELIXIR Hub. Dedicated posts within the ELIXIR Hub secretariat will be required to support various aspects of the implementation of the ELIXIR Industry Strategy, in particular the Innovation and SME programme, which will require logistical coordination (Table 3). The ELIXIR Hub budget could be used to cover the cost of a certain number of these events hosted by ELIXIR Nodes.

Additionally, ELIXIR Nodes themselves will also be involved in implementing the strategy, not least through organizing and hosting ELIXIR Innovation and SME Programme events and through their direct collaboration with industry as well as the local bioregion organisations.

6.3 Innovation and SME Programme

One of the clear messages from ConnectedDiscovery's analysis was that ELIXIR can play a major role in supporting SMEs throughout Europe, particularly through 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.

- Attendance of 50-70 SMEs at each event, ensuring ELIXIR Node builds and maintains lasting links with local SME users
- Resourced through ELIXIR Hub budget, with some contribution from ELIXIR Nodes
- Event programmes to include training to SMEs or information on relevant resources and services across the ELIXIR network
- Agenda developed by ELIXIR Nodes and Hub and tailored for local interests, i.e., pharma, biotech or food-agriculture.

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, once the Programme is operational it could be scaled to take place more regularly.

Table 3. The roles of the ELIXIR Hub and Nodes in implementing the ELIXIR Industry Strategy

ELIXIR Industry Strategy	Role of ELIXIR Hub	Role of ELIXIR Nodes
Industry dissemination network	• Overall coordination and outward communication of activities	• Local interactions with industry and communication to Hub on new services and news
Innovation and SME events	• Overall coordination of programme • Providing ELIXIR Hub financing to cover costs	• Hosting SME events locally
Development of Embassy Cloud for SME model	• Overall co-ordination of programme	• Definition of scope and concept • Direct involvement in activities; local interactions with companies
High-level interest groups	• Coordination and organisation of meetings • Secretariat support	• Hosting of and participation in meetings
Global landscape mapping	• Coordination of activity	• Input into activity
Engagement with IMI and EFPIA	• Overall coordination of engagement with EFPIA and IMI • Finding, communicating and supporting opportunities for Nodes in collaborations	• Direct involvement in R&D collaborations with industry

6.4 Overlap with other areas of ELIXIR Programme

In addition to a dedicated set of activities within the Industry Strategy, engagement with industry and SMEs will also form part of the **ELIXIR External Funding Strategy**, with the objective of developing strategic collaborations via EU-funded grants. Initiatives such as IMI, applied topics within Horizon 2020

and emerging initiatives such as BRIDGE2020 will all present ELIXIR with opportunities. Industry is therefore also a key stakeholder in **ELIXIR's External Stakeholder Engagement Strategy**. To this end, engagement with European industry associations and bodies such as European Technology Platforms will also help ELIXIR's positioning as regards industry and SMEs.

The development of an ELIXIR kite-mark is an integral part of the **ELIXIR Services and Themes Programme**. Direct input from industry will be sought when ELIXIR partners undertake the development of the ELIXIR Kite-Mark concept. Likewise, with respect to the recommendations on the need for clarity as regards ELIXIR's position on data integration and HPC, partners will seek to address this through the development of the **ELIXIR Services and Themes Programme**.

7 The ELIXIR Hub

The ELIXIR Hub hosts the ELIXIR Executive Management and Secretariat and contains the external relations, communications and grant-application functions necessary for a large, distributed research infrastructure and the effective interfacing and coordination with other biological, health and e-Science infrastructures. While the ELIXIR Nodes will deliver the core service delivery and interoperability, the ELIXIR Hub will handle the organizational, technical and infrastructure 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.

Following the ratification of the ELIXIR Consortium Agreement by five Member States and EMBL the ELIXIR infrastructure will be legally incorporated as an *EMBL Special Project* as defined in the ELIXIR Consortium Agreement; this construct has significant advantages in that ELIXIR will benefit from EMBL's legal persona (including VAT exemption and light procurement regulations), staff-rules and financial regulations, as well as the operational benefits and associated lower costs of using an established administrative infrastructure for the ELIXIR Hub. Nevertheless ELIXIR and the ELIXIR Hub will have operational independence from EMBL, through governance by the ELIXIR Board and a separate funding mechanism. Thus: ELIXIR must at all stages be independently funded, maintain a positive cash-balance and EMBL will charge ELIXIR for the usage of EMBL services, personnel and overheads. The rapid construction of ELIXIR will benefit from the opportunity to draw on services from the EMBL administrative and technical teams and hence the following sections will describe the functional requirements and necessary roles without prejudging the location of these resources.

As ELIXIR is established there is an intense set of activities to build the ELIXIR Executive Management and Secretariat as well as develop and support the Collaboration Agreements with the ELIXIR Nodes. A key role for the ELIXIR Hub will be to identify and develop grant applications together with the Nodes for the Horizon 2020 research programme and manage the effective collaboration and coordination with other ESFRI Research Infrastructures to secure the necessary funding. During the Co-ordination Phase of building the ELIXIR infrastructure the focus is to construct an effective secretariat; as the ELIXIR Nodes are established and the infrastructure moves into permanent operation there will be a gradual scale-up of technical operations. Technical and scientific services from the Hub can be delivered directly by ELIXIR Hub staff, through commissioning services from ELIXIR Nodes, or through external procurement; again the following sections make no assumptions on the delivery mechanism but rather focus on the user requirements and service needs.

The emerging ELIXIR infrastructure will need to initiate a set of working groups to develop plans for effective coordination and development of EU wide services in key areas such as annotation of new genomes for a wide range of organisms, storage and computing access for the emerging, very large biological datasets as well mechanisms for secure access to human data. High quality data-curation is a long tradition and key feature of European data resources and ELIXIR will provide for a framework where the research communities can effectively collaborate and deliver well annotated datasets in areas of national expertise such as plant species or other economically important organisms. Although these activities form a key part of ELIXIR Node services, there will be a need for technical support from the Hub for standardisation and interoperability.

Technical and scientific training is a key component and critical success factor for ELIXIR but as several of the Node applications propose significant training components this plan only contains a small training resource required for support and technical management of ELIXIR training by the ELIXIR Hub.

7.1 Key functions at the Hub

The ELIXIR Executive Management, assisted by the Secretariat Staff, 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 the support to ELIXIRs standing committees, maintaining accounts, auditing and submission of the annual budgets and financial reports. The ELIXIR Hub coordinates outreach and communication and maintains an effective interaction with key science policy functions at the European level as well as actively monitoring collaboration and funding opportunities to ensure that ELIXIR services are relevant, effective and sustainably supported. The ELIXIR Hub will help to ensure the provision of training both by providing support for and coordination of suitable trainers in the Nodes, and by supplying an infrastructure for the management of training materials, so that they can be used and re-used by all of ELIXIR's trainers. Thus, within the ELIXIR Secretariat the operations can be group in three major units further outlined below: *ELIXIR Strategy, Policy and External relations, ELIXIR Programme Management and Operations,* and *ELIXIR Technical Coordination and Delivery (*Figure 6*).* As outlined above, the *EMBL Special Project* status enables the ELIXIR Hub to source financial, human resource, administrative and purchasing services from EMBL, hence there are no staff provisions for these functions within the Hub but they are budgeted for in the accompanying Financial Plan.

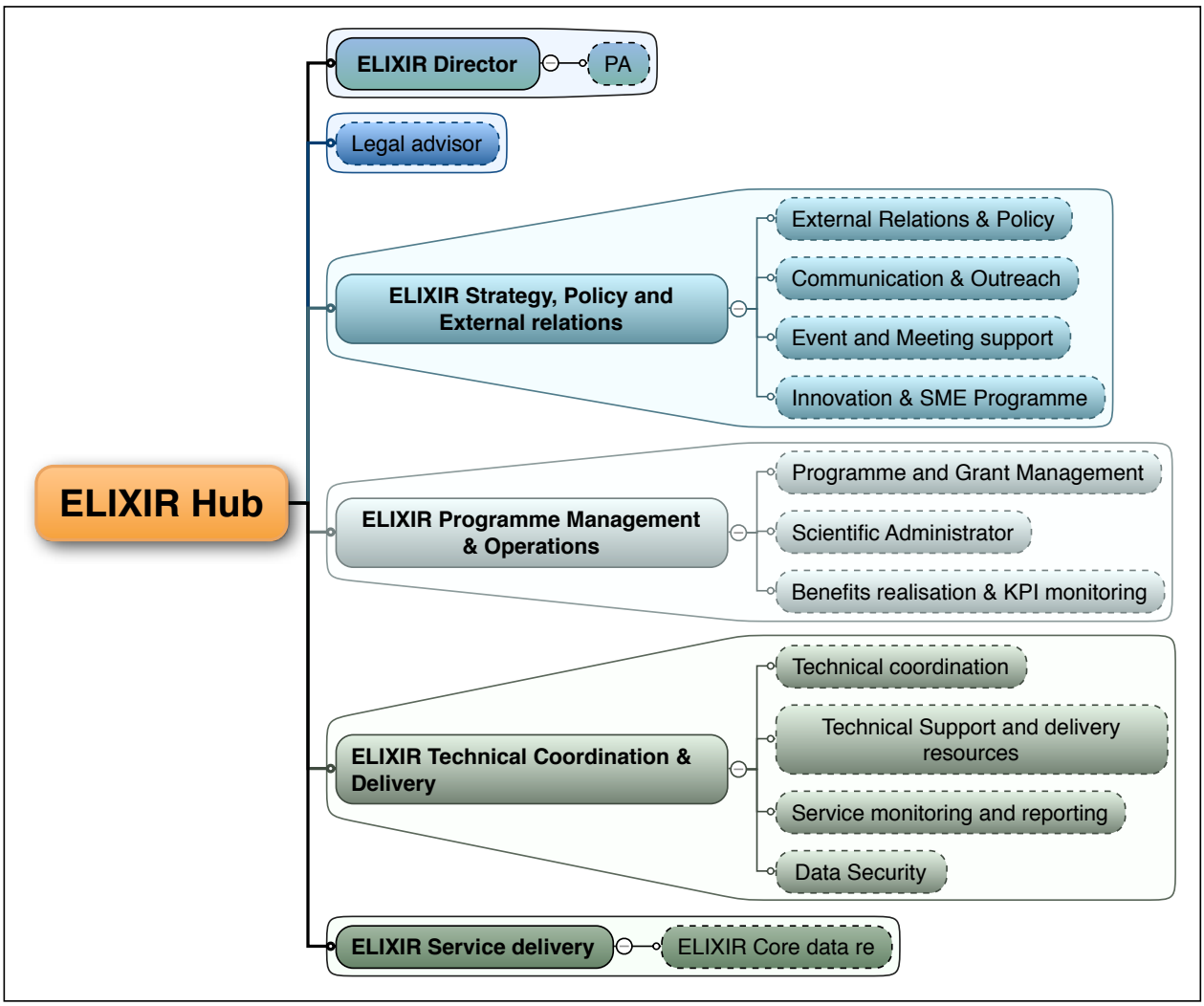


Figure 6: Functional Units of the ELIXIR Hub.

7.1.1 ELIXIR Hub resource and staffing requirements

To match the outlined activities, the operating budget of the ELIXIR Hub would need to grow from 1.4M€ in 2013 up to 5.6M€ in 2018, with most of that growth in the Sustained Operations Phase beyond 2016. The Financial Plan also includes capital requirements to support acquisition of additional compute and storage resources in 2015 and 2018. Key resource requirements and expectations on the ELIXIR Hub in the three Phases are:

Co-ordination

- Following recruitment of ELIXIR Director, build an effective secretariat for outreach, engagement with future Member States and coordinate applications for H2020 infrastructure grants.
- Promote ELIXIR in Europe and beyond to support the application and integration of new Member States and Nodes.
- Interface with other ESFRI Research Infrastructures to ensure effective coordination and alignment.
- Drive the establishment of strong links with industry.
- Deliver the agreed ELIXIR technical pilots and support the technical build-up of the Node services through accelerated execution of first and second wave pilot actions.

Service Deployment

- As routine delivery of ELIXIR services begin, drive delivery of interoperable data, services and software through execution of ELIXIR interoperability infrastructure including technical support to Node services.
- Establish services and data for ESFRI BMS infrastructure interoperability through technical advice and operational support with data handling.

Sustained Operations

- Consolidation of first and second wave Node applications into a coordinated and interoperable research infrastructure based on open and established standards across a EU-network of life-science computing capabilities
- Begin production phase of the first wave of "new ELIXIR services" on EU and global scale.
- Continued support to BMS infrastructures in terms of technical advice, operational data provision and standards development

7.1.2 ELIXIR Directorate and Legal counsel (2.5 FTE)

In addition to the administrative support for the ELIXIR Director, ELIXIR has a critical need for a legal advisor to support the development of Collaboration Agreements, Memorandum of Understanding and other bilateral agreements. Legal support is provisioned for by the EMBL International legal specialist as well as independent advice from international law firms.

7.1.3 ELIXIR Strategy, Policy and External relations (4 FTE)

The core role of the *ELIXIR Strategy, Policy and External relations* function is to ensure that ELIXIR is to monitor, and as appropriately engage and influence, developments in the European and Global science policy landscape. This function is also responsible for the effective interactions and communication with the Member States and engagement and on Boarding of new Member States.

The Outreach and Communication activities from the ELIXIR Hub also reside in this function, including ensuring web and social media presence, development of press releases, communication materials and annual reports. Coordination and support of ELIXIR presence at events and scientific meetings, organisation of ELIXIR workshops is also resourced from this function as well as the responsibility for effective interactions with industry including the Innovation and SME programme. In today's multi-media society ELIXIR must develop a multi-media strategy, especially in designing and maintaining an informative web site. Therefore a web designer and software engineering resources are needed to handle

the technical details of the web site and its development and maintenance, together with a communications officer to monitor and develop the information content.

A key role for the ELIXIR Hub is to develop and maintain a strategy to ensure sustainable funding of service development and delivery. The Strategy, Policy and External relations function will maintain an overview of the funding landscape and as appropriate engage with stakeholders to communicate ELIXIR needs, requirements and future opportunities, including collaborations with other research and e-Infrastructures.

7.1.4 ELIXIR Programme Management and Operations (3 FTE)

The *ELIXIR Programme Management and Operations* function provides the on-going coordination and project management support required to coordinate service delivery across the distributed infrastructure. This function is also responsible for the scientific administrative support to the ELIXIR governance bodies.

ELIXIR Programme management will ensure the coordination of a complex portfolio of interdependent *Programmes of Work*, monitoring of KPI and documenting benefits realisation and maintain an updated risk register with associated mitigation plans. This function will also play a key role in coordinating the development of ELIXIR grant applications and ensure the timely action on emerging opportunities.

7.1.5 ELIXIR Technical Coordination and Delivery (2 FTE)

The *ELIXIR Technical Coordination and Delivery* function carries the core responsibility, in close collaboration with ELIXIR Node technical staff, to manage and develop the portfolio of services and ensuring that ELIXIR offers globally competitive and interoperable data and tool resources. The *ELIXIR Technical Coordination and Delivery* function will also maintain the portfolio of ELIXIR Pilot actions and ensure that opportunities are brought forward for potential implementation.

This function will also support the development of the technical and scientific details of ELIXIR Collaboration Agreements as well as ensure that quality standards and life-cycle management plans for services are agreed, implemented and monitored. An important part of this role is to lead the development and implementation of the ELIXIR technical and interoperability strategy. Critically, the *Technical Coordination and Delivery* function will handle the technical and infrastructural interactions with the Nodes and the other ESFRI BMS infrastructure projects, ensuring that ELIXIR becomes a coordinated infrastructure that can ensure seamless access to biomedical data resources and the tools required to make use of them from throughout Europe. This will require coordination of standards and tools with the National Nodes and other infrastructures to improve the integration and access to data. The Technical Coordination and Delivery function of the secretariat will also ensure that ELIXIR has a coherent and effectively implemented data security policy.

Additional software engineers will also be needed to develop and operate mechanisms to monitor resource connectivity, reliability, usage etc. over the whole infrastructure. These functions will not only be at the heart of the operation but will also be used to quantify the success of ELIXIR. Monitoring will cover aspects like web hits, unique users, jobs run, data downloaded etc. for all the different parts of ELIXIR. For a large distributed infrastructure, such monitoring is not trivial (5 staff will be required by 2018), but must be addressed as part of the infrastructure's need to measure both its functionality and its progress towards the goals set by the ELIXIR Board.

7.1.6 ELIXIR Core Service delivery (growing to the equivalent of 19 FTE in 2018)

Effective delivery of the core life-science data services needed by the life-sciences community in Europe is the central mission of ELIXIR and this section sets out the resource requirements at the Hub to ensure sustainable service delivery as well as development of resources to meet the challenges of the current life-science data surge. It is envisioned that ELIXIR Hub service delivery will be provisioned through a

combination of *Commissioned Services*, core staff at the ELIXIR Hub as well as competitive procurement from third parties.

While it is expected that the details of *ELIXIR Core Services* and associated resource requirements will be refined during the development of the *ELIXIR Programmes of Work* (as outlined in section 1,"Developing the ELIXIR Programme") the following indicative areas and resource requirements have been identified during the Preparatory Phase and presented to the steering group and Interim Board.

ELIXIR Core data resource delivery

The most significant responsibility for ELIXIR is the delivery of the ELIXIR core data resources to European life scientists with very high levels of availability and robustness. Such data delivery requires expert systems staff, but also software engineers to run these external services and monitor them on a daily basis. Initially most of the data to be deployed at the London Data Centre corresponds to resources developed at EBI in collaboration with European partners such as SIB (Swiss Institute for Bioinformatics) or in world-wide consortia, like the DNA or protein data resources. Over time, this will be increasingly supplemented by data collected and curated at other ELIXIR Nodes. EBI has moved its core services to the externally hosted data-centres over the last years, and has therefore developed expertise in streamlining the process to ensure that databases and tools are transferred as quickly and efficiently as possible. Such engineering is essential for robust services. For ELIXIR a reasonably large team of software engineers will be needed to develop, deploy and maintain all the ELIXIR services from the London Data Centre and to ensure that international transfer also runs robustly.

Providing Coordinated Access to ELIXIR Data including cloud services

This is the biggest challenge for ELIXIR – to develop an ecosystem of resources, which can be seamlessly traversed so that data can be retrieved from sites around Europe and merged together to answer biological questions. The vision for ELIXIR is to make this happen across Europe for many data resources managed on different sites and by different scientists. One aspect of course is to develop semantic-web compatible resources – and this option will be explored, but in addition, by close collaboration between ELIXIR Nodes and the Hub, also tools interoperability and other methods of data integration should be possible. To achieve this, ELIXIR will require a team of service providers with this goal as their central mission. The coordinated access will also address methods to "bring the tools to the data" - cloud services and standard interfaces for the "transportable algorithms" that are currently actively developed to aggregate and analyse data that for regulatory or ethical consent reasons cannot be integrated.

The technical services in this domain should also, through collaborative efforts, address the interoperability with other research infrastructures, biobanks and seed collections to support genomics driven research in a variety of fields. The requirement for researchers to interact with data resources beyond Europe will require effective global collaboration.

Research data management and BMS Infrastructure services

ELIXIR is coordinating the BioMedBridges grant that provides tools for improving communication and developing standards, ontologies for interoperability as well as establishing key services to enable effective data-bridges between the infrastructures. The long-term maintenance and refinement of these services requires ELIXIR to develop an intimate interaction with all the BMS infrastructures to ensure that data can be easily transferred between them via ELIXIR. One possible option to ensure this support is to allocate a domain-specific bioinformatician to each of the other 12 BMS infrastructures to act as the critical personal link between the stakeholders of each infrastructure and ELIXIR. With such personal contact and advice, each infrastructure will gradually become coordinated, ensuring that common standards and best practice protocols for data security are adopted, the latter being particularly crucial in

the biomedical field. Other options are the establishment of central service function to support the local integration efforts underway within many National Nodes.

Translational Infrastructure

A major driver for the rapid increase in life-science data-volumes and complexity is the rapid adoption of sequencing based and metabolomics profiling in medical and translational research. Personalized medicine approaches as well as broad population studies are underway in most European countries addressing a wide range of complex diseases such as cancer, asthma and metabolic and cardiovascular disease. Corresponding efforts are also underway in agricultural research and from an infrastructure perspective there will be a significant overlap in service requirements. It is expected, through close collaboration with ESFRI infrastructures, IMI projects as well as global initiatives (e.g. TranSMART Foundation, Global Alliance, Sage BioNetworks) that ELIXIR will provide a set of core services that interface and support translational initiatives.

Establishing the Compute infrastructure

The ELIXIR Hub will seek to enable the best compute infrastructure for life scientists throughout Europe. As data sets grow in size, the need to co-locate data and compute will increase. Therefore, it is inevitable that the large data Nodes will also become large compute Nodes and vice versa. ELIXIR will seek to establish the best solutions and ensure good interconnectivity, building where possible on the e-infrastructures already developed in Europe. Developments in areas such as cloud computing and virtual machines must be adopted rapidly throughout ELIXIR. ELIXIR has also engaged in the European Technology Platform for computing, ETP4HPC, to ensure that the needs from the life-sciences community are articulated and catered for.

8 ELIXIR Pilot Actions

ELIXIR Pilot Actions are short projects or initiatives to address key scientific and technical issues. The ELIXIR pilots should leverage or join-up established resources or on-going activities to deliver demonstrator or proof-of-concept studies. 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.

8.1 Principles and process for starting new pilot actions

ELIXIR Pilot Actions are formally endorsed 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 HoN via an email procedure for endorsement by the HoN group 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 HoN 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 of Pilot Actions based on the following criteria (in order of importance):

1. **Portfolio balance and fit to ELIXIR strategy,** strive to give a broad cover to the key areas for ELIXIR;
2. **Delivery date,** to prioritize pilot actions such that they can support the development of ELIXIR scientific programme and H2020 applications;
3. **Pilots that leverage on-going / delivered pilot actions,** i.e. if there are focussed activities that could broaden the utility and/or involve additional Nodes this will be prioritized;
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.

Furthermore, a key criteria for a pilot action is inclusiveness i.e. that the pilot should be open to participation and learning for other Nodes; ELIXIR pilots are not competitive research projects but rather efforts to develop strategies or test-drive technology to inform and shape future ELIXIR efforts and grant applications. 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 other Nodes.

9 ELIXIR Programmes of Work

The scientific and technical infrastructure service delivery within ELIXIR will be organised into a set of broad *ELIXIR Programmes of Work* that will address a set of specific challenges and user requirements. The following section of the draft ELIXIR Programme gives a brief outline of each area together with indicative activities; each of the *ELIXIR Programmes of Work* will be discussed in detail during the Heads-of-Nodes meeting and developed in a Work stream until January 2014. The Work streams are expected to deliver a detailed view of service requirements, including an overview of current and planned services (including Hub and Nodes), as well as specific objectives, deliverables and work plans until 2018.

9.1 Core services & Data resources

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 provision together with the fact that significant volumes of biological data provision is no longer constrained to a small number of laboratories and research institutes, the gaps are significant and will require substantial investment.

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 dependence graphs.

The Work stream of core services & data resources will address and develop:

1. A recommendation on the ELIXIR Kite-mark that addresses issues such as the kite-marking process, requirements and service levels; will there be different levels of kite-marking (ELIXIR Gold, Silver, Bronze?)
2. Agreed criteria for ELIXIR Core resources and the process to identify the set of core services
3. Recommendation on service monitoring, life-cycle management and quality control / peer review
4. A development / enablement path for additional / future core services.

9.2 Tools Interoperability and Service Registry

Identification of tools and resources was highlighted as a key challenge in the report on industry needs and expectations from ELIXIR; to meet this need ELIXIR will develop and provide a *Service Registry* (led by elixir.dk). 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. This will build on e.g. the ELIXIR pilot-project on interfacing the Human Protein Atlas with other proteomics resources. This Work stream will address and deliver concrete objectives and proposals to address:

1. Overview of the Service Registry and how this will interface and exchange with workflow-based systems
2. A plan to address Tools Interoperability within ELIXIR
3. Address the standard, 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).

9.3 ELIXIR Technical Services

Accessing and computing with the large data-volumes from modern biology comes with a specific set of challenges on the technical infrastructure and the *ELIXIR Technical Services* Programme of Work will develop solutions of wide-spread utility in the life-science domain. 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 are also cloud provisioning from several ELIXIR Nodes. One major conclusion from the ELIXIR technical coordinator workshop in May 2013 was that while cloud technologies and virtual machines offer great potential to host a variety of computing needs close to the data, there is an urgent need to develop services and templates on top of the current cloud solutions to meet the majority of user needs.

Further important aspects of the ELIXIR Technical Services will be the close collaboration with translational and biobanking 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.

ELIXIR will clearly have to collaborate closely with the European e-Infrastructures to address user needs and fully integrate into the European technology landscape. An important part of this Programme of Work will be to define the interface and collaborative models with e-Infrastructures.

A final challenge in this domain is to develop sustainable models for resource management and exchange including transnational models for charging capital and operational costs for e.g. cloud provisioning. This work will interface closely with the long-term sustainability Work stream in the ELIXIR Management Programme (section 9.8).

Specifically the Work stream to deliver the *ELIXIR Technical Services Programme* will address

1. ELIXIR Cloud offerings and services
2. Secure data access and exchange
3. e-Infrastructure synergies
4. Resource management and exchange policies

9.4 Research data management

Recognising the importance of open access to research data, and indeed open access to public data overall, there is an increasing requirement from research funders to publish and make the findings from publicly funded research broadly available. There is also another societal aspect of data sharing, particularly pertinent to the health sector, where the release of data that enables an open assessment of patient safety or treatment efficacy is increasingly seen as a basic ethical requirement. An often-implicit assumption in these requirements is that the data will be made available in a machine readable and reusable format – a prerequisite to realise the value of the produced data in new contexts. Hence, the issue of data sharing and re-use is tightly coupled to effective research data management and presence of processes and infrastructure to support data coordination, meta-data curation and deposition in suitable archives. An example of this is the tight coupling between the EBI and Sanger Institute that allows direct deposition of genomic sequences into the Ensembl and Short-read-archives with meta-data assignments from laboratory management systems.

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. There is also an urgent need to establish principles for optimisation of the existing data capacity to meet the demand – such as developing tools and recommendations to effectively archive genomic sequence data.

ELIXIR will also play a key role in the development of practice-based science policy for data management and accessibility and will need to work closely with initiatives such as RDA and EUDAT to ascertain an effective and transparent solution for life-science users. ELIXIR will also engage in DG CONNECT's efforts to implement a pilot action on Open Data in Horizon 2020, sharing best practice from Nodes where appropriate.

This Programme of Work will address and deliver:

1. Overview of national efforts for research data management in the life-science with a concrete plan for practice exchange, policy development and spread of best-practice
2. Plans for effective collaboration and joint efforts with RDA and EUDAT
3. Identification of services and resources (e.g. vocabularies and ontologies) needed to support national programmes.

9.5 Data interoperability, vocabulary and ontology services

The application of bioinformatics analysis and prediction in many life-science fields require the effective integration of data from a large set of different resources; ensuring the standards, vocabularies and services that enable data interoperability is of vital importance to the success of ELIXIR. Indeed this requirement was one of the major conclusions from the analysis of industry needs within ELIXIR that was recently developed with an outside partner (Annex 3).

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).

The Data interoperability, vocabulary and ontology services will deliver a concrete plan and specific objectives for semantic integration within ELIXIR including:

1. Identification of critical vocabularies and ontologies for ELIXIR services and required maintenance resources
2. Identification of key European identifier resolution and data interoperability services
3. Requirements from other BMS RI such as ISBE on data exchange services

9.6 ELIXIR Training Programme

The need for training to ensure optimal usage of the data and associated tools is paramount. Experts based at the Hub and Nodes, working closely with the data and tool developers, will provide the training. This will be delivered either electronically or through face-to-face courses held throughout Europe. The ELIXIR Hub will help to ensure the provision of appropriate training both by providing support for and coordination of suitable trainers in the Nodes, and by supplying an infrastructure for the management of training materials, so that they can be used and re-used by all of ELIXIR's trainers. A key aspect of the training programme is a train-the-trainer strategy to scale with demand throughout Europe as well as supporting capacity building in all the Member States. Most ELIXIR Nodes have identified training as part

of the offered services. The ELIXIR Programme Work stream is expected to deliver a ELIXIR-wide training plan that coordinate training capabilities across the Nodes while recognizing that Nodes have funding and capacity targeting the local needs. The plan delivered by the Work stream will include:

1. Plans for coordination of training to build on specific Node capabilities
2. Plans to optimise a range of training methods including using online resources across Nodes
3. Plans for coordination of training with global efforts such as GOBLET.

9.7 ELIXIR Domain Specific Services

9.7.1 Translational Services

It is essential that ELIXIR is an infrastructure able to support the translation of basic science into medicine and agriculture. A significant part of these activities will occur through our interactions with other BMS infrastructures (e.g. BBMRI), but ELIXIR should be active in its own right in this area. These services will support the significant translational science investments in the academic sector but will also be highly relevant to industrial research in both medical and e.g. agricultural research. Currently many commercial companies make heavy use of the core biomolecular resources, including pharmaceutical companies, the biotechnology sector, largely driven by SMEs, the health sector and increasingly the food industries. As the amount of data increases and becomes impossible for companies to manage internally, it is clear that in the future this use will grow, rather than decrease.

Specifically these services should address the integration of genomics and other high-throughput biomolecular data sources with clinical, in-vivo and phenotypic data. There are currently significant investments in translational data services and data interoperability within the IMI projects and ELIXIR will work closely with IMI consortia, as well as other large infrastructures (e.g. the TranSMART foundation) to address the need for stable translational services.

Building such collaborative projects that lead to robust public data resources and tools will require considerable external investment, and the Hub could help to lead this engagement with specialists for the different sectors. Such funding could pump-prime new developments to establish a proof-of-principle maybe before commercialization. With the emergence of pre-competitive research, there is an excellent opportunity for ELIXIR to work with such companies to develop public domain resources that will benefit all, but provide specific expertise and advantage to Europe's bio-economy.

9.7.2 Food and BioIndustries services

Agricultural research is experiencing a transformation similar to medical and human health research with the opportunities presented by the advances in plant and microbial genomics to realize the long term vision of integrating data from gene banks, plant genomic research and phenotyping efforts, together with meteorological and geospatial data into a coherent and accessible infrastructure. Within ELIXIR there are solid provision of services for plant genomic data as well as in infrastructure for reference genomes and variant sequences. However, there is currently a significant gap in provision of portable systems that could be uniformly rolled out to seed and gene banks. The cost of actually sequencing the stocks is falling to levels where in the 2014-2018 timeframe the data will become available to populate such a system.

Obviously there are many groups worldwide working on aspects of this problem, but there's no obvious home even for the genetic data (EBI is collecting variant information from the public domain for some species, but is resourced to provide an integrative solution to all gene banks at present). Further, the extension to phenotype and environment is not fully scoped, although it's something that will become a big reality in the lifetime of Horizon 2020.

9.7.3 Marine and Biodiversity services

High-throughput biomolecular data-acquisition is fundamentally changing biodiversity, ecological and marine bioprospecting research; projects such as the "Barcodes of life" as well as marine metagenomics project are producing large volumes of data and face significant data management and integration challenges. In addition to the generic infrastructure components there is a need for domain specific services addressing the specific challenges associated with data from a wide variety of species across all taxonomic groups. Infrastructures such as LifeWatch dealing with biodiversity will benefit from the ELIXIR project as well as EMBRC and facilities such as GBIF. This Work stream should articulate a specific plan and deliverables for the domain specific services including how the needs from industrial users and biomaterials research are met.

9.8 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. These Work streams are outlined below.

In addition ELIXIR needs to foster the active collaboration and exchange between the National Nodes and within Programme of Work will address the development of an active exchange programme for scientist and engineers involved in delivery of ELIXIR Services.

9.8.1 Develop ELIXIR Strategy, Performance indicators and Operational Principles

While the value of the individual ELIXIR services is clear with straightforward indicators around scientific uptake and impact there is a need to bring together the different Node offerings into a transparent service offering with uniform performance indicators. This will include a scheme for classifying Nodes by the services provided, a gap analysis for services and clearly present how the services interrelate through the ELIXIR coordination and governance.

While these Work streams will be initiated as a priority they will draw heavily on the outcome from the development of Node Collaboration Agreements and hence is not expected to conclude until end of 2014.

1: Work stream to finalise Collaboration Agreement template and the Service Delivery Plans.

2: Work stream to develop the ELIXIR risk register with mitigation procedures.

3: Work stream to prepare ELIXIR Communication plan including internal stakeholders.

4: Work stream to develop ELIXIR's External Stakeholder Engagement Strategy.

5: Work stream to develop ELIXIR scorecards for Service Deployment and Sustained Operations Phases

9.8.2 Accessible and Open data policy, Software licence policy and Intellectual Property policy

This area was highlighted by SAB in June and requires a Work stream to drive conclusion with clear recommendation for ELIXIR Policy and Charter.

6. Work stream to define ELIXIR Open data and software license policies.

7. Work stream to define and deliver the IP policy of ELIXIR.

9.8.3 Define ELIXIR User Strategy

An extensive analysis of user needs and requirements was performed during the Preparatory Phase but Bioinformatics is a rapidly moving field. To ensure a user-centred perspective on services and the relevance of ELIXIR to life-science research there is a need to revisit the data and refresh input. A significant change in the bioinformatics landscape over the last years is the establishment of *Bioinformatics core resources* in many large research centres. Maintaining an overview and good contact with these groups is of paramount important to ELIXIR.

8. Work stream to set up User Survey.

9. Work stream to define ELIXIR User Strategy.

9.8.4 Develop the ELIXIR External Funding Strategy

While ELIXIR has a core long-term funding from the Member State contributions, this is a complex area where the ELIXIR funding is likely to come from a portfolio of sources. The AEG recommended a Work stream to develop and refund the External Funding Strategy, led by an experienced Science Policy Advisor and initiation of this work will be a priority for ELIXIR.

10. Work stream to develop the ELIXIR External Funding Strategy.

9.8.5 Investigate and propose Ethical Framework for ELIXIR

Handling of personal genomics data has stringent regulatory and ethical requirement and while there are extensive activities within BioMedBridges to develop both a Data Security model as well as recommendations on an Ethical governance framework, there is a need for ELIXIR to analyse how to implement this within a distributed infrastructure. This Work stream should deliver a recommendation to the Board on how to implement Ethical Governance and review within ELIXIR.

11. Work stream to investigate and propose Ethical Framework for ELIXIR.

9.8.6 Devise implementation plan for ELIXIR Industry Strategy

As outlined in Section 6 of this Programme there are a number of recommendations from the Industry analysis performed by ConnectedDiscovery (Annex 3). Two specific Work streams will be started on the basis of this analysis:

12. Implementation Work stream for ELIXIR Innovation and SME Programme.

13. Work stream devise implementation plan for ELIXIR Industry strategy, including formation of Industry Board.

List of Annexes

- Annex 1: The ELIXIR Action Plan (recommended to be printed on an A3)
- Annex 2: The ELIXIR Risk Register
- Annex 3: The report: "Developing ELIXIR Interactions with Industry" delivered by ConnectedDiscovery.

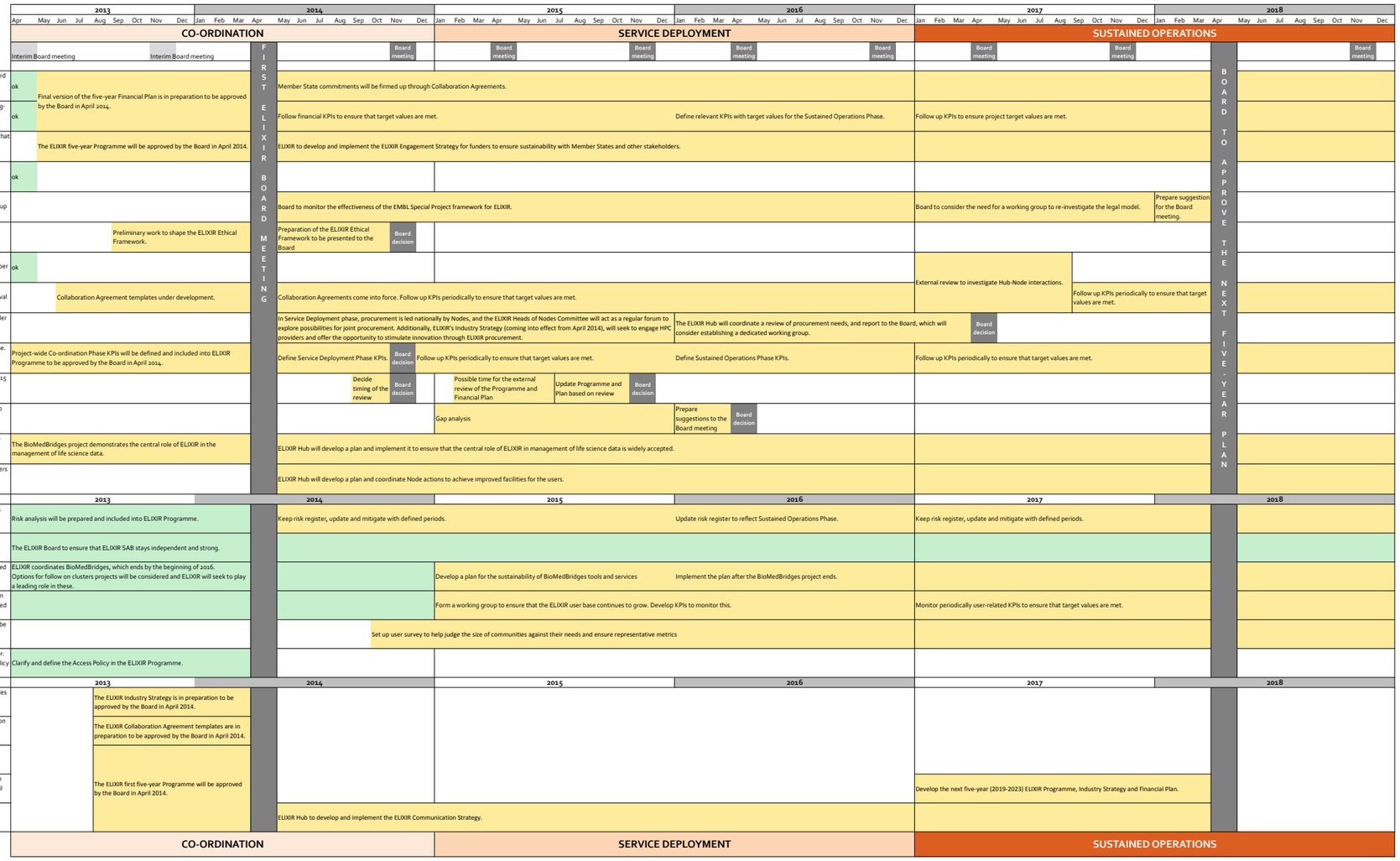
ELIXIR Action Plan based on recommendations from the Assessment Expert Group

9. October 2013

Type	AEG report observations on weaknesses, specific to ELIXIR	ELIXIR Action
Financial	Out-dated financial figures in the report.	Updated figures for the Hub five-year Financial Plan were provided to the Interim Board in April 2013. Updated figures for Nodes will be provided with the Collaboration Agreements.
Financial	Plans for a stronger long-term financial relation with Member States and national efforts should be considered.	In the five-year plan, provided to the ELIXIR Interim Board in April 2013, a plan for long-term financial relation with Member States was presented.
Financial	The organisation of long-term funding should be treated as a work-package of high complexity, led by an experienced science policy maker.	This is a key role for the ELIXIR Hub, and will be included into the ELIXIR Programme that is currently in preparation. External expertise will be sought where necessary.
Legal	Clarification of the double role of EMBL as partner and contractor is needed.	The ELIXIR Consortium Agreement (ECA), which was approved 23.4.2013, clarifies the role of EMBL.
Legal	It should be considered whether a future move to ERIC is realistic, or if the present legal framework is sufficient in the long term.	The ELIXIR Board will monitor the effectiveness of the EMBL Special Project and will during the Sustained Operations Phase consider establishing a dedicated working group to investigate an alternative legal framework.
Legal	An overarching Ethical Board is recommended, especially in the area of data access policy, where both ethical and data protection issues should be taken into account.	The Hub will prepare a suggestion regarding the Ethical Framework of ELIXIR to the Board, which will decide on the procedures.
Managerial	A plan to strengthen gradually the Hub functions should be developed.	Since May 2013, the ELIXIR Hub has grown and will shortly number nine full and part time members of staff. Further, the five-year Financial Plan includes increase in Member State contributions to enable a stronger Hub.
Managerial	The role of the hub is weak, it will not be able to impose standards or ensure the coordination of the distributed infrastructure.	In June 2013, ELIXIR formed a Collaboration Agreements working group to draft the agreement, including KPIs to monitor progress. The templates will be ready for approval in the April 2014 Board meeting.
Managerial	A more structured approach to organise procurement and recruitment in order to maximise the project impact while minimising costs. An international procurement coordination task force might be envisioned.	When the project enters the Sustained Operations Phase, the ELIXIR Board will consider establishing a dedicated working group for this.
Managerial	List of KPIs provided, but no target values of KPIs are available.	Project-wide KPIs for the Co-ordination Phase will be defined in the ELIXIR Programme. Service Deployment and Sustained Operations Phase KPIs will be defined later, once Collaboration Agreements between the Nodes and the Hub are in place.
Managerial	Once an activity plan and related budget breakdown is developed, it is recommended that it is independently (externally) reviewed.	The Board will decide when the external review is performed, likely to take place in 2015 during the Service Deployment Phase.
Managerial	A top down-analysis to identify possible gaps in the infrastructure. What the project as whole wants to achieve in terms of science and societal impact?	The ELIXIR Hub will perform the analysis during 2015. Further, the ELIXIR SAB will also be asked to identify 'service' gaps in ELIXIR infrastructure.
Political	Getting the central data role of ELIXIR accepted outside of the traditional EMBL constituency will take time and effort.	This is a key role of the ELIXIR Hub, which will continue to work towards achieving the goal. Activities such as annual ELIXIR AGMs are planned and these will help to engage users from all constituencies.
Political	ELIXIR should continue and further develop its leading role in coordinating all biomedical research data related activities leading to improved facilities for the users.	This is a key role of the ELIXIR Hub, which will coordinate the Nodes to ensure that users are provided with improved facilities. The Nodes will develop strategies to ensure this nationally.

Type	AEG report observations on strengths, specific to ELIXIR	ELIXIR Action
Legal	Risk in data protection properly taken into account.	The ELIXIR Programme will include a project-wide risk analysis, and the risk register is planned to be updated regularly.
Managerial	Strong and independent SAB.	The ELIXIR Board and Member States will ensure that the SAB stays independent and scientifically strong
Managerial	ELIXIR has an important role as the leader of BioMedBridges.	The sustainability of tools and services developed within BioMedBridges will be planned with ELIXIR and the other BMS RIs during 2015.
Users	Large and growing user community, extending beyond EMBL-EBI.	A working group will be formed during Service Deployment Phase to develop a plan on how to expand the user community. User-related KPIs with target values will be defined for the project.
Users	The survey from preparatory phase describes user needs.	The survey will be updated to ensure that user needs are being met. New surveys will be undertaken when necessary, e.g. during the gap analysis. User-related KPIs will be defined for the project.
Users	Access is free for academic, controlled access for industry.	ELIXIR provides free access to its tools and service for all users, including private sector. The access to data in Personal Medical Records is controlled, and the ELIXIR Ethics Policy will consider this.

Type	AEG report, general remarks	ELIXIR Action
Industry	Consider forming an Industry Advisory Board if industry is a demand-pull partner.	The ELIXIR Industry Strategy, to be approved by the ELIXIR Board in April 2014, includes the establishment of an Industry Board, which will include representatives of users, suppliers and publishers.
Legal	Ensure the protection of IP.	This is a key topic, and a plan to protect IP will be included into the ELIXIR Collaboration Agreement templates that are currently being developed with IP and technology transfer experts.
Managerial	Investigate synergies in building up common operational ICT platforms and address aspects of storage.	The BioMedBridges project already includes ICT partners. A plan for further synergies will be included into the ELIXIR Programme. The plans for storage solutions will be included in the Programme.
Managerial	Develop activity plan with deliverables, WBS, KPIs, Financial Plan, and risk analysis, including a coherent European-level action plan to demonstrate the value added by the infrastructure.	The first five-year ELIXIR Programme is to be approved by the Board in April 2014. The second five-year ELIXIR Programme (2019-2023) will be approved by the Board in April 2018.
Political	Develop a Communication strategy.	This is a key role for the ELIXIR Hub, and the plans to establish a project-wide Communication strategy will be included into the ELIXIR Programme.



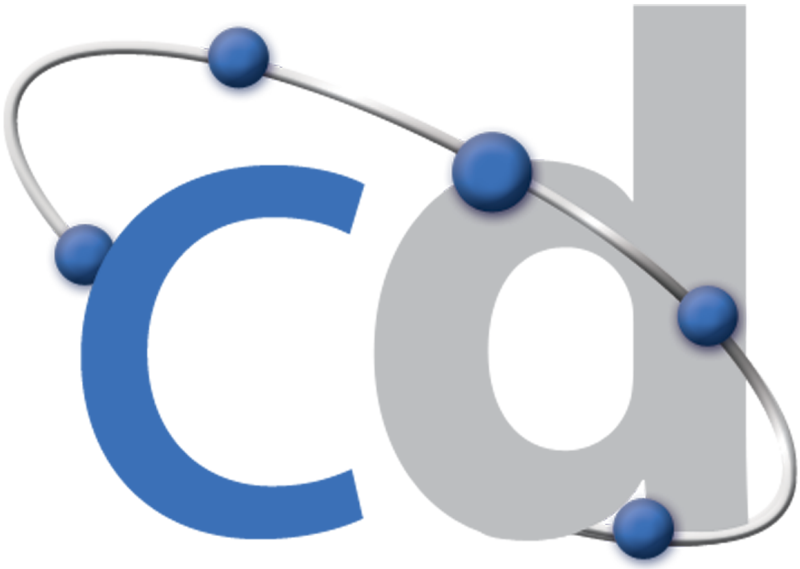
ELIXIR Risk Register

LAST UPDATED: 09/10/2013

Risk ID	Category	Priority	Risk short name	Risk description	Possible effect	Mitigation measures
1	Operational	1	IP Infringement	IP infringement, could be deliberate or non-intentional.	Legal consequences	IP clauses in the Collaboration Agreements.
2	Operational	1	Data protection	Reputational risk, e.g. related to data protection/ confidentiality.	Legal consequences	Devise and monitor Data Security Policy in collaboration with BioMedBridges.
3	Operational	1	Service failure	Reputational risk, e.g. services fail.	Users lose trust in the services provided by ELIXIR.	Build up process for service monitoring and quality control.
4	Financial	1	Cash flow	Issues related to cash flow, e.g. Member States paying at different time points during the year.	Budget imbalance due to currency changes.	Timely invoicing. Build up cash reserve during early phases of operation (>25%).
5	Political	1	Stakeholder engagement	Failed to effectively engage with stakeholders.	Benefits realisation fails.	Develop Stakeholder Engagement Strategy to make sure to interact with the right stakeholders.
6	Political	1	Reputation	Reputational risk, related to technical issues.	Stakeholders loose trust in ELIXIR.	Define communication strategies, corporate social responsibility policies.
7	Political	1	Ethical	Ethical risks	Stakeholders loose trust in ELIXIR.	Develop ELIXIR Ethical Framework, consider establishing an external Ethical Board.
8	Users	1	Added value	Users not using the services.	The added value of ELIXIR is not realised.	Develop User Strategy based on the existing preliminary work.
9	Operational	2	No delivery	Programmes of Work not delivering the set deliverables.	Delays in Programme progress.	Define roles and responsibilities clearly, create a WBS to be attached to the Programme.
10	Financial	2	Currency	Currency exchange rate changes between the different currencies used by the Member States and the Hub.	The Hub receives less money than planned in the budget.	Move the ELIXIR Hub budget into GBP.
11	Scientific	2	Scientific value	Fail to show the scientific value of ELIXIR.	The impact of ELIXIR fails to meet its full potential.	Engage with the user and stakeholder communities, develop a Communication Plan. Develop KPIs to demonstrate the value.
12	Technical	2	Cloud computing	Risks related to cloud computing.	Users lose trust in the services provided by ELIXIR. Legal consequences.	Devise and monitor Data Security Policy in collaboration with BioMedBridges.
13	Financial	3	Leaving Members	Countries drop out of ELIXIR.	Decrease in the capacity to implement the common strategic goals.	Articulate the value of ELIXIR clearly.
14	Financial	3	Lack of new members	Countries do not join ELIXIR.	No capacity to implement the common strategic goals, role of the Hub not fulfilled.	Develop Stakeholder Engagement Strategy.
15	Organisation	3	Authority	Lack of effective execution of authority.	Project is delayed, lack of decisions to move forward.	Ensure that responsibilities and authority are clearly defined.
16	Organisation	3	Node communication	Lack of communication through nodes.	Failure to deliver the objectives.	Develop Communication Plan. Ensure that HoN Committee and Technical Coordination Group function effectively.

Developing ELIXIR Interactions with Industry

A commissioned project quantifying opinions and understanding of ELIXIR amongst researchers in the private sector



Lee Harland and Bryn Williams-Jones
Connected Discovery Ltd
http://www.connecteddiscovery.com

Key recommendations:

- Current understanding of the scope and activities of ELIXIR amongst industry stakeholders is limited and confused. Clarity around the scientific programme and service footprint is urgently needed.
- ELIXIR needs to clarify its position with respect to data interoperability. Will this be a major element of the initiative, and if so what activities and standards will be used?
- ELIXIR should hold a workshop to address the question "What Is ELIXIR for Industry". This would have the main aims:
    - Present the role of ELIXIR services, what "services" actually means, what levels of support will be provided. Critically, the workshop should feature presentations from as many nodes as possible on what they do, and how they could engage with industry.
    - Define the technical outline of the granular needs, expectations and requirements of ELIXIR services from the industry partners.
- ELIXIR should establish a dissemination network, whereby industry-relevant activities across the nodes are brought together and disseminated to all relevant parties (big industry, SMEs etc).
- ELIXIR should develop a set of standard legal agreements to simplify service development and consumption, and collaboration building.
- ELIXIR should develop a focussed SME outreach programme that leverages the node networks and fosters a collaborative network for open services development.
- ELIXIR should develop SME Embassy Cloud sandbox environments to drive innovation and tool development.
- ELIXIR should develop specific interest groups for example, engaging with publishers, HPC users, or senior figures within industry. Such working groups can provide a much more focused discussion.
- ELIXIR should specifically develop its vision in the context of other global initiatives, and in particular IMI. Industry should be able to understand how ELIXIR fits into the bigger picture AND how ELIXIR will interface with and differs from other global resources.
- ELIXIR should develop a "kite-mark" and start to brand existing services with this mark, to start to convey what is part of the ELIXIR service network.

Background, Introduction and Methods

Biology is entering an era of big data, and for individual scientists and teams in applied biological research this creates multiple challenges: relevant reference data needs to be identified, often diverse types of data across disciplines needs to be integrated and analysed and visualised in the context of newly generated findings. The volume of data now available has shifted the knowledge bottleneck back to data handling and integration, with the majority of computational scientists now being swamped with large amount of difficult to handle data. The challenge remains in turning this into actionable knowledge that can impact biological research in ways that are understandable and workable to the non-expert computational scientist.

The wide uptake of big data approaches combined with a growing focus on open innovation precompetitive approaches, the emphasis is shifting to horizontally integrated data infrastructures, which are focussed on answering key scientific questions. These questions range across different data types, and crucially embrace the diversity of data producers and consumers involved in life sciences today.

European companies across multiple industry sectors (large and small pharma, biotech, consumer healthcare, agrochemical, scientific publishing, content, and software providers) are a key consumer of public biomedical data. Yet, the efficient use of these resources by industry is inhibited by a number of challenges such as discovery ("what resource do I need?"), exploitation ("how do I query this?") and integration ("how can I gain a more holistic picture"). Continuous strategic interactions between ELIXIR as a research infrastructure for biological data, and commercial data and service consumers in these sectors will be essential for the successful delivery of an actionable implementation plan for ELIXIR.

Industry researchers are a critical partner in, and user of ELIXIR infrastructure. As such, it is vital that the needs of industry and their use cases are fully understood. Building on previous work in the preparatory phase (2009) we have carried out a limited project interviewing key personnel from across the life science industry to understand:

- What they know about ELIXIR
- How they want to engage with ELIXIR
- What ELIXIR needs to provide and develop to meet the needs of Industry
- How they want to participate in the development of ELIXIR
- What is the best model for continuous ELIXIR-Industry interaction

Over the course of the last 3 months we have interviewed around 40 senior industry researchers and informaticians (stakeholders) to work through a series of questions to understand their knowledge, expectations and requirements of ELIXIR as an infrastructure for biological data.

For the purposes of simplicity, throughout this report 'Industry' is used to represent users of data, services and infrastructure from the food, pharma,

biotech, publishing, and consumer healthcare sectors, and from organisations of varying size from global corporations to SMEs.

Many of these topics that emerged from these interviews are developed in more detail individually in this report, but representative comments from stakeholders include:

- What are the ELIXIR nodes, and what is their resource, service and tool portfolio – how do I find out who's doing what? What is the process for analysing gaps and developing new node capabilities that will meet my needs? In particular, I have an idea, how do I "talk" to ELIXIR about it?
- The industry is very dependent on SMEs and BioIT companies; yet academic outreach to these is very patchy. How can Industry best partner with ELIXIR to bring these together in a way that benefits all concerned?
- The current infrastructure is very fragmented, and there is a lot of inertia to overcome to enable real interoperability of data. What concrete steps are planned for the scientific programme to address this, and how can industry bring their experience to the table? Specifically, can ELIXIR spell out its mission around data integration, as well as service provision "what difference will ELIXIR make here?"
- Is there real buy-in from the node academics that industry is a key partner in developing future infrastructure, or do they see ELIXIR as source of funds alone?
- Industry support for services usually boils down to simple questions
    - Is this valuable enough that we would spend money on it?
    - Is this something that the business (end consumer of scientific data/analysis) wants?
    - The industry pays for things that are valuable, and can really help ELIXIR develop services that add value
- At present, ELIXIR seems valuable only as a tool for computational scientists. More needs to be done to market ELIXIR as an initiative that has wider benefits to patients, farmers, and customers.
    - With this in mind, what is the mechanism for a researcher as an end user making needs clear to ELIXIR? How do the nodes plan to engage with the real consumers of biological data and understand their needs?
- Industry needs real help in understanding Horizon 2020 and engaging the academic community to develop new projects and collaborations. What role will ELIXIR play in facilitating this? Could ELIXIR develop a mechanism to move ideas to real funded projects? Industry can't do this alone, but it could be a vital win-win for both parties.
- How does ELIXIR fit with the push to semantics, widely interoperable data, and develop capabilities and services that are greater than the sum of the ELIXIR parts? How can industry assist this?
- ELIXIR is part of a complex global landscape of initiatives, projects and collaborations. What are the unique features that would influence Industry to engage at a meaningful level? At present, public domain services are maintained and managed mostly independently of

industry, who come along and consume them. Does ELIXIR offer a different role for industry engagement?

Key Themes:

1) Clarity on the ELIXIR Mission, Scope and Relationship to Current Landscape
Many of the stakeholders are confused about the mission of ELIXIR and its relationship to the existing biological data infrastructure and services landscape. Whilst there is understanding that the current focus is on securing the ELIXIR Consortium Agreement, there is a wide feeling that for Industry to engage constructively with ELIXIR, there needs to be concerted outreach with concrete facts and information on how ELIXIR will be implemented and delivered.

Many stakeholders highlighted interoperability of data as the major requirement for ELIXIR. For ELIXIR to be a success for Industry, it really needs to address horizontal interoperability across the multiple public data silos. What is not clear to many Industry stakeholders is that the need for interoperability is also understood as a priority from the academic side. In large part, this is due to inconsistent messages from industry about what the essential services are and clarity around how these services should be defined, provisioned and operated. Many stakeholders expressed a need for, and willingness to participate in a focussed workshop to collectively define industry needs and requirements of services aimed at cross-ELIXIR data interoperability. Critically, ELIXIR must spell out how it would transform this space, rather than continuing with the status quo.

1a) Specific Concerns Regarding ELIXIR and EBI
Many of the stakeholders interviewed were unclear as to the relationship between ELIXIR and the EBI, and how this would evolve as ELIXIR develops. Whilst they recognize the value of a wider network of services and tools, there needs to be clear demarcation of responsibilities between ELIXIR and the EBI particularly – much more so than is the case for other nodes. The development of the "ELIXIR Brand" is crucial here and needs to start sooner rather than later.

Stakeholders who are part of the EBI Industry Programme (EBI-IP) are clear in their desire to ensure that ELIXIR Industry engagement should be synergistic with the EBI-IP. They also recognize a clear and pressing need to engage more widely across ELIXIR, and include commercial organisations of different sizes and from sectors. These stakeholders also emphasized that the EBI-IP should be used to share best practice for industry engagement across nodes, focus engagement with SMEs, and drive outreach to service providers alongside ELIXIRs own capabilities in the long term.

For stakeholders who are not part of the EBI-IP, there is real confusion around the relationship of ELIXIR and the EBI. Again this speaks to the need for concerted outreach and engagement with the broader Industry community with concrete information around the capabilities of ELIXIR once the ECA is signed. In the short term, this outreach should focus on differentiation of ELIXIR and EBI

services and establish a roadmap for Industry engagement in the ELIXIR scientific programme.

There is some confusion as to whether the Embassy Cloud (see later) is an EBI project, or an ELIXIR project. In particular, those who have regular contact with the EBI had a lot more questions around how could resources beyond the EBI could be built into Embassy Cloud capabilities. It was clear that there was much excitement and support for the Embassy Cloud initiative, providing an excellent opportunity to build on this interest for the future.

In order for wide Industry support of ELIXIR, the node make-up, service footprint, differentiation and scientific programme needs to be clarified and communicated as soon as possible. ELIXIR needs to provide a clear, succinct "capability footprint" that describes the activities and services across the nodes.

2) Services, Standards, SLAs and Agreements

For ELIXIR to gain wide industry engagement, the prime focus has to be on services. A common question was "will all services from the node be at the same level"? In particular, will levels of support be consistent across services, or will there be a "Gold/Silver/Bronze" type grading? For critical services, ELIXIR should work with industry to define what an 'Industrial Grade' looks like, particularly expectations about security, uptime and service level agreements. As the technical services develop, any services that the Industry could consume should have clear information available around each of these points. Combined with this, the industry needs clear routes to work with the nodes on the development of new services either locally or through the ELIXIR technical hub. A mechanism for the identification of gaps, and collaborative development of new services is also required and close industry involvement in this should be written into the ELIXIR scientific plan.

Many stakeholders emphasised that the provision of ELIXIR services should be driven by demonstrable business cases – i.e. address the real pain points for a particular service consumer who in many cases would be willing to fund or co-fund development. Where this is done, it should be tracked and collated for dissemination to demonstrate the real-world value of ELIXIR.

The lack of these real service capabilities is currently the major barrier for wider Industry engagement with the data infrastructure as it exists today, and drives the mirroring and internal integration activities that many stakeholders are forced to undertake. This is arguably inhibiting the development of scientific use case driven services in the public domain at present, and inhibiting innovation. To some extent, this is due to inconsistent messages from industry about what the essential services are, and clarity around how these services should be defined, provisioned and operated. The EMBASSY Cloud initiative offers great potential for industry users to discover use-cases and say "I'd like one of those too please".

For ELIXIR to successfully bridge this service gap, then a comprehensive suite of standard service agreements, security policies, collaboration agreements, MTAs, Public Private Partnership agreements *etc.* that are used across the ELIXIR nodes will be needed. The internal overhead in individual negotiation is a real limit for many stakeholders, and ELIXIR would benefit greatly from work to put these key-enabling pieces in place.

Most of the stakeholders indicated willingness to fund and 'buy-in' services that meet the 'industrial quality' criteria – in fact would welcome the opportunity given internal resource constraints. The current landscape of 'academic best efforts' is understandable but such services need to be clearly differentiated from the many that are provided under a much more comprehensive service policy. The path to transforming a service from "best-effort" to "fully-supported" needs to be laid out.

Beyond service standards, many stakeholders highlighted the need for ELIXIR to visibly take the lead in the development of comprehensive data standards. As interoperability of data is perceived as a real bottleneck in the applied use of multiple resources for industry users, there would be real support for the development of robust but workable and pragmatic data standards under the ELIXIR banner. ELIXIR could play a big role in bringing together both the academic and industrial community to develop workable standards that drive wider use. There was also willingness to explore how semantic technologies could simplify data interoperability in areas perceived to be of very high industry value such as translational biology/medicine. There are clear examples of this across the ELIXIR nodes, yet even in the same companies some were unaware that their colleagues were engaged in this activity with an ELIXIR node. This is another area where dissemination of ELIXIR activities to industry could be of great benefit.

Some large stakeholders highlighted the need to develop a comprehensive set of services that have a real applied focus, and that Industry would be very willing to leverage their experience in mapping the need for services in key areas for future development.

3) Embassy Cloud

Several stakeholders identified the Embassy Cloud (EC) as a key feature of ELIXIR. For larger organisations the attraction of on-call compute was seen as very positive and welcome development. Some indicated that internal restraints – either hardware/configuration or funding – were limiting their ability to operate with big data. For smaller companies, and particularly SMEs the Embassy Cloud capabilities could be transformational and really drive innovation. ELIXIR needs to develop a few "shining examples" of where SMEs engaged with this infrastructure to drive their business forward.

One key requirement highlighted is the need for properly organized interoperable data to be made available alongside the Embassy Cloud given that most of the use cases for biological data are very I/O intensive. There is

considerable scope for shared learning opportunities for ELIXIR from companies who have wide experience of commercial cloud providers and the idiosyncrasies of biological data. Several organisations involved in the Pistoia Sequence Services project highlighted this as a particularly relevant learning for ELIXIR. There were many more technical questions around this initiative and it may form the basis for a specific ELIXIR workshop in the future. In particular, questions were asked whether all ELIXIR data will be available via the EC, and if not, what the policy here was. Furthermore, how could industry contribute back to this infrastructure, with either learning or software?

As an alternative means for users to consume ELIXIR services, there was clear understanding that a centralized 'industrial grade' infrastructure must be a cornerstone for ELIXIR to gain wide industrial uptake. Several stakeholders also highlighted that Embassy Cloud instances could also provide a method of industry 'publishing' or otherwise releasing private data into the public domain over time. Also, successful implementation of Embassy Clouds would make it far simpler to develop 3rd party tools and services precompetitively, and drive innovation in new areas.

3) ELIXIR and Engagement with SMEs

For SMEs, the current infrastructure landscape is very confusing, particularly if they are not computational or data specialists. Arguably the benefit of ELIXIR for smaller companies is far greater than large companies who usually have their own internal data capabilities.

For scientific service/analytics companies, there is little awareness of, or drive to use resources that are not 'the big obvious public databases'. To go beyond these into new resources there would need to be real clarity around the cost-benefit. For ELIXIR to drive uptake of services to 'non-obvious' data sources, there needs to be very visible and published use cases clearly highlighting the value of going beyond the obvious resources. This would also greatly help outreach to a wide community of users who may be remote from the nodes, and/or conference communities.

The development of Embassy Cloud capabilities generated a lot of interest from companies of all sizes, but particularly amongst SMEs. One of the real pain points emphasised by SMEs is integrating and handling the essential public data. The development and availability of sandbox environments in the Embassy Cloud, with organized public data sets through ELIXIR would be a significant enabler of SME algorithm or service innovation. This also brings about the question of human-support – while SMEs understand the need to have a certain level of technical proficiency to work in this environment, how much support would be provided by ELIXIR to assist SMEs making the most of the data? And would this support be technical (how to configure) or scientific (how to exploit) or both? Several SMEs highlighted that funders were quite clear in their expectation that publically funded data should be made widely available, but were really struggling to find partners who could help non-specialists to do this. Again, outreach to SMEs through the nodes could really impact SME innovation whilst

making more data available through repositories and services to the whole community.

There is a clear role for ELIXIR to play in the development of outreach activities to SMEs. The node network is ideally placed to locally highlight ELIXIR services and capabilities to biotech clusters. By sharing best practice of the EBI-IP which has previously received limited funding for SME activities, there is a real chance to build a vibrant and innovative SME programme that will have benefits way beyond the use of ELIXIR services. Beyond the node network, there are also several organisations for bioregion clusters/hubs and SMEs that ELIXIR should join and leverage to facilitate outreach to SMEs. We note ELIXIR is already engaging with biotechnology groups such as OneNucleus in the UK. This was encouraged by respondents, however, ELIXIR should ensure it has a clear answer to the question of technical and scientific support as well as support of "end-user" scientists. ELIXIR should monitor the rapidly developing "bio-incubator" sites across Europe and connect to these via the nodes to feed back specific needs and use-cases.

4) ELIXIR Special Interest/Advisory Groups

In addition to the specific recommendations already discussed above, several stakeholders highlighted the need for ELIXIR to develop interactions with constituencies beyond traditional data/infrastructure users. Alongside regular interactions, periodic interactions with a limited number of special interest groups was suggested by several stakeholders:-

4a) ELIXIR and High Level Industry Thought Leaders

Given the state of flux within the pharma industry, and the rate of change in other sectors, the biological infrastructure needs of industry as a whole is something of a moving target. With increased externalisation activities, precompetitive working and the development of public service infrastructures, ELIXIR needs to develop a close interaction with high level industry thought leaders. These thought leaders are senior computational/informatics/data specialists or budget holders that have strategic influence over the development of data capabilities in their own organisations.

Currently it is not clear that any node in the ELIXIR network has a well-developed strategy to engage with high-level industry leadership and gauge how they see needs changing and evolving in a 5-10 year time frame. Whilst there are forums that do this in the Pharma industry (PRISME and others) ELIXIR needs a wider focus and should seek to develop a strategic industry group to provide high-level strategic guidance. This could also help position ELIXIR in the complex global environment (see later) where managing competing needs across multiple external projects usually falls to these industry thought leaders.

4b) ELIXIR and Publishers

Publishers really support the use of existing data resources and infrastructure, and are faced with a complex environment of standards and resources that brings a large overhead for both themselves and authors. Differing standards,

submission pipelines, archiving and curation rules have developed into a bottleneck that is limiting willingness to take on new resources. This also drives a move to 'single resource' publications to keep handling simple, but doesn't necessarily reflect the science. Curation can have unintended consequences if provenance is not retained, and some very niche resources are difficult to extract meaningful data from.

Publishers are also really starting to develop an interest in horizontal integration of resources and data interoperability, and would engage with ELIXIR if this were a central focus. There are also emerging moves towards precompetitive projects that ELIXIR could underpin and support. Arguably the landscape facing publishers as producers and consumers of data is similar to many industry data users, and they represent an important constituency of ELIXIR service users and collaborators. Engagement would significantly aid the understanding of the ELIXIR "brand" if links and data were to be branded on these publisher sites.

5) ELIXIR as a Network for Access to Expertise

There is a clear role for ELIXIR in developing networking opportunities for academics, industry scientists, SMEs and other biological data users. With the increased emphasis on precompetitive working from big-pharma starting to drive change in other sectors, the majority of stakeholders indicated that precompetitive collaborative development was expected to play an increasing part of their activities.

The key enabler of precompetitive development is a forum for discussion and idea development with a neutral organisation/honest broker and many stakeholders identified this as a key role for ELIXIR. This has attractions not just from the industry perspective where collaboration and consortium participation will be an increasing part of 'business as usual', but also meet the need of academic researchers who can often struggle to find the 'right people to talk to in industry'.

For emerging technologies (e.g. Next Generation Sequencing for Agriscience) there are big capability gaps in the industry that are both training and consultancy needs. The usual industry practice is to bring an expert in, but it can be challenging to find the right expertise. Through leveraging the node network, ELIXIR could play a real role in connecting scientists with the relevant expertise.

Taken together with other recommendations, this is more of a soft deliverable and of course should not be the sole focus of ELIXIR. However, it should be seen as a natural consequence of open infrastructure development with collaborative partners, and for many organisations is seen as a key benefit of ELIXIR. Thought should be put to how ELIXIR can enable these collaborative interactions, and how to capture impact stories for future learning.

6) ELIXIR, IMI, ESFRI Projects and Other Initiatives

For many global companies, ELIXIR is part of a complex network of initiatives and project that have to fit into a value-based investment strategy. For some companies, internal computational experts are limiting and so they have to make complex choices around which projects receive support. Whilst it currently seems harder to generate funding for big public infrastructure in the US than Europe, this could of course change. ELIXIR needs to develop a global map of the ecosystem, coordinate with emerging initiatives and projects, and focus on tangible differentiators around services and delivery. Many interviewees asked "has ELIXIR spoken to …" and "how does ELIXIR fit with …". Such a map would immediately address this and provide confidence to industry that ELIXIR saw itself as part of this ecosystem. Critically, industry leaders and ELIXIR should develop a vision around European data infrastructure might look for both Horizon2020 and future IMI initiatives.

Similarly, the industry has already made big investments in some areas, and ELIXIR could play a role in sustaining and further developing these. Several IMI projects are generating tools, infrastructure and data, and could be partners for new grant activities with ELIXIR. The ESFRI landscape is complicated, and many industry stakeholders indicated that they are disheartened by lack of delivery in some of these. ELIXIR should hold specific conversations with relevant members and major projects to develop plans and statements around interactions.

In the US, organisations such as the Genomic Alliance, Sage Bionetworks and Broad Institute offer some valuable services to commercial users. Several stakeholders asked whether ELIXIR was related to EMBNET - a previous node-driven network of computational resources. Similarly, Agrinet is an emerging precompetitive network for chemical biology in the crop science space. Understanding how ELIXIR fits with initiatives such as Force-11, Pistoia, and many others is also important to industry as they map out their interfaces to the public service space.

The ecosystem is large and complex, and ELIXIR should develop specific plans to map and define interactions with infrastructure and data initiatives globally. Many of the industry stakeholders indicated a willingness to help, and that this in itself would be a valuable benefit of ELIXIR to them. For ELIXIR to successfully engage across industry, clear definition of value compared to alternatives in the global landscape will be essential. Again for SMEs, this is a particularly complex and resource intensive problem that ELIXIR could have a real impact on.

7) ELIXIR and High Performance Computing

There are very divided views on HPC amongst the stakeholders interviewed. In part, this is because everyone's definition of "HPC" is different. Some see classical, large scale compute farms for data intensive jobs. Others see HPC as Amazon-cloud like where they can increase their compute capacity on demand without acquiring physical hardware. In some ways, this is more industry's issue than ELIXIRs, but ELIXIR should tread carefully in this space, understanding that there is much confusion in this area.

Whilst some can see the value of elastic access to resources for particular use cases, others highlight that HPC itself is an immature technology for life sciences. Industry practice is patchy and not well disseminated throughout a particular organisation. Thus, asking the question "are you interested in HPC" will differ vastly with different individuals from the same organisation. Use-cases tend to be very specific, and more compute intensive simulations are not that common beyond a few well-defined use cases in chemical docking. Interestingly, those organisations with some experience of HPC indicated that the issues encountered were never down to a lack of compute but rather of configuration or vendor specific issues. A similar view was echoed by many who reported that their problems were in connecting many diverse data, rather than processing one large, single type of data.

Several larger companies have their own collaborations with supercomputing centres focussed on specific use cases i. e. protein folding and ligand docking, or mapping neuronal connections in the brain. These are generally seen as non-core activities and long-term research projects rather than capability building. There is both an opportunity and a risk here for ELIXIR. On the positive side, this is a new, emerging area and so ELIXIR can in some way take the lead by identifying opportunities to engage industry at an early stage. These can be technical – for instance describing the latest developments in this infrastructure would be of great benefit to industry users setting something like this up at their site. Or, scientific, providing real-world use cases where HPC has been used to solve a scientific problem. ELIXIR should consider a workshop in the future when there are enough examples from both categories to present. However, ELIXIR should understand that the vast majority of correspondents did not identify this as an area of major interest. It is something that is very important to a few, and not important at all to the many. It was generally agreed that ELIXIR should promote the use of HPC but not be seen as being "about" HPC. This could alienate a significant body of potential users if the two were to become synonymous.

There is recognition in that in the longer term should sufficient semantically interoperable data become available then activities like disease network mapping and other more blue-sky use cases would be interesting. At the moment, most users were concerned that investments in this area are risky, may struggle to generate return on investment even as part of larger consortia, and that vendor lock-ins and other factors are a real risks. If ETP4HPC did develop then there would be advantage in ELIXIR providing use cases and guidance for life science HPC, but the risk should remain firmly with the HPC vendors. It is however necessary for ELIXIR to maintain a watch on this area and become the first port of call for life science HPC so that developments in future could be undertaken with close industry involvement. But the bottom line is, that ELIXIR could definitely play a role in showing how HPC is evolving, but that this should be presented to the *right* people in industry to gain the right traction and value.

8) Other Important Questions:

In addition to the themes already highlighted, several important individual observations or questions are worthy of further consideration in the development of the scientific programme:

- What is the process for ELIXIR users to nominate services that could be part of the ELIXIR, but are not provided by an ELIXIR node?
- How can industry work with ELIXIR to develop service key performance indicators that can drive funding in future?
- How can the industry help ELIXIR nodes develop from being product centric into customer centric organisations?
- What role is envisaged for data angels/eBiologists (specialists with both scientific and computational skills to bridge the gap between developers and users) in ELIXIR. Several organisations have experience with how this focus can drive service use cases and uptake, how can they help ELIXIR do the same?

About connecteddiscovery

connecteddiscovery is a computational knowledge management company specialising in industry-relevant knowledge management technologies and practices. With over 30 years experience running knowledge management, informatics and data mining activities in major pharma, we are well positioned to help companies develop strategies in coping with the "data revolution" underway in the field. Due to the phenomenal increases in data generation and online resources, many scientists fear they are not making best use of the resources beyond their walls. connecteddiscovery's role is to work with scientists to understand current business challenges, and then identify the best informatics solutions to address these. Most often, the output will be a report detailing critical data sources and recommending specific workflows or software systems to address the specific problem. In addition, connecteddiscovery has licences to a number of innovative new knowledge management technologies which we can design, implement, and administer knowledge management systems, and train staff to gain maximum benefit from an investment in this business-critical area.

Lee Harland, Founder and CTO

Lee received his B.Sc. (Biochemistry) from the University Of Manchester, UK and Ph.D (Epigenetics & Gene Therapy) from the University Of London, UK. Lee has over 13 years of experience leading informatics within major Pharma. His experience spans data management, integration & warehousing, vocabulary & ontology, text-mining, competitor intelligence, knowledge and information management, data mining, bio- & chemo-informatics and software development.

Bryn Williams-Jones, Founder and COO

Bryn received his B.Sc. (Biochemistry) from Liverpool John Moores University, UK. Bryn has 18 years of experience working in drug discovery research and computational science in global pharma. Bryn has extensive experience of many facets of the drug discovery process from leading exploratory projects in multiple different disease areas, to bioinformatics and knowledge management.

http://www.connecteddiscovery.com