



Call topic: INFRADEV-3-2015: Individual implementation and operation of ESFRI projects

Title of Proposal: ELIXIR-EXCELERATE: fast-track ELIXIR implementation and drive early user exploitation across the life sciences.

List of participants

Participant Nº	Legal Name of Participant Institution	Country
1	EUROPEAN MOLECULAR BIOLOGY LABORATORY	DE
2	THE CHANCELLOR, MASTERS AND SCHOLARS OF THE UNIVERSITY OF OXFORD	UK
3	THE GENOME ANALYSIS CENTRE	UK
4	THE UNIVERSITY OF MANCHESTER	UK
5	TARTU ULIKOOLO	EE
6	STICHTING NETHERLANDS BIOINFORMATICS CENTRE	NL
7	FUNDACION CENTRO NACIONAL DE INVESTIGACIONES ONCOLOGICAS CARLOS III	ES
8	FUNDACIO CENTRE DE REGULACIO GENOMICA	ES
9	FUNDACION DE LA COMUNIDAD VALENCIANA CENTRO DE INVESTIGACION PRINCIPEFELIPE	ES
10	FUNDACIO INSTITUT DE RECERCA BIOMEDICA (IRB BARCELONA)	ES
11	UNIVERSIDAD DE MALAGA	ES
12	BARCELONA SUPERCOMPUTING CENTER - CENTRO NACIONAL DE SUPERCOMPUTACION	ES
13	AGENCIA ESTATAL CONSEJO SUPERIOR DE INVESTIGACIONES CIENTIFICAS	ES
14	UNIVERSIDAD POMPEU FABRA	ES
15	FUNDACIO PRIVADA PARC CIENTIFIC DE BARCELONA	ES
16	FUNDACIO INSTITUT MAR D INVESTIGACIONS MEDQUES IMIM	ES
17	FUNDAÇÃO CALOUSTE GULBENKIAN	PT
18	INESC ID - INSTITUTO DE ENGENHARIA DE SISTEMAS E COMPUTADORES, INVESTIGACAO E DESENVOLVIMENTO EM LISBOA	PT
19	INSTITUTO DE BIOLOGIA EXPERIMENTAL E TECNOLOGICA	PT
20	CENTRO DE CIENCIAS DO MAR DO ALGARVE	PT
21	CSC-TIETEEN TIETOTEKNIIKAN KESKUS OY	FI
22	UNIVERSITETET I BERGEN	NO
23	NORGES TEKNISK-NATURVITENSKAPELIGE UNIVERSITET NTNU	NO
24	UNIVERSITETET I TROMSOE	NO
25	UNIVERSITETET I OSLO	NO
26	SIB SWISS INSTITUTE OF BIOINFORMATICS	CH
27	CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE	FR
28	INSTITUT NATIONAL DE LA RECHERCHE AGRONOMIQUE	FR
29	CENTRE DE COOPERATION INTERNATIONAL EN RECHERCHE	FR

	AGRONOMIQUE POUR LE DEVELOPPEMENT	
30	INSTITUT PASTEUR	FR
31	CONSIGLIO NAZIONALE DELLE RICERCHE	IT
32	LINKOPINGS UNIVERSITET	SE
33	UNIVERZA V LJUBLJANI	SI
34	NACIONALNI INSTITUT ZA BIOLOGIJO	SI
35	USTAV ORGANICKE CHEMIE A BIOCHEMIE, AV CR, V.V.I.	CZ
36	MASARYKOVA UNIVERZITA	CZ
37	CESNET ZAJMOVE SDRUZENI PRAVNICKYCH OSOB	CZ
38	VLAAMS INSTITUUT VOOR BIOTECHNOLOGIE	BE
39	DANMARKS TEKNISKE UNIVERSITET	DK
40	BIOMEDICAL SCIENCES RESEARCH CENTER ALEXANDER FLEMING	EL
41	THE HEBREW UNIVERSITY OF JERUSALEM	IL

Table of Contents

1	Excellence	4
1.1	Objectives.....	7
1.2	Relation to the work programme	10
1.3	Concept and approach	15
1.3.1	Concept.....	15
1.3.2	Related Projects and Initiatives.....	17
1.3.3	Approach and methodology.....	19
1.4	Ambition.....	23
1.4.1	Current state of the art in life-science data infrastructure.....	23
1.4.2	How ELIXIR moves beyond state of the art	25
2	Impact.....	29
2.1	Addressing the expected impact set out in the work programme	29
2.1.1	Contribute to the realisation of the Innovation Union flagship Commitment n. 5: "to complete or launch the construction of 60% of the ESFRI projects by 2015"	29
2.1.2	Strengthen the ERA position and role in the global research environment; reinforce the partnership between the Commission, Member States, Associated Countries and relevant stakeholders in establishing pan-European research infrastructures.....	30
2.1.3	Foster capacity building and Research Infrastructure human capital development in targeted/relevant regions	30
2.1.4	Enhance the role of the Union in international organisations and multilateral forums	31
2.1.5	Support progress towards the development of global research infrastructures.....	31
2.1.6	Enable researchers to address societal challenges with a global dimension such as climate change	31
2.1.7	Raise the technological level of the European industry and SME's, thus improving their competitive position, through their involvement in research infrastructures development and service provision.	32
2.2	Measures to maximise impact.....	33
2.2.1	Dissemination and exploitation of results	33
2.2.2	Measuring impact nationally, in Europe and globally.....	33
2.2.3	A consolidated data management strategy for European life science.....	33
2.2.4	Knowledge management and innovation strategy.....	35
2.2.5	Communication activities	35
3	Implementation.....	36
3.1	Work plan — Work Packages, deliverables and milestones	36
3.1.1	Work Package descriptions (Tables 3.1a).....	39
3.1.2	List of Work Packages (Table 3.1b)	76
3.1.3	List of deliverables (Table 3.1c)	77
3.2	Management structure and procedures	80
3.2.1	Overview of the ELIXIR Governance framework relevant for the ELIXIR-EXCELERATE project	80
3.2.2	ELIXIR-EXCELERATE Governance	81
3.2.3	Advisory bodies	81
3.2.4	Operational management of project delivery	82
3.2.5	Project meetings	84
3.2.6	List of milestones (Table 3.2a)	85
3.2.7	Critical risks for implementation (Table 3.2b)	89
3.2.8	Minimising the carbon footprint	91
3.3	Consortium as a whole	91
3.3.1	Gender aspects and promotion of equal opportunities	91
3.4	Resources to be committed.....	92
3.4.1	Summary of staff effort (Table 3.4a)	94
3.4.2	'Other direct cost' items (Table 3.4b)	97

1 Excellence

ELIXIR - the European life-science Infrastructure for Biological Information - is a unique and unprecedented organisation that consolidates Europe's national centres, services, and core bioinformatics resources into a single, coordinated infrastructure.

ELIXIR brings together Europe's major life-science data archives and, for the first time, connects these with national bioinformatics infrastructures throughout ELIXIR's member states. By coordinating local, national and international resources the ELIXIR infrastructure will meet the data-related needs of Europe's 500,000 life-scientists¹. ELIXIR supports users addressing the Grand Challenges in diverse domains ranging from marine research via plants and agriculture to health research and medical sciences.

ELIXIR-EXCELERATE will fast-track the implementation of key scientific and organisational aspects of ELIXIR and facilitate the integration of Europe's bioinformatics resources (Figure 1): it will deliver 'excellence' to ELIXIR's users by 'accelerating' implementation of what is considered by ESFRI and European Council one of Europe's three priority Research Infrastructures.



Figure 1: ELIXIR: Coordinating Hub and participating ELIXIR national centres (Nodes)

The grand challenges facing humankind involve living creatures; finding solutions to the challenges rests on innovative basic and applied research in the life sciences. Our environment, from carbon dioxide levels in the oceans through to the expansion of deserts is fundamentally intertwined with the processes of life. Societal changes, and changes in climate, water and land-use have a direct impact on this web of life. For example, the growing population of the world - and its demand of land, water and energy resources - require constant innovation in crops and agriculture to secure a stable food-supply in a changing climate. Offshore energy extraction – also carbon neutral resources such as wind-farms – alter habitats and necessitate new approaches to environmental monitoring. Rising living standards and transformed agricultural practice has led to improvements in both health and nutrition but change the use of land and water. Reduction in global child mortality meant that for the first time in 2005-2010 more than half of all deaths were attributable to old age. Managing the rising burden of cancer, chronic and age-related diseases is now not just exclusive to healthcare systems in high-income countries but also a significant problem facing low- and middle-income countries. To maintain our hard-won improvements in all facets of human health, including standard surgical practice, we need to find solutions to combat a multitude of

¹ EC: Enabling Science, 2013: <http://ec.europa.eu/research/infrastructures/pdf/enabling-science.pdf>

obstacles – such as multi-resistant bacteria and spreading viral infections. Europe needs a highly innovative trajectory to meet the future needs of our society.

Biological research is being transformed from a laborious and costly data-gathering discipline to a highly collaborative science driven by systematic and (relatively) inexpensive data acquisition followed by complex analysis. Similar to other data-intensive sciences such as high energy physics, astronomy and oceanography, large datasets drive discoveries and form a bedrock of information on which life-scientists plan, execute and understand future investigations. The remarkable diversity addressed in life science encompasses 7 billion people worldwide, over 8 million eukaryotic species and at least 10 million bacterial species. Among eukaryotes, individuals themselves can be a complex assemblage of cells, tissues and commensal organisms. This complexity and diversity is altered continuously through the process of evolution, making data management a daunting undertaking. Living organisms respond and interact with their environment, often through mechanisms that are only partly understood. This observational and experimental complexity makes metadata and provenance acquisition complex but critical. It also means that life science arguably provides the most complex and heterogeneous datasets that science can currently imagine.



Figure 2: Self-reported high-throughput data generation capacity for genome sequencing²

Life science needs a new approach. The onset of high-throughput sequencing technologies (Figure 2) has created a deluge of data³. Most life-science data archives double every 9-12 months with some disciplines growing even faster, for example proteomics databases currently double in size every 4-5 months. With high-content biology and, in particular, sequence-based biological assays becoming routine at every major bio-research centre, and accessible by most of Europe's life-science researchers, we need to connect data management, standards, and services between all stakeholders - from local research institutes through to global core reference data archives. Data-driven analysis and research relies on a large and growing number of reference data resources and biological knowledge-bases that serve all life-science disciplines⁴ and provide focused resources that are small but critically important for a single community. ELIXIR's Preparatory Phase report found there are around 1,800 bioinformatics resources in Europe alone⁵ (Figure 3).

² Next Generation Genomics: World Map of High-throughput Sequences: www.omicsmaps.com

³ Marx V. The Big Challenges of Big Data Nature News. [doi:10.1038/498255a](https://doi.org/10.1038/498255a)

⁴ e.g. Uniprot (www.uniprot.org), Ensembl (www.ensembl.org), PDB (www.pdb.org)

⁵ Database Provider Survey report for ELIXIR, 2009:

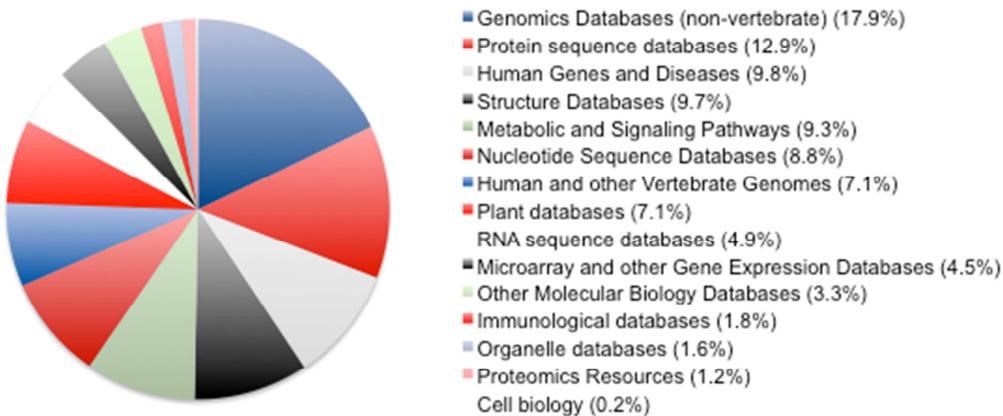


Figure 3: Diversity of molecular biology databases⁶

Data needs to be Findable, Accessible, Interoperable and Re-usable (FAIR) to generate value for a research community beyond the initial researcher's laboratory. The importance of long-term stewardship is highlighted by the observation that the odds of retrieving the data from a scientific publication decline by 17% per year⁷. Life science data infrastructure must be able to cope with the aggregation, annotation and functional integration of data from thousands of laboratories across Europe, as well as the access demands of users worldwide (e.g. the Human Protein Atlas received more than 750,000 visits during 2013). These challenges are too great for any one institute or country to address alone. It requires a coordinated response from researchers in academia and industry, national infrastructures and our global data resources, such as the Protein Data Bank (PDB), which has safeguarded the high-resolution structures of proteins, nucleic acids and complex assemblies since 1974. It also requires alignment and engagement of policy makers, and research funders at both national and EU levels. ELIXIR represents Europe's collective response to this challenge, and ELIXIR-EXCELERATE provides the vehicle to deliver an immediate response.

Use and re-use in academic and industrial research. ELIXIR's services – data resources, tools infrastructure, standards, compute and training – are used across the life sciences⁸; not just by bioinformaticians and computational biologists but also geneticists, biochemists, clinical specialists, and plant, environmental and marine scientists. In addition, industry is a major user of ELIXIR resources; users originate from pharmaceutical and biotech industries through to crop science and aquaculture, and encompass both multinational corporations and SMEs alike. The Danish and Dutch ELIXIR Nodes are formal public-private partnerships with direct industry involvement and many other Nodes have tight and long-standing industry collaborations. As these industries are major employers in Europe, these collaborations greatly contribute to the generation of wealth and support the transformation to a knowledge-based economy.

Defragmenting Europe. A coordinated approach to data management in life sciences is a priority for Europe. Many national and European life-science research programmes, as well as public private partnerships (PPPs) such as the Innovative Medicines Initiative (IMI), make significant investments in data and knowledge management infrastructure. Nevertheless, fragmentation and overlapping investments in data management are not inevitable. By implementing standards and, importantly, a Europe-wide framework of experts and supporting structures, ELIXIR will use ELIXIR-EXCELERATE to drive the coordination efforts at both national and international levels.

Formally established as a legal entity in January 2014, ELIXIR has been identified by ESFRI and the European Council as one of three priority infrastructures for Europe. The *ELIXIR 2014-18 Scientific Programme*⁹ sets out the strategic priorities for ELIXIR and the *ELIXIR 2014-18 Financial Plan* provides core funding by Member States to begin coordinated operations and identify suitable technical approaches via the implementation of pilot actions.

http://www.elixir-europe.org/system/files/D2.1%20WP2Provider_Survey_Report_0.pdf

⁶ Fernández-Suárez, XM *et al.* The 2014 Nucleic Acids Research Database Issue and an updated NAR online Molecular Biology Database Collection. [doi: 10.1093/nar/gkt1282](https://doi.org/10.1093/nar/gkt1282).

⁷ Gibney E, Van Noorden R. Scientists losing data at a rapid rate. Nature News. [doi:10.1038/nature.2013.14416](https://doi.org/10.1038/nature.2013.14416).

⁸ ELIXIR Preparatory Phase User report: <http://www.elixir-europe.org/documents/user-communities-report>

⁹ ELIXIR Programme 2014-2018 : <http://www.elixir-europe.org/about/elixir-programme-2014-2018>

The ELIXIR-EXCELERATE project will increase the velocity of this integration by comprehensive technical and scientific service harmonisation, coordinated approaches to build capacity and human capital in the ELIXIR Nodes, and addressing critical bottlenecks in operations and management of a large and distributed research infrastructure.

ELIXIR is an Open Infrastructure: it does not “own” all data resources in Europe. Through ELIXIR-EXCELERATE we will set up and provide a coordinated ELIXIR Interoperability Backbone that allows partners (e.g. other ESFRI Research Infrastructures, national resources, institutional archives) to make use of existing resources and connect and interoperate their own resources. Providing a sustainable infrastructure that manages data identifiers, secures data archiving and access, and ensures mappings between resources will enable long-term, cost-effective, data management and drive “standards as the default” across the life sciences.

1.1 Objectives

The aim of **ELIXIR-EXCELERATE** is to fast-track the implementation of ELIXIR and to drive early exploitation by user communities across the life sciences. The project will support the implementation of the *ELIXIR Scientific Programme* and the delivery of ELIXIR’s main mission:

To build a sustainable European infrastructure for biological information, supporting life-science research and its translation to society, the bio-industries, environment and medicine.

Supported by its European member states, ELIXIR will functionally integrate the key European bioinformatics and data resources and services into an effective, sustainable infrastructure. ELIXIR will drive the efficiency gains necessary to scale the infrastructure in line with the rising data volumes by applying uniform metrics and performance indicators. While there is broad agreement on the value of open access to publicly funded research data, a concerted and coordinated effort is necessary to make sure the infrastructure scales with the challenge and systematically captures the use, impact and value in the research communities. Only this will support long-term sustainability.

ELIXIR-EXCELERATE will drive the development, maturation and integration of national ELIXIR Nodes (Figure 4). By distributing resources across Europe and coordinating the efforts in ELIXIR’s member states, where most of the research investments are made, ELIXIR can scale and sustain resources. By coordinating national and international efforts ELIXIR leverages the individual strengths of Europe’s bioinformatics resources while providing the flexibility necessary to meet national priorities and the future demands of researchers.

Ultimately, the outcomes of ELIXIR-EXCELERATE build on the significant success and impact of the European bioinformatics community to ensure that life scientists, whether from academia or industry, have user-friendly, continued and reliable access to core data resources and associated tools and services that support long-term stewardship, access, integration and re-use of advanced research data in the life sciences.

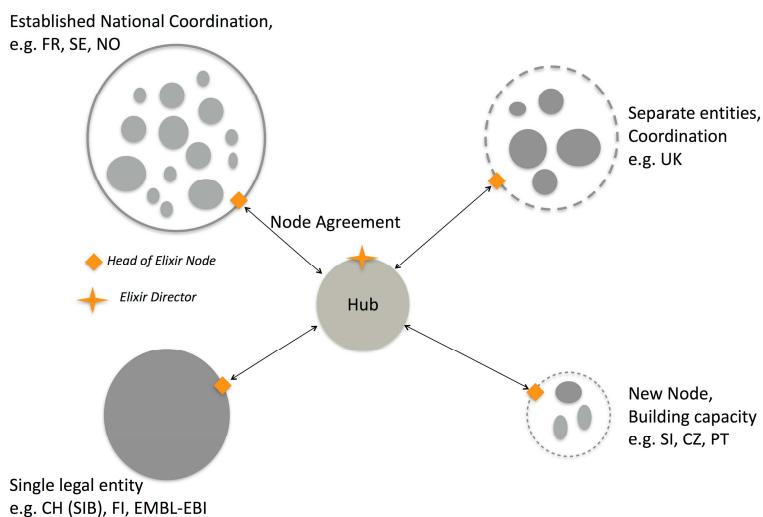


Figure 4: Organisational models of established and emerging ELIXIR nodes

The goals of ELIXIR-EXCELERATE are to be by 2018:

- 1. Deliver world-leading data services for academia and industry**
- 2. Increase bioinformatics capacity and competence across Europe**
- 3. Complete the management and organisational processes for an efficient distributed infrastructure**

The Goals will be achieved by implementing a series of ‘SMART’ Objectives, which map directly against ELIXIR-EXCELERATE Work Packages (WPs). The tasks, milestones and deliverables of these Work Packages will allow ELIXIR partners and external reviewers to effectively assess performance towards achieving these Objectives.

The Objectives of ELIXIR-EXCELERATE also complement the Strategic Objectives of the *ELIXIR Scientific Programme*. This provides the most effective way of ensuring that ELIXIR-EXCELERATE funding has the desired effect of meeting the INFRADEV-3 call: to support the implementation and initial operation of ELIXIR.

Goal 1: Deliver world-leading data services for academia and industry

The purpose of a state-of-the-art data infrastructure is to deliver reliable and sustainable high-impact services to end users. This goal is at the centre of ELIXIR-EXCELERATE. By establishing a single, integrated European approach to data management in the life sciences ELIXIR will provide a coordinated response to global efforts, ensure a strong position and role for European resources and further progress development of global standards, registries and data-exchange mechanisms.

Infrastructure must be deployed in the context of the scientific questions and realities. To ensure this ELIXIR has identified four representative and high-impact exemplar Use Cases (WP6 to 9) that bring together representative communities across our Nodes. Our Use Cases are selected to deliver infrastructure impact across the life sciences and support basic and translational research towards the grand challenges of food security, a sustainable bio-economy, and human health research.

The expected outcome is a fully functional integration of the ELIXIR services – local, national and European - into a coordinated research infrastructure, with impact being clearly demonstrated in the Use Cases. This Goal will be met by achieving the following Objectives:

- Establish an ELIXIR discovery portal that provides a transparent route to tools and services for data access and exploitation by users (WP1 and 2)
- Implement the coordination of data archives and literature sources within ELIXIR (WP3)
- Establish the technical infrastructure across ELIXIR to enable effective data deposition, access, exchange and compute (WP4)
- Establish the ELIXIR Interoperability Backbone encompassing standards, and services that implement the standards, for data archiving, integration and reuse (WP5)
- Demonstrate impact of public research data in applied academic and industrial biotechnology research by establishing standardised metagenomics services in the marine and aquaculture sectors (WP6)
- Establish a distributed genotype–phenotype annotation that supports agriculture research and industrial development (WP7)
- Demonstrate, in partnership with the Rare Disease community, how aligned ELIXIR resources enable research, avoid fragmentation and support the development of sustainability models for resources created by the community research projects (WP8)
- Coordinate data archives, technical services and support with legal/ethical requirements to provide secure, streamlined, and comprehensive solutions for long-term archiving, sharing and re-use of human research data (WP9)
- Stimulate innovation by supporting industry uptake of ELIXIR resources, particularly in SMEs (WP1 to 9, WP13)

Outcome Statement

Completion of these Objectives will ensure that ELIXIR is recognised as a world-leading data infrastructure for the life sciences, with its services becoming indispensable to users in academia and industry. ELIXIR's resources, standards and technical infrastructure will be trusted by the community and recognised globally. ELIXIR will become the *de facto* European initiative for structured collaboration with major initiatives globally, and play a fundamental role in supporting data management practice in national, EU and IMI research projects and other ESFRI Research Infrastructures and e-Infrastructures.

Goal 2: Increase bioinformatics capacity and competence across Europe

Europe has many globally competitive bioinformatics resources that are built and funded through existing national efforts. Recognising the value of these resources is at the heart of ELIXIR's strategy. These national infrastructures form the ELIXIR Nodes - the building block of the ELIXIR infrastructure. The shape and legal form of these national infrastructures is variable and reflect local requirements; however they share the fundamental ambition of organising the national communities into a single entity with an agreed lead institute. The focus of ELIXIR-EXCELERATE is to align the capacity of each individual Node and, through joint training programmes, strengthen professional skills, with the final aim to safeguard the overall quality and integrity of the data generated and stored within Europe.

This will project strong infrastructure expertise towards our users from across all ELIXIR member states. This Goal will be achieved by meeting the following Objectives:

- **Implement a programme of organisational capacity building in newly formed ELIXIR Nodes, including sharing of best practice between partners in accessing EU Structural Funds (ESIF) for operating infrastructure (WP10)**
- **Construct and coordinate ELIXIR-wide ‘communities of practice’ that support and develop the professionals who deliver advanced data and bioinformatics support and services in ELIXIR Nodes (WP10)**
- **Deliver the coordination framework and trainer capabilities (train-the-trainer) needed for the pan-European ELIXIR training programme, and use this training infrastructure to build critical skills within data and service provision (WP11)**

Outcome statement

Completion of these Objectives will ensure that ELIXIR Nodes are more effective and collaborate more closely with each other, and that operators and users are better trained and more aware of resources. Our advanced user experts will be supported by ‘communities of practice’ that provide peer-to-peer exchange and a framework to influence standards and services.

The desired outcome will be the release of skills from the European bioinformatics community for the good of all by strengthening organisational and cutting-edge research infrastructure capacity in the network of all ELIXIR Nodes.

Goal 3: Complete the management and organisational processes for an efficient distributed infrastructure

To truly be a world-class infrastructure for users, the efficiency and effectiveness of ELIXIR as a distributed infrastructure requires clarity of operational roles and processes whilst preserving the agility necessary to respond to the needs of fast-changing scientific disciplines. ELIXIR-EXCELERATE will complete the coordination and communication functions that establish ELIXIR as a world-leading, distributed infrastructure. This Goal will be met by achieving the following Objectives:

- **Conclude the establishment of ELIXIR’s operational processes and ELIXIR Hub-Node agreements so ELIXIR can operate as a single, efficient, recognised and reliable entity (WP12)**
- **Comprehensively address the ESFRI and High-level AEG recommendations (WP12)**

- Deliver the long-term strategy for sustaining Europe’s core data resources, data standards, data-legislation (WP12)
- Grow ELIXIR’s membership, within Europe and outside, and establish long-term partnerships with global science initiatives (WP13)
- Develop and implement an ELIXIR-wide Communications strategy that will engage users and disseminate project outcomes to user communities (WP13)

Outcome Statement

The expected outcome is the completion of the governance and organisational model for a truly distributed infrastructure. ELIXIR will partner as a single entity in future life science research projects and consortia. The internal processes required to deliver effective and agile operations will be completed and users will benefit from greater visibility and usage of ELIXIR services.

1.2 Relation to the work programme

ELIXIR-EXCELERATE represents ELIXIR’s submission to the Call ‘INTRADEV-3-2015: Individual implementation and operation of ESFRI projects’. Following an extensive and rigorous prioritisation process carried out by the European Strategy Forum on Research Infrastructures (ESFRI)¹⁰, ELIXIR has been recognised by ESFRI and the Council of the European Union¹¹ as one of the top three strategic priority research infrastructures for Europe. ELIXIR is considered to be pushing the boundaries of scientific excellence, holding strategic relevance for European research, and ready for immediate action. Following this decision, ELIXIR has been invited to submit an application to INTRADEV-3-2015 for targeted implementation and initial operational support. The proposal has been constructed to ensure that ELIXIR accelerates towards its implementation and sustained operations phase and delivers excellence for its scientific users. The activities that will be carried out within ELIXIR-EXCELERATE complements agreed core ELIXIR actions, funded by ELIXIR member states, and are within scope of the aims of the Call text and strongly support delivery of the Horizon 2020 Research Infrastructures Work Programme.

All ELIXIR Members (Czech Republic, Denmark, Estonia, Finland, Israel, Netherlands, Norway, Portugal, Sweden, Switzerland, UK and EMBL-EBI) are involved in this single, coordinated application. In addition, ELIXIR-EXCELERATE includes participation from a further six ELIXIR Observer countries (Belgium, France, Greece, Italy, Slovenia, Spain) which are close to ratifying the ELIXIR Consortium Agreement and are expected to become full ELIXIR Member States before, or during, the grant period.

The ELIXIR Node proposals from both ELIXIR Members and Observers have been reviewed by our international Scientific Advisory Board (SAB) and only infrastructure services that form part of a reviewed Node proposal were considered for inclusion in this grant. ELIXIR Nodes represent the national infrastructures and are all lead by a senior bioinformatics PI (Table 1) appointed by the national funding organisations or research ministries. Thus, the ELIXIR Heads-of-Nodes committee, our senior management team, and the PI group for this grant are of exceptional quality, underlining the transformational potential of this effort.

¹⁰ Prioritisation of Support to ESFRI Projects for Implementation:

http://ec.europa.eu/research/infrastructures/pdf/ESFRI_projects_for_impl_7_april_2014.pdf#view=fit&pagemode=none

¹¹ Conclusions on the implementation of the roadmap for the European Strategy Forum on Research Infrastructures.

COMPETITIVENESS Council meeting Brussels, 26 May 2014:

http://www.consilium.europa.eu/uedocs/cms_Data/docs/pressdata/en/intm/142794.pdf

Table 1: ELIXIR Node Leadership

Node	Head of Node (HoN), Organisation and Function	Node Status
BE	Yves Van de Peer , VIB; Division Coordinator	Observer
CH	Ron Appel , SIB; Executive Director	Member
CZ	Jiří Vondrášek , UOCHB AVCR; Group Leader	Member
DK	Søren Brunak , DTU; Professor	Member
EE	Jaak Vilo , UTARTU; Institute Head	Member
EMBL-EBI	Rolf Apweiler , EBI; Joint Associate Director Ewan Birney , EBI; Joint Associate Director	Member
ES	Alfonso Valencia , CNIO; Director	Observer
FI	Tommi Nyrönen , CSC; Project Director	Member
FR	Jean-François Gibrat , IFU; Institute Head	Observer
EL	Babis Savakis , BSRCAF; President & Scientific Director	Observer
IL	Michal Linial , HUJ; Director	Member
IT	Graziano Pesole , CNR; Director	Observer
NL	Barend Mons , NBIC; Professor	Member
NO	Inge Jonassen , UiB; Professor	Member
PT	Jose Pereira Leal , FCG-IGC; PI	Member
SE	Bengt Persson , LIU; Professor	Member
SI	Brane Leskošek , UL; Assistant Director	Observer
UK	Chris Ponting , UOXF; Professor	Member

ELIXIR-EXCELERATE requests a budget of EUR 19.05 million. This is in line with the recommendations of the Call text for between EUR 15-20 million, respects the pressures placed on the overall INFRADEV-3 budget heading of EUR 97 million, and is appropriate to ensure that the activities are achievable and will create meaningful impact for ELIXIR and its users.

The proposal builds firmly on ELIXIR's EU-funded Preparatory Phase¹² findings (Table 2). The understanding of ELIXIR's user communities, the various technical feasibility studies carried out and the learning from the database provider's survey allows ELIXIR to focus, in this application, on implementing services, coordinating the infrastructure and continuing to develop the operational and governance steps required of a world-leading data infrastructure. Since the conclusion of ELIXIR's Preparatory Phase, ELIXIR has agreed a *Financial Plan* for 2014-2018 including the operational budget for the ELIXIR Hub as well as ELIXIR's *Scientific Programme* for 2014-2018. ELIXIR-EXCELERATE will be the major vehicle to help implement this ambitious scientific programme, augmented by core funding and, where applicable, additional grants.

In addition, the Work Programme specifies additional activities under Part B of the section "Specific features for Research Infrastructures", which are considered relevant for INFRADEV-3 (

¹² ELIXIR Prep Phase reports: <http://www.elixir-europe.org/members-area/documents?page=1>

Table 3).

ELIXIR-EXCELERATE will also be used to address the longer-term recommendations of the High-level Expert Group on Assessment¹³, which assessed ELIXIR in 2012-2013 (the more immediate recommendations have been addressed in the intervening period such as the role of EMBL within ELIXIR was clarified through the now approved Work Programme of EMBL). Finally, ELIXIR-EXCELERATE addresses the ESFRI recommendations of 2014. In Table 4 we describe how direct tasks in corresponding Work Packages will address these. Collectively, these steps will ensure that ELIXIR-EXCELERATE represents a high-quality application for the implementation of a critical ESFRI research infrastructure.

In section 2.1, we show in detail how ELIXIR-EXCELERATE addresses the expected impact required by the Call text for INFRADEV-3.

Table 2: ELIXIR-EXCELERATE Work Packages mapped against suggested activities of Work Programme

INFRADEV-3 scope: suggested activities	Addressed by ELIXIR-EXCELERATE Work Package
<i>Enlargement of the membership</i>	WP13
<i>Enhancement of the technical architecture</i>	WP1 to 5
<i>central coordination</i> <i>data management (including possible open access to data)</i> <i>training</i> <i>outreach</i> <i>international cooperation</i> <i>research and innovation activities</i>	WP12 WP1 to 9 WP11 WP13 WP13 WP6 to 9
<i>Interaction with end-users; notably industry and SMEs</i>	WP6 to 9, 13
<i>Activities that foster the development of Regional Partner Facilities</i>	WP10

¹³ Report of the High-level Expert Assessment group: <http://ec.europa.eu/research/infrastructures/pdf/jd-final-aegreport-23sept13.pdf>

Table 3: ELIXIR-EXCELERATE Work Packages mapped against suggested activities of Specific features for Research Infrastructures

Part B of the section “Specific features for Research Infrastructures”	Addressed by ELIXIR-EXCELERATE Work Package
Central management and coordination, including setting up and initial running of the central coordination office	WP12
Organisation of the logistic support for researchers, definition of access policies for researchers and management of IPRs and ethical issues	WP12
Integration of the new entity in the European landscape of related facilities, and in the local context	All WPs, key role of WP10. In addition, WP8 (Use Case D) demonstrate how ELIXIR integrates with EC Integrating activities and other projects in a domain
Innovative solutions for data or sample collection, management, processing, curation, annotation, and deposition, <i>including relations with publishers</i> for supporting data and sample deposition services	WP1 to 5, WP7 and WP9 provides demonstrator data management use cases aligned towards health and agriculture research
Deposition and stewardship services for data and publications generated by Horizon 2020 actions	Core role of ELIXIR: WP3 – archives WP4 – technical backend WP5 - standards WP6 to 9 demonstrator Use Cases WP10 – local capacity WP11 – training & skills WP12 – organizational and policy support, long term sustainability
Definition of standards, protocols and interoperability; benchmarking	WP2, WP4, and WP5
Access provision to research communities following the rules specified for integrating activities	Note that ELIXIR services are ‘open access’ hence direct costs rather than ‘Unit costs’ are applied. However, developing a Unit cost/resource allocation model for e.g. cloud storage is part of WP4
Integration of distributed resources into virtual facilities.	Integrated throughout, but particularly WP10 (building and connecting Node capabilities) and WP12 (Organisation)
Spreading of good practices, consultancy and training courses to new users; outreach; exchange of personnel and training of staff	Integrated throughout but WP10 to 13 establish long-term mechanisms within ELIXIR
Coordination with national or international related initiatives and support to the deployment of global and sustainable approaches in the field	Integrated throughout, but especially WP13. ELIXIR has established links with NIH BD2K initiative ¹⁴
Activities to increase the potential for innovation, including social innovation	Integrated throughout, but especially WP13 – ELIXIR Innovation and SME programme

Table 4: How ELIXIR-EXCELERATE addresses the ESFRI recommendations

¹⁴ NIH Big Data to Knowledge (BD2K): <http://bd2k.nih.gov/#sthash.JDJBNRvf.dpbs>

ESFRI Recommendation	ELIXIR-EXCELERATE Response	Work Package
Strengthen central coordination function with the goal of creating a single robust European research infrastructure	Establish an effective project management unit and conclude the implementation of ELIXIR coordination, governance and operational structure	WP12, Task 1 WP10 WP13
Clarify the role of EMBL as a partner and contractor and consider the long term sufficiency of the present legal framework as EMBL special project;	Establish high-level working group that, based on independent evaluation by external consultant, assesses the performance and suitability of ELIXIR's implementation as part of EMBL	WP12, Task2
Fully develop the different service streams of the infrastructure across the Nodes for data resources, compute provision, tools infrastructure, training and standards infrastructure	Establish an implementation Work Package for data resources, compute provision, tools infrastructure, training and standards infrastructure that addresses Node coordination and identified bottlenecks from infrastructure gap analysis and SAB recommendations	WP1, WP2 ('Tools') WP3 ('Data') WP4 ('Compute') WP5 ('Standards') WP11 ('Training') guided by WP6 to 9 ('Use Cases')
Fully develop data access procedures and associated data protection, as well as ethical, legal and societal aspects	Integrate services across ELIXIR through a user-driven project (WP9) that secures effective implementation of processes with 4 different medical research cohorts	Our user-driven project (WP9) builds on data resources (WP3), technical services (WP4) and established ELSI partnerships with BBMRI (WP12) and Global Alliance (WP13)
Develop common procurement and recruitment procedures in order to maximize the impact and decrease costs, therefore achieving significant impact in production and access to services	We focus on the opportunity for commercial cloud offerings as an option for biological computations and address issues such as coordinated access terms, provision of biological reference data and SME access	WP4, WP12
Establish a fully dedicated program for industry engagement including the pharmaceutical and agriculture sectors, supporting industry and SME innovation through service provisioning, joint projects and public-private partnerships and a 'pan-European SME and innovation programme'	Strengthen Node partnerships with Industry and SME by sharing and further develop best practice from established Node PPP Establish ELIXIR SME and Innovation Programme across national Nodes	WP10 (Capacity Building) WP13 (SME programme)
Take steps to expand the ELIXIR's membership	Our response is three-fold: 1. Support the involvement of new countries by reaching out within ERA members 2. Establish practical guidance and support for developing national ELIXIR Nodes 3. Develop scientific collaborations with countries outside of Europe and foster global initiatives	WP13 (Communications and Outreach) WP10 (Node Capacity building) WP13 (Develop internationalisation strategy)

1.3 Concept and approach

1.3.1 Concept

The core concept of ELIXIR is to provide coordinated capacity through a distributed organisation – to establish services and expertise close to our large and distributed user-base. ELIXIR-EXCELERATE will accelerate the implementation of this by addressing specific implementation and coordination bottlenecks in organisation and coordination; in human capital and Node capacity; and in harmonisation, coordination and delivery of research infrastructure services.

The specific actions and tasks in this proposal were identified through extensive consultations during the ELIXIR Preparatory Phase and the development of ELIXIR's *Scientific programme*. In addition, ELIXIR will address the long-term recommendations of the EC High-level Expert group on Assessment for Research Infrastructures and ESFRI recommendations of 2014. Resolving these bottlenecks through ELIXIR-EXCELERATE will contribute to providing world-class services to ELIXIR's users.

ELIXIR-EXCELERATE is organised around three project goals, each supported by Work Packages (Section 1.3.3) that deliver specific, quantifiable outcomes:

Deliver world-leading research infrastructure services (WP1 to 9, 12M€) The overall outcome of the efforts towards this objective is fully functional integration of the ELIXIR services – local, national and European – into a coordinated research infrastructure, with impact clearly being demonstrated in the Use Cases.

Increase bioinformatics capacity and competence across Europe (WP10 and 11, 4M€) Strengthening national bioinformatics capacity, human capital and enable industry usage across Europe. The overall outcome will be the provision of strong infrastructure expertise to our users – in academia and industry – across all member states.

Complete the management and organisational processes for an efficient distributed infrastructure (WP12 and 13, 3M€) Ensure effective, agile and purposeful operations of ELIXIR and establish the coordination framework both within Europe and with sister initiatives internationally. This work will also ensure that the recommendations of the ESFRI prioritisation report will be implemented through all stakeholders.

To support integration and capacity building the Work Packages will be resourced and led by ELIXIR Nodes throughout ELIXIR's member states. Overall, the project is led by the ELIXIR Hub; the ELIXIR secretariat at the Hub will deliver the project management functions needed for a large infrastructure consortium.

ELIXIR-EXCELERATE is an infrastructure implementation project: project governance will be tightly linked to the established ELIXIR governance model. Therefore, no new governance bodies will be created through this project. For example: Scientific progress will be reviewed by ELIXIR's existing SAB and linked to the evaluation of the overall infrastructure. The Heads-of-Nodes remain the senior scientific management team. The team of Technical Coordinators already established between the current Nodes will govern technical implementations. The established Industry Advisory Committee (IAC) will guide collaboration between academia and industry within the ELIXIR context. The ELIXIR Board, with representatives from Member and Observer States, remains the ultimate decision making body for ELIXIR.

Coordinated capacity in a distributed organisation: Despite significant investments in life sciences in Europe, existing biological data centres lack a centrally coordinated strategy that results in efficient integration of, and synergy between, different countries and programmes. The outcome is that beyond the core resources, which are mainly based at EMBL-EBI, the European bioinformatics landscape has thus far remained largely fragmented across borders and institutions.

The ELIXIR Preparatory phase began to address this fragmentation. Inspired by long standing national efforts such as the SIB Swiss Institute of Bioinformatics (founded in 1998), the FP7-funded Preparatory Phase of ELIXIR catalysed the formation of national infrastructures in many European countries. ELIXIR Nodes represent the national bioinformatics infrastructure, funded by national research organisations, and support life science data management and exploitation nationally. By coordinating national and international efforts, ELIXIR also leverages the individual strengths of members and has flexibility to respond to national priorities and future demands.

The maturity of the ELIXIR Nodes is variable: some of the national infrastructures and networks that form ELIXIR Nodes have been operating for many years, while other Nodes are currently being formed by receiving initial support through national research infrastructure roadmaps. Thus, a key concept in this project is to provide support to the newly formed Nodes by spreading good practice through a focused Node capability building effort:

- The ELIXIR-EXCELERATE project will drive community building and development of organisational capabilities required to operate and further develop ELIXIR Nodes. Through this project ELIXIR will also establish ELIXIR-wide communities that support the ELIXIR Node professionals that deliver advanced user support through means such as data management practice and genome annotation. In addition, it will support Nodes in developing Regional Partner Facilities, enabling them to better access ESIF for research infrastructure operation.
- The concept of the ELIXIR Innovation and SME programme has been developed based on extensive industry discussions and surveys (including discussions with local bio-regions and accelerator organisations) and tested through a first workshop in 2014. Aligning this programme with developing Node capacity represents a step-change in bioinformatics industry outreach and addresses the major bottleneck in creating sustainable impact by aligning industry and national capacity and fostering strong local links.

Through ELIXIR-EXCELERATE we deepen and enhance the coordinated scientific and technical infrastructure of ELIXIR. The ELIXIR preparatory phase project, through wide engagement and analysis across European life sciences, delivered the basic architecture for an effective, sustainable data management infrastructure. To comprehensively address the challenges the infrastructure needs to cover: Data, Tools, Standards, Compute and Training (Figure 5).

The *ELIXIR Scientific Programme 2014-18* sets out the strategic objectives for each of these areas. ELIXIR-EXCELERATE builds on this and will, together with national investments in ELIXIR Nodes, ELIXIR core funding, and potentially other grants, achieve the strategic objectives.

Through ELIXIR-EXCELERATE we will deepen and enhance the coordination aspects of this '*Data, Tools, Standards, Compute and Training infrastructure*' to ensure maximal value of the member state investments:

- ELIXIR aims to provide an infrastructure that supports the community in managing access and discoverability of **Tools** (through the registry), encourage robustness and sustainability (through metrics recommendations and services) and support *community-led* efforts to benchmark and enhance scientific quality. In this project we will enhance federation and support incorporation of local tools registries (WP1) and establish the Community benchmarking infrastructure (WP2).
- Europe has many established **Data** resources that are brought together in ELIXIR. Through this project ELIXIR will harmonise the collection of technical and impact metrics across the infrastructure by establishing uniform and robust metrics services and address the critical bottleneck of data-literature integration by demonstrating connection of a semantically enriched Europe PubMed Central (Europe PMC) with our core resources (WP3).
- The **Standards** infrastructure in ELIXIR requires a comprehensive set of scalable services to support local, national and Europe-wide data management. Many of these services exist, developed through national projects and international collaborations, but there is a need for coordination of ELIXIR-wide implementation. Through ELIXIR-EXCELERATE we will bring together the key services into the ELIXIR Interoperability Backbone that, together with our core resources, will provide scalable and sustainable services for data interoperability and national data management (WP3).
- **Compute** services, together with effective data access and exchange, need to scale for large and distributed datasets. Collaborating closely with e-Infrastructures (GÉANT, EUDAT and EGI among others), ELIXIR will ensure effective access to compute services, co-located with our data. User access is also a key component and to provide optimum user support it is necessary to track who is using the ELIXIR platform. With the access control layer we will be able to manage, resource and steer the services in the long-term. While most of the ELIXIR services provide virtual access, a specific user account with login will be a requirement for some, for example cost tracking purposes (compute cloud cluster accessing a core data resource). Through ELIXIR-EXCELERATE we will establish a *research platform for life science* implementing geographically and

organisationally distributed Cloud, Compute, Storage and Authentication and Access infrastructure services collected in the ELIXIR registries (WP4).

- ELIXIR Nodes have established national **Training** programmes. In ELIXIR-EXCELERATE we will align training with national capacity building and establish the coordinating infrastructure that will allow initiatives to join up and effectively exchange training materials, courses and events throughout the infrastructure. We also address specific skills shortages within the infrastructure ('train the developer / trainers') and provide the training needed to achieve full impact in the project's Use Cases (WP10 and 11).

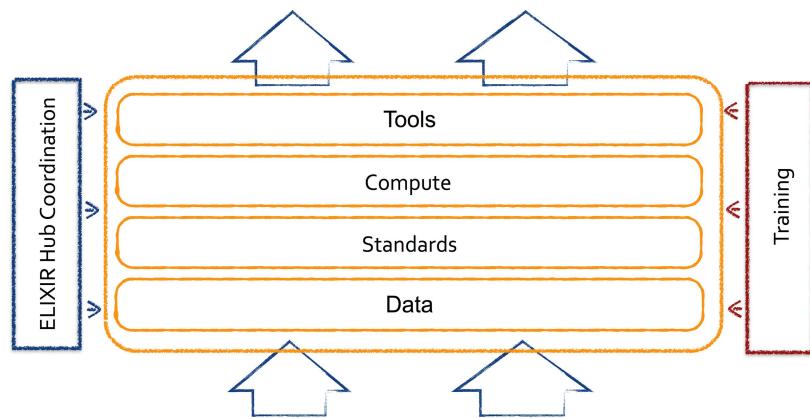


Figure 5: A bioinformatics infrastructure for European needs. Services for Data, Standards, Compute and Tools powered by skilled people that operate and make use of the services (Training), and supported by an efficient and agile coordinating organisation (ELIXIR Hub Coordination).

ELIXIR-EXCELERATE embeds ELIXIR within the ERA's life-science communities. The project includes four driving 'Use Cases' (WP6 to 9) that capture user needs from key communities and guide the coordination and enhancement of ELIXIR's key resources. These user-driven exemplar projects coordinate the development of complex and secure data-sharing, integration and re-use through end-user engagement (5). Our 'Use Cases' (WP6 to 9) build on work from emerging communities within ELIXIR- with the technical approach validated through ELIXIR Pilot Actions¹⁵- and in addition to addressing important infrastructure challenges for individual communities will also serve as exemplars that deliver solutions that will be scaled to other communities. Two Use Cases address the 'data in' challenges facing ELIXIR whereas the other two typify community requirements for effective data re-use. All Use Cases will illustrate how ELIXIR serves as a platform for collaborating and interfacing with community research projects, other ESFRI research infrastructures and e.g. EC 'integrating activity' projects for regional and national infrastructures.

Science is a global effort and life science is no exception, indeed most of the flagship life-science projects are truly global endeavours: the wheat genome consortium has over 1000 partners in 57 countries, and the International Cancer Genome Consortium includes 74 projects in 15 countries¹⁶. International collaboration is of paramount importance for effective exchange of data within these projects and includes not only technical standards for data exchange but also agreements on compatible legal and ethical frameworks for exchange on human data. ELIXIR is considered by the G20 countries to be of global significance and with great potential for membership by countries outside of Europe. Building on the established links with partners outside of ELIXIR (such as other ESFRI RIs, IMI projects, and the NIH BD2K initiative in the United States) we ensure that ELIXIR solutions are well recognised and integrated globally. Through ELIXIR, and this project, we bring together these resources to ensure that Europe speaks with one voice on issues such as data standards, data-legislation and exchange mechanisms.

1.3.2 Related Projects and Initiatives

The landscape of life-science data resources is fragmented into many initiatives and resources. An important role for ELIXIR is to provide a long-term vision and architecture to allow the different projects, stakeholders and related infrastructures to integrate and make use of our coordinated resources and interoperability backbone.

¹⁵ ELIXIR pilot projects: <http://www.elixir-europe.org/about/pilot-projects>

¹⁶ The International Wheat Genome Sequencing Consortium: <http://www.wheatgenome.org/>

ELIXIR has made extensive use of pilot actions during the initial phase of implementation. These short (3-6 Months) technical ‘sprints’ inform and validate the architectural decisions within ELIXIR. This approach has been exceptionally successful: the first generation pilot actions are now all implemented as production services with e.g. the single-sign-on services and now form the basis of secure user management within ELIXIR, across BMS RI (see below), and also have been implemented in Social Science archives¹⁷. Similarly the pilot actions involving the EGA and integration of marine resources give us strong confidence in user delivery and impact of WP6 to 9 Use Cases.

ELIXIR, as the data infrastructure, has an important role in connecting all the ESFRI Biological and Medical Research infrastructures (BMS RI). ELIXIR coordinates the EU-funded *BioMedBridges*¹⁸ project (ending in Dec 2015), which is building the fundamental data bridges and standards required for infrastructure interoperability. Through this project the BMS RI has mapped out their data management practices¹⁹, delivered a proof-of-concept tools registry with an extensive corpus of annotated tools, developed programmatic interfaces, and generated metadata standards catalogues. These deliverables, through the strong connection with data experts and users in other infrastructures, guide and embed the ELIXIR services throughout the community, and will be further extended and sustained in this project (e.g. ELIXIR Discovery Portal (WP1), Interoperability backbone including identifier management, WP5).

Beyond 2015 the eleven BMS RIs hope to further deepen their integration through *CORBEL*, a proposal for a new “cluster project” (through INFRADEV-4) that would build on *BioMedBridges*. Currently still under evaluation, *CORBEL* would develop four shared collaboration platforms for the BMS infrastructures: Integrated user access, Data management, ELSI Services and Innovation support services. The *CORBEL* project is based on the notion of a shared user pipeline: advanced biomedical research (e.g. biomarker discovery) needs to make extensive use of services across infrastructures (imaging, biobanks, molecular structures) and through the *CORBEL* project the BMS RI will develop effective interfaces and the needed joint services for this pipeline. The *CORBEL* Project is coordinated by ELIXIR and co-chaired by BBMRI.

Three aspects of BMS RI interoperability and *CORBEL* are of particular importance to this project:

- ELIXIR will, through *CORBEL* and through bilateral agreement with BBMRI, make extensive use of the joint ELSI platform; BBMRI has established a Europe wide network of ELSI expertise, ELIXIR will not duplicate this and will develop the internal processes needed to effectively partner and make use of BBMRI ELSI Common Services (WP12).
- All BMS RI need strong internal data management to deliver effective user services, ELIXIR will not own, nor be responsible for, the internal needs of other infrastructures. Hence, the *CORBEL* data management Work Package will continue to identify and map the identifiers and ontologies needed for integration and interoperability of BMS RI data resources. This will interface, and depend on, the core ELIXIR interoperability layer (WP5, the ELIXIR interoperability implementation).
- CORBEL also addresses the critical issue of secure access to sensitive human research data by mapping secure access practices and streamlining applications across infrastructures; this project will ensure that ELIXIR resources have a coordinated approach to data security, meets good practice and provides scalable services for our sister infrastructures (WP4).

National and regional infrastructures for many life-science user communities are currently integrated through integrating activities in FP7 or H2020 projects (e.g. Integrating Activities in marine sciences and aquaculture, plant sciences and rare diseases and Virtual Research Environments covering particular life science domains). ELIXIR has a key role in supporting the continued integration and development of national practice. This is addressed in WP13 through a task that will develop a transparent, long-term mechanism for integration of Community projects. WP8 provides a Use Case template for this in the field of rare disease (interfacing with *RD Connect*, *Innorare*, etc.).

¹⁷ Guideline at ESFRI level: <http://www.clarin.eu/content/federated-identity>; implementation example in Social Sciences Archive in Finland: <https://services.fsd.uta.fi/?lang=en>

¹⁸ BioMedBridges: Building data bridges from biology to medicine in Europe: <http://www.biomedbridges.eu/>

¹⁹ Principles of data management and sharing at European Research Infrastructures, 2014: <http://dx.doi.org/10.5281/zenodo.8304>

The ELIXIR resources and services have to build, strengthen and forge additional international alliances in order to harmonise efforts, share resources, avoid duplication/redundancy, and to maximise effectiveness. Globally, the specific organisations and initiative of relevance include NIH and NSF funded Bioinformatics resources in North America (for example, model organism databases and the Big Data to Knowledge centres), the Global Alliance, relevant resource and service providers in the Asia-Pacific region, as well as associations in other parts of the world. Many of the ELIXIR Core resources (Section 3.3) have long standing agreements with sister resources globally (e.g. through INSDC²⁰, IMEX²¹). The EGA²² archive has established an exchange mechanism with dbGaP²³ for human data consented for research; EMBL (Paul Flicek) leads the work on data security standards and implementation within the Global Alliance for Genomics and Health²⁴. Several partners also have roles within emerging NIH BD2K centres (e.g. BioCADDIE) and further integration is planned (e.g. through joint workshops on data identifiers).

ELIXIR also has strong links and interdependencies with the e-Infrastructures, for instance through current joint pilot actions that evaluate the EUDAT services. ELIXIR recently endorsed the GÉANT Data Protection Code of Conduct for managing and exchanging user identities in compliance with European data protection regulations. Our Compute, Data access and Data security (WP4) will establish strong links to the e-Infrastructures; the leading partners within ELIXIR (CSC, MU) also have strong and long standing roles within the European e-Infrastructures.

1.3.3 Approach and methodology

The organisation of this project is based on the extensive analysis of the European life-science data landscape during the ELIXIR preparatory phase²⁵. In addition, during 2014 ELIXIR ran a series of work-streams with participation of over 100 scientists across ELIXIR Nodes for the development of the *ELIXIR Programme*²⁶. This was complemented by an external analysis of the user needs of SMEs and industry Through a process of evaluation by ELIXIR's technical task-forces (ELIXIR Technical Coordinators and training experts) and final prioritisation by ELIXIR Heads of Nodes (assisted by SAB evaluation), the critical bottlenecks and high value activities have been selected for accelerated implementation through this project. The feasibility and proof of concept of many of these activities has been demonstrated through ELIXIR Pilot actions, giving us confidence in our approach.

To reinforce the coordination and capacity building aspects of this project, ELIXIR has identified co-chairs from the ELIXIR Nodes to lead each of the implementation Work Packages.

ELIXIR-EXCELERATE project implementation has been designed around three goals (Figure 6):

Goal 1: Deliver World-leading Research Infrastructure Services

Goal 2: Increase bioinformatics capacity and competence across Europe

Goal 3: Complete the management and organisational processes for an efficient distributed infrastructure

²⁰ International Nucleotide Sequence Database Collaboration: <http://www.insdc.org/>

²¹ IMEx: The International Molecular Exchange Consortium: <http://www.imexconsortium.org/>

²² EMBL-EBI European Genome-phenome Archive: <https://www.ebi.ac.uk/ega/home>

²³ dbGaP: The database of Genotypes and Phenotypes: <http://www.ncbi.nlm.nih.gov/gap>

²⁴ Global Alliance for Genomics & Health: <http://genomicsandhealth.org/>

²⁵ The ELIXIR Strategy for Data Resource: http://www.elixir-europe.org/system/files/elixir_strategy_for_data_resources_report.pdf

²⁶ ELIXIR Programme 2014-2018: http://www.elixir-europe.org/documents/elixir_scientific_programme

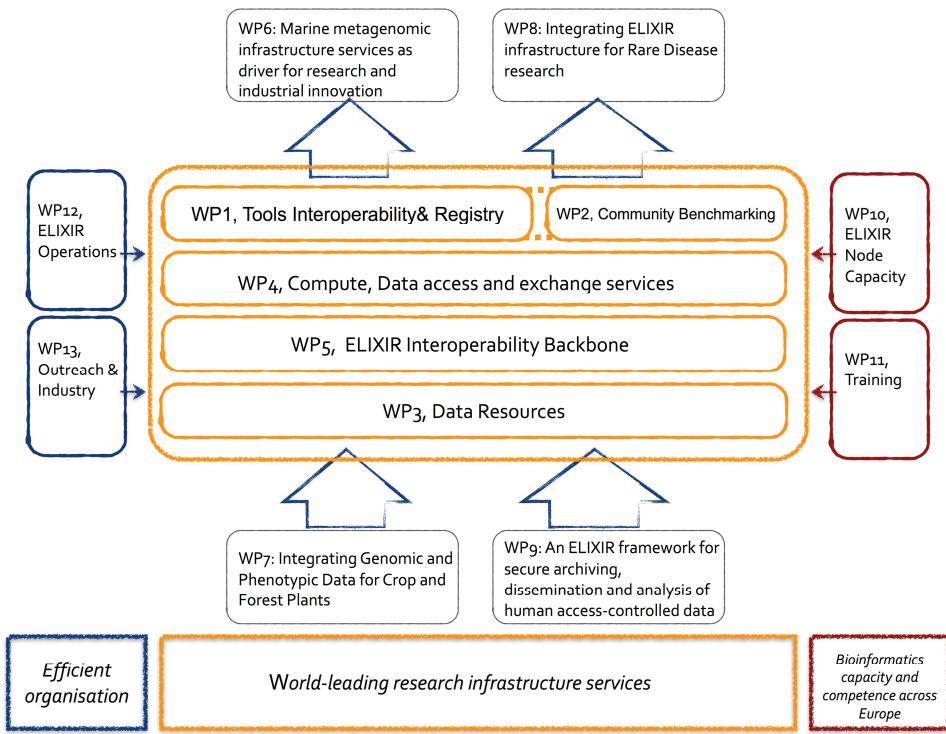


Figure 6: Organisation of ELIXIR-EXCELERATE Work Packages around the three project Goals. Our ‘Use Case’ Work Packages (WP6 to 9) have been selected to represent data management and archiving (‘data stewardship’) as well as data exploitation and integration for research communities addressing the grand challenges of health, food security and a sustainable bio-economy.

Goal 1: Deliver World-leading Research Infrastructure Services:

WP1, Tools Interoperability & Registry (Søren Brunak, DK and Alfonso Valencia, ES): This WP will deliver a discovery portal for bioinformatics resources world-wide, built upon federated curation of a wide range of key resources. It will represent service monitoring, resource integration, and interoperability aspects of services.

WP2, Infrastructure for Community-led Benchmarking (Alfonso Valencia, ES and Søren Brunak, DK): This WP will develop and provide an infrastructure that supports community centred benchmarking efforts. By bringing together Data resources, standards and tools (including community tools) ELIXIR can offer a stable environment that supports community benchmarking efforts of tools and algorithms. Community led benchmarking exercises are critical to understand algorithm performance and trade-offs and guide tool usage, in addition they produce high value reference datasets for the research community. These are high-impact outcomes that ELIXIR will support by developing a stable infrastructure.

WP3, Data Resources and Services (Christine Durinx, CH and Jo McEntyre, EBI): Implementing uniform metrics and benchmarking standards is fundamental to the integration of Europe’s data archives and knowledge bases. This WP will deliver a joint framework that inform and drive the sustainable development of Europe’s core life-science data resources based on agreed, uniform KPI. It will also close the current gap of literature-data integration and cross-linking by supporting curated knowledge-bases through Europe PubMed Central, the literature archive.

WP4, Compute, Data access and exchange services (Tommi Nyrönen, FI and Ludek Matyska, CZ): This WP will link the *ELIXIR Scientific Programme* for 2014-2018 to the day-to-day technical service work in the distributed Nodes. It will create a sustainable and supported *research platform for life science* for implementing geographically and organisationally distributed Cloud, Compute, Storage and Authentication and Access infrastructure services collected in the ELIXIR service registry. The WP will manage external technical dependencies with e-Infrastructures and Nodes with ELIXIR Technical Coordinator group for services delivered as a priority for the ELIXIR-EXCELERATE Use Cases (WP6 to 9).

WP5, ELIXIR Interoperability Backbone (Barend Mons, NL and Carole Goble, UK): The WP will focus on standardization, data publishing and data connectivity - that is “functional interlinking” the resources offered in

ELIXIR. The WP will drive the accelerated implementation of interoperability services necessary for the WP6 to 9 Use Cases and those prioritized in WP3 by (i) provisioning a Backbone of services and resources, and (ii) consolidating and sharing sustainable standards and practices, which contributes to capacity (WP10).

WP6 to 9, Integrated ELIXIR delivery exemplars: These WPs will align ELIXIR services from the technical implementation WPs to ensure embedding into the work of user communities through a portfolio of 4 “vertical” integrated ELIXIR delivery projects. These projects serve as implementation exemplars and partner with on-going scientific research projects in different areas of life science. They demonstrate how fully integrated ELIXIR services support users and promote scientific excellence. The application exemplars demonstrate broad applicability across life sciences (Biotechnology, Agriculture and Plants, Biomedical and Health) but also represent different stages of the data life-cycle (Secure Data deposition (WP9), Distributed data annotation (WP7), and advanced data re-use (WP6 and 8). Hence these users -driven exemplar projects (“Use Cases”) will tie together ELIXIR services (and project Work Packages) into effective workflows that partner with key user communities. The Use Cases will also help to define and test training curricula and infrastructure capacity building.

WP6 Use Case A: Marine metagenomic infrastructure services as driver for research and industrial innovation (Nils P Willassen, UiT and Rob Finn, EBI): The main objective for this Use Case is to develop a sustainable metagenomics infrastructure to enhance research and industrial innovation within the marine domain. Metagenomics methodologies need to overcome a number of challenges related to standardization, development of relevant databases and bioinformatics tools. This Use Case will develop research infrastructure and service provision specific for the marine domain in order to enable metagenomic approaches responding to societal and industrial needs. The outcome of the proposed Use Case will meet the major needs expresses by the marine domain (e.g. ESF Marine board Position Paper 17 “*Marine Microbial Diversity and its role in Ecosystem Functioning and Environmental Change*” and Position Paper 15 “*Marine Biotechnology: A New Vision and Strategy for Europe*”) and supports on-going user projects in academia and industry.

WP7 Use Case B: Integrating Genomic and Phenotypic Data for Crop and Forest Plants (Paul Kersey, EBI and José Leal, PT): Massive sequencing and genotyping of crop and forest plants (and their pathogens and pests) is generating large quantities of genomic variation data. These efforts are likely to accelerate in the near future, with further expected reductions in the cost of sequencing and international efforts (such as the DivSeek Initiative²⁷⁾ aiming to catalogue all genetic diversity present in global germplasm resources. However, structural variation in most crop plants is enormous (more so than in humans), and phenotypic characterisation data is (i) often inaccessible (ii) diverse and non-standard (iii) lacks any route of unified access. Both technical and sociological progress in data definition and sharing are lagging experimental progress. To address this problem, this Use Case will harness the domain-specific expertise and data held in a distributed fashion across many national Nodes with interests in agriculture. Seven ELIXIR Nodes will jointly establish a technical infrastructure and associated social practices to define an open model for the publication and sharing of plant genotype-phenotype data.

WP8 Use Case C: Integrating ELIXIR infrastructure for Rare Disease research (Ivo Gut, ES and Marco Roos, NL): This Use Case will address the data integration needs of the rare diseases community through the coordination of the ELIXIR resources described in WP1-5. The aim is to interface and empower on-going (e.g. RD Connect) and future RD research projects by addressing data interoperability and management bottlenecks and support sustainability of the resources created by these projects in the long-term.

WP9 Use Case D: An ELIXIR framework for secure archiving, dissemination and analysis of human access-controlled data (Justin Paschall, EBI and Arcadi Navarro, ES): This Use Case will deliver a workflow (Figure 7) that support data submitters and ELIXIR Node coordination efforts on data deposition to the EGA archive, enable data release to authorized individual users from the archive, and to partner downstream secure ELIXIR data analysis platforms.. Importantly, data ownership and access is maintained in the hands of the original resource owner (who has acquired consent from study participants). This workflow will also allow resource owners to focus on their unique areas of data generation and analysis expertise while being able to rely on EGA and the ELIXIR infrastructure for their common big -omics data storage, coordination and distribution needs under appropriate legal and data security frameworks.

²⁷ DivSeek: <http://www.divseek.org>

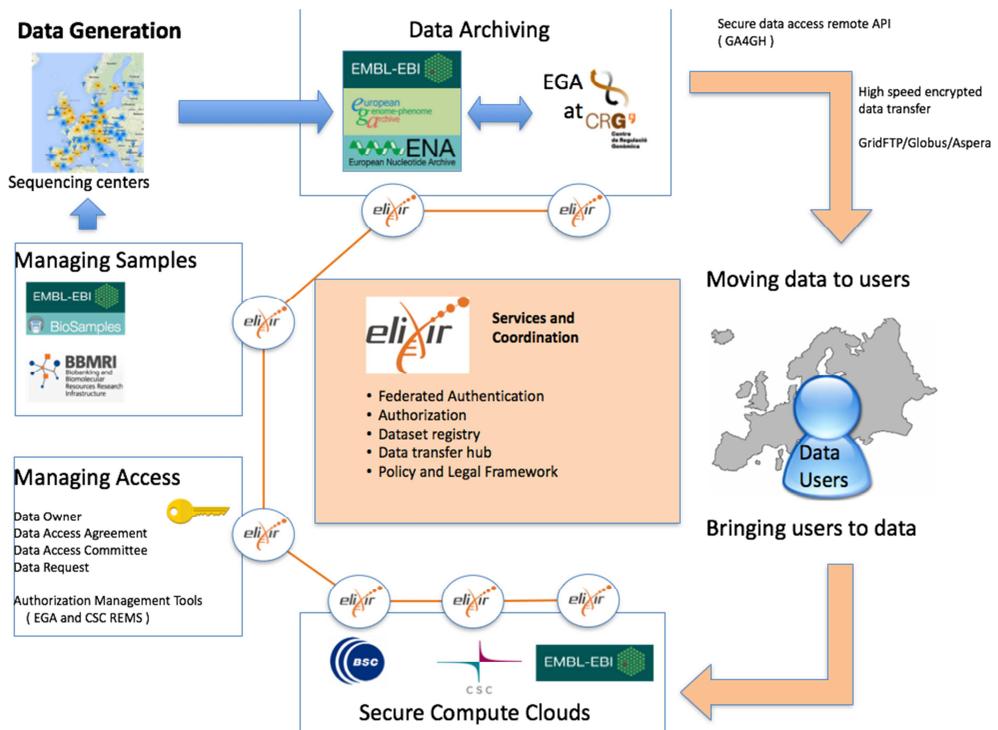


Figure 7: Schematic to illustrate the workflow pipeline for archiving, disseminating and analysing human data (WP9 Use Case D)

Goal 2: Increase bioinformatics capacity and competence across Europe

WP10, ELIXIR Node Capacity Building and Communities of Practice (Jiří Vondrášek, CZ, and Bengt Persson, SE): Effective and scalable data management practices in Europe requires the connection of institutional, national and international efforts²⁸. This Work Package includes activities aiming at strengthening the ELIXIR infrastructure by supporting coordinated activities and increasing the capacities of individuals and organisations in the Nodes. The WP will build organisational and community building capacity, develop a systematic approach for developing practical synergies between relevant ESIF operational programmes and smart specialisation strategies (RIS3) but importantly also develop networks to support our experts delivering advanced user support through formation of ELIXIR wide communities of practice.

WP11, Accelerating the ELIXIR Training Programme (Chris Ponting, UK and Patricia Palagi, CH): The life science community is estimated at 500,000 people, drawn from across the spectrum of individuals ranging from researchers and bioinformaticians, to tool developers and infrastructure operators. This WP has two principal objectives: (1) To build a sustainable training infrastructure for ELIXIR's community - encompassing both a technical infrastructure and training expertise, as well as mechanisms for guaranteeing quality of training. (2) To develop and deliver training in topics selected as training gaps within the ELIXIR community in selected application areas (namely WP6 to 9 Use Cases).

Goal 3: Complete the management and organisational processes for an efficient distributed infrastructure

WP12, Excellence in ELIXIR Management and Operations (Niklas Blomberg, ELIXIR Hub): This WP will further develop ELIXIR Management and Operational processes that are necessary for deep and effective integration of ELIXIR Nodes. In addition, this WP will establish a world-class project management unit, consolidate, and further develop ELIXIR operational processes, including gender equality and diversity management. This includes risk management as well as sets of tasks that specifically address the ESFRI and AEG recommendations on the research infrastructure operations.

²⁸ The Royal Society, 2012. Science as an open enterprise:

https://royalsociety.org/~media/royal_society_content/policy/projects/sape/2012-06-20-saoe.pdf

WP13, Communications, Industry and Community Engagement (Andrew Smith, ELIXIR Hub): This WP will ensure that ELIXIR reaches out effectively to users across different communities. This will be achieved by carrying out activities that address three complementary objectives: Develop and implement an ELIXIR-wide Communications Strategy, support innovation through development and delivery of an SME focused innovation support programme, and develop a coordinated International Strategy for ELIXIR to reach out and collaborate globally on life-science research data management.

Managing dependencies and synergies:

The activities in the Work Packages are tied together through management tools (e.g. meetings, work-shops and aligned deliverables, Section 3) and, via the scientific delivery of the Use Cases, establish expert communities and the permanent ELIXIR Working Groups (Technical, Data and Training coordinators). For instance:

- In WP3 a list of core resources has been defined that are prioritized for optimal interoperability in WP5. These will be preferentially addressed in close collaboration with the authoritative data owners in the respective ELIXIR Nodes. WP5 will form the collaborative effort between the Nodes involved in the interoperability efforts of these prioritized data sources, including active and personal assistance at request for data source owners in the ELIXIR network.
- In WP6 three ELIXIR Nodes aim to establish interoperable data resources for the marine microbial domain. In WP7 seven ELIXIR Nodes aim at interoperable plant genotype-phenotype data, using established repositories for genomic data and sample metadata. In WP8, eight ELIXIR Nodes aim to share interoperable data across patient registries, biobanks, and -omics databases. WP10 will establish a transformative, ELIXIR-wide ‘Data Nodes Network’ for good data stewardship: practices for submission to archives and practices for optimal interoperability. This will tie together the national experts required to establish and manage standards for describing and accessing datasets, reporting data, matching and comparing data content, and building linkages between datasets, making extensive use of the infrastructure components delivered by WP1-5.

1.4 Ambition

1.4.1 Current state of the art in life-science data infrastructure

For the past 30 years biological and medical research has progressively transformed into a data-intensive science. This transformation is accelerating as technologies for data generation are progressively miniaturised, automated, and ‘industrialised’. The current estimate is that Biological research produces as much data as earth sciences or astronomy, and is rapidly approaching the levels of high-energy physics. This is reflected in rapid growth of our data archives: every month sees the addition of 30TB of secure human data consented for research use; the European proteomics database doubles every 4 months. Cancer research is expected to sequence over 50,000 genomes over the next five years and high throughput, data intensive, assays are also of critical importance to agricultural research and environmental sampling. Reference datasets are reaching a size where ‘download-and-compare’ is no longer viable, not even for regional or national data-centres: the European data-centre at EMBL-EBI currently manages over 50 petabytes (PB) of storage. Biological data is also fragmented, with large data-sets produced at thousands of sites across Europe, often widely differing in formats, annotations and the standards used. A further challenge is that interpretation often requires extensive integration of very large datasets. In addition to the massive data storage and exchange requirements from these efforts, they will also necessitate effective access to reference data and tools of critical importance to individual researchers and health-care professionals.

The current organization of Bioinformatics has been able to cope with the organization and analysis of data due to its distributed structure and the open accessibility to data and analysis methods. The rich tradition of well-curated data resources has resulted in services of wide applicability and usage. For example, the UniProt resource of manually aggregated functional data (The UniProt Consortium, 2014) has over 800,000 daily requests²⁹. At the national level, the ELIXIR Nodes have to meet increased support needs for management and publication of research data. It is therefore essential that ELIXIR, through national Nodes as well as coordinated European investments, develops coherent strategies to sustainable life-science data management, archiving and services. This requires a coordinated response from researchers, research funders and science policy bodies. A key issue in this context is transparency of the different cost drivers for research data management and archiving.

²⁹ Source: UniProt web logs

Projects are increasingly faced with the challenge of integrating multiple, complex large-scale datasets and compare the results with rapidly growing reference datasets. These datasets are often acquired over large number of samples, from different sites, organisations and countries. Despite the life-science data surge, the major challenge is not the volume of data but the complexity: Biological data is heterogeneous, information dense, and has complex assumptions and error models that depend on experimental details. Biological research across life-science domains increasingly also requires the integration of multiple *-omics* technologies with imaging and biochemical data that are increasingly sampled through high-throughput metabolomics techniques. This creates an underlying driver for diverse data formats, annotations and metadata standards. Consequently, the aggregated resource needs required for deposition and integration of such data into the overall data landscape remains a significant challenge not only to dedicated bioinformatics infrastructures, but also to European life-science researchers as a whole. It remains a risk that large numbers of stakeholders and programmes within Europe drive fragmentation of data management and technology platforms. For the investments to be effective the infrastructures need to work closely with user communities and research funding networks. In the biomedical domain there are also investments in large public private partnerships, such as the IMI, that bring public research institutions and academic researchers together with industrial R&D organisations. ELIXIR should have an important role as a bridge between the infrastructures and initiatives through coordination and consolidation of fundamental data services.

The fragmentation of biological data with many divergent formats creates a significant barrier to data integration and reuse. Data interoperability has been highlighted for many years as a key bottleneck for value creation³⁰. In the life sciences there are many examples where community standards drive reuse and allow new scientific discoveries. A good example is the series of “minimal information” standards that articulate metadata requirements for data depositions. In the functional genomics field, where this was pioneered through the “MIAME” standard, there have been systematic evaluations of data reuse: it is estimated that on average one of every three deposited datasets contributes to new publications³¹. Data reuse underpins and validates additional experimental studies as well as provides a foundation for further³², value added resources. With a ‘state of the union’ of over 1800 dispersed and frequently non-interoperable data sources³³, ranging from high-end knowledge-bases and archives (e.g. UniProt, HPA, ENA) to custom data-bases that deliver valuable services to a single community functional data integration and analysis has become a formidable challenge. Only well-equipped and solidly funded larger groups and companies may still be able to handle this exploding situation with local ‘download-and-integrate capacity’. Research across data resources is hampered by the lack of standards and the poor adoption of existing standards by stakeholders. Data Interoperability overcomes the barriers of syntactic access with semantic use in one implementation. Optimal Interoperability is only attained when access and use can be completely automated: programming and interfaces conform to standards that specify consistent syntax and formats; and data are associated with metadata and terminology identifiers and codes that support computational aggregation and comparison of information that resides in separate resources.

The Open data culture in the life sciences has created an infrastructure of services and skills. In large, international genome consortia, it is well-established good practice to have a designated data coordination centre that generally involves one or several of the public data repositories. As the scale of data generation has grown so has the need for well curated metadata and computationally accessible knowledge-bases, dictionaries and ontologies. Findability, Interoperability, Accessibility and Re-usability of open managed access research data requires robust annotation of conditions and experimental context and easily navigational reference data, as well as professional licensing, authorization and authentication services across Europe and globally. It is therefore essential to guarantee the future sustainability of our capacity to optimally analyse, share, functionally integrate, sustain and re-use these extremely valuable data resources. Datasets in current life sciences range from those that hold raw data (Experimental Repositories) through to those that hold fully annotated and curated data (Biological Knowledge bases):

Experimental Repositories hold experimental reports, deposited in public archives or repositories (examples: EUDAT, ENA, PRIDE, PDBe, Metabolight, HPA, BioSamples). These sources typically

³⁰ Searls, DB. ‘Data Integration: Challenges for Drug Discovery’, *Nature Reviews Drug Discovery*, 4 (2005), 45–58.

³¹ Piwowar H *et al.* ‘Data Archiving Is a Good Investment’, *Nature*, 473 (2011), 285–285.

³² Rung J, and Brazma, A. ‘Reuse of Public Genome-Wide Gene Expression Data’, *Nature Reviews Genetics*, 14 (2012), 89–99.

³³ Database Provider Survey report for ELIXIR, 2009:

http://www.elixir-europe.org/system/files/D2.1%20WP2Provider_Survey_Report_0.pdf

provide a series of time date stamped and stable records. They are the bedrock of the knowledge bases and the archives of Biology.

Biological knowledge bases hold curated and annotated biological entities and their relationships representing current biomedical knowledge and analytical processes over other datasets (examples: UniProt, Ensembl, Reactome, ChEMBL, Orphanet). The knowledge is dynamic; the records are updated, may diverge or merge over time, and may change considerably in their interpretation as new knowledge is acquired. For example, a novel gene-rare disease relationship is reported or the consequences of a single nucleotide change in a regulatory genomic region are better understood.

ELIXIR infrastructure must recognise the fact that biological knowledge is incomplete. Data incompatibilities are often the products of incomplete datasets (example: draft genomes, or incomplete coverage of experiments), compounded by complex biological scenarios (example: structural genomic variation, or protein structure unit cell components), plus the practicalities of release cycles and coordination. Links between concepts and entities are complex (example: protein sequences in PDB are not strictly the same as UniProt and they are not strictly the same as Ensembl). Therefore, ELIXIR must be able to handle and report metadata in this rapidly evolving field, and make explicit the variations and scope of data and identifiers: for this Linked Data's descriptive flexibility is ideal. Integrating complex datasets will require a set of services to handle identifiers for data and biological concepts (phenotypes, diseases). It will also require tools that allow users to map data between different sources and make sure that users can find and apply the right standards. Many of these tools exist. ELIXIR-EXCELERATE will bring these services together.

Industry usage of bioinformatics resources is high but unevenly distributed. In a requirement-mapping analysis of bioinformatics need in industry companies performed in 2013³⁴ it was noted that the industry is very dependent on SME and BioIT companies as a source of innovation, yet academic outreach to these is highly variable. Industry representatives have also noted concern that the current infrastructure is very fragmented, and there is a lot of inertia to overcome to enable real interoperability of data. 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.

1.4.2 How ELIXIR moves beyond state of the art

Europe is home to some of the world's leading bioinformatics institutes and resources. ELIXIR represents the collective efforts of its members to coordinate and sustain these vital resources, ensuring a level of interoperability that will transform scientific discovery.

Coordinating local, national and European resources will transform life-science data management. ELIXIR will ensure that users – who may be individual scientists, large consortia or indeed other research infrastructures – can easily access data resources that are sustainable, built on strong community standards, and safeguarded in the long-term. This is transformative and unique: never before has such a concerted effort, backed by long-term commitments from member states, been made to connect national infrastructures that reach out to local and regional centres with Europe-wide reference data resources and supporting services for data-standards. This is amplified by a significant investment to increase bioinformatics capacity and competence across Europe and backed by a sustainable and effective organisational framework. Researchers need to explore large and heterogeneous data from multiple sources. By coordinating Europe's national and international capabilities into a coherent infrastructure our 500,000 users will seamlessly navigate an ecosystem of life science data services.

ELIXIR reaches scientists across Europe by strengthening national Nodes. Many European countries have long-established national bioinformatics infrastructures, which provide nationally coordinated services. One of the lasting benefits achieved by the preparatory phase of ELIXIR is the formation of further national infrastructures; through this project ELIXIR will build and integrate of additional bioinformatics capacity (WP10), scientific (WP1-9) and training (WP11) expertise close to our large and distributed user base. Bioinformatics platforms in the Nodes are tightly linked to research projects within their national life science communities. For example, they

³⁴ ELIXIR: Understanding industry's needs: http://www.elixir-europe.org/industry/understanding_industrys_needs

might provide biomedical cloud platforms, compute services or participate in collaborative scientific projects in response to calls for proposals from funding agencies. Because of the distributed model, ELIXIR is linked with many user communities and is well placed to provide a platform that build on national research strengths and priorities and supports data-driven life science throughout Europe.

Create a distributed data architecture that connects European life science. The core strategy of ELIXIR is to use our distributed organisation to release the collective capabilities of our Nodes to construct a distributed infrastructure that enables full data integration of the collective, expanding capacity across the continent. This requires a robust and scalable architecture that allow coordinated collaboration between the many different partners and stakeholders. Strong collaboration between Nodes will address current data challenges that include quality curation of data (WP3), interoperable service interfaces (WP5) and long-term sustainability of data resources (WP3, WP12). Of particular importance is to ensure strong links remain at the national level between ELIXIR and the other distributed BMS RIs and ensure uniform implementation of the shared services and strategies between infrastructures (WP10, ‘Nodes Data Network’). ELIXIR do not own or control user’s data - data ownership remains with the data generator – but provides the infrastructure for long-term data stewardship. ELIXIR acts as the bridge between the infrastructures and initiatives through coordination and consolidation of the fundamental data services.

Meet the challenge of secure access to identifiable data in health research. Personal genomics and multi-omics data archives are a key resource for translational research, personalised medicine and the elucidation of the mechanisms underlying complex diseases. Nevertheless, the long term archiving and re-use of personal -omics data poses significant challenges in the management of ethics, consent for reuse, review of access, data security and governance. Specifically, ELIXIR-EXCELERATE will perform a comprehensive review of current practice for archiving and access to sensitive data together with an analysis of gaps and major risks to deliver recommendation on best practice (WP9 supported by WP4 and WP12).

ELIXIR will also collaborate closely with the Global Alliance for Genomics and Health³⁵ to support the development and implementation of the emerging global code-of-conduct for management and sharing of personal research data and develop training modules for implementation of the Global Alliance for Genomics and Health (GA4GH) security framework. Critically, ELIXIR-EXCELERATE will deliver a set of tools to allow the effective management of data in the secure archives by Data Access Committees.

Standards as the default – establish Europe wide expert networks for data management. Life science research increasingly requires the effective integration of heterogeneous datasets. These datasets span molecular, cellular and (patho-) physiological processes for downstream analysis. Data integration requires standards for fundamental biological entities (such as samples, genes, proteins and cells) and experimental components (e.g. cell-lines, strains, conditions and encoding of quantitative, structural and imaging data), and standards and well defined descriptions of complex concepts such as bioprocesses, tissues and diseases. The ELIXIR wide Communities of Practice (WP6) and permanent Data, Technical and Training Coordinator groups (WP12) have a key role in driving the uptake, use and extension of data standards through their roles in national archives and projects.

ELIXIR will identify a set of Core Resources that are globally competitive and of critical importance to the life science community and actively promote their integration and sustainability. The Core Resources will be delivered under their own brands, as services from the Nodes, and form the backbone of the ELIXIR data infrastructure. To qualify as a Core Resource a resource must be well maintained, with capacity and processes for professional service delivery, as well as plans for the life-cycle management and understanding of service. Europe provides many of the resources used globally today; three examples the European Nucleotide Archive (founded in 1984 as the EMBL Data Library), the Swiss-Prot resource in the 1980s (becoming part of UniProt) and the Ensembl Genome browser.

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

³⁵ Global Alliance for Genomics & Health: www.genomicsandhealth.org

tightly coupled to the provision of ELIXIR Core Resources; ELIXIR will close the cycle of data annotation, deposition, provision and integration for reuse.

Making the implicit explicit and sustained: ELIXIR Core Resources establish authorities for data identifiers and identifier mappings. This includes data templates, best practices and guidelines. The ELIXIR core resources have managed identifiers, identifier mappings definitions, insights, fault corrections (amongst others) in and between their resources for years and are clearly the most qualified to ‘take authority’ in their core expertise area. With ‘ELIXIR recognized authorities’ providing identifier schemes, mappings, data templates for genomics, proteomics, metabolomics and other data services, Findability, Accessibility, Interoperability and actual Re-use of data will rest on a foundation of long-term, collectively managed resources supported by an internationally coordinated training programme. The ELIXIR Core Resources will thus serve as the core of the ELIXIR Interoperability backbone.

Connecting an ELIXIR Interoperability Backbone

The implementation and use of uniform descriptors, standard terms and encoding in the generation and storage of data is truly fundamental, yet it is not enough for an effective integration infrastructure. There is also a need for a knowledge management architecture that provides services for users of these standards. The emphasis is on machine actionable metadata as well as human consumable metadata, and the instigation of metadata standards and services that ensure that data is ‘FAIR’. This requires specific interfaces, standards and integration services between research infrastructure data-types (coordinated through WP5) and comprises:

- Catalogues or registries that support “findability”, i.e. the description, location and look-up for basic entities such as compounds, samples, genes (persistent identifier resolution services).
- Services to identify the key vocabularies, classification and ontologies (ontology lookup services) that support the linking and interoperability of data.
- Repositories which implement and maintain data-ontology mapping components.
- Services that support the mapping and connection of vocabularies and ontologies to enable complex search of data in distributed data systems.

This is Use Case driven: In WP6 three ELIXIR Nodes aim at establishing interoperable data resources for the marine microbial domain. In WP7 seven ELIXIR Nodes aim at interoperable plant genotype-phenotype data, using established repositories for genomic data and sample metadata. In WP8 eight ELIXIR Nodes aim to share interoperable data across patient registries, biobanks, -omics databases. In each of these use cases work is required to establish and manage standards for describing and accessing datasets, reporting data, matching and comparing data content, and building linkages between datasets.

One deliverable of WP10 is to establish a transformative, ELIXIR wide ‘Data Nodes Network’ for good data stewardship: practices for submission to archives and practices for optimal interoperability. The ELIXIR Data Coordinators group is the glue in this network: it is tasked with coordination, development of shared strategies and identifying good practice. WP5 will underpin the collaborative effort between the Nodes involved in the interoperability efforts of these prioritized data sources by constructing the ELIXIR Interoperability Backbone, including active and personal assistance at request for data source owners in the ELIXIR network.

The ELIXIR Interoperability Backbone is user-data driven (integrations working with the datasets based on use cases defined by end users). ELIXIR will support cross-resource questions, bridging genomic and phenotypic data for variant identification, using machine processable (RDF/XML) representations of the (meta) data. For the data resource and service management of ELIXIR Core and Named Resources we emphasize self-described datasets and explicitly described and published life-cycle metadata, using machine processable representations and common APIs for accessing it. Many of the necessary services exist across ELIXIR Nodes. ELIXIR-wide coordination of this currently fragmented process will ensure an aligned and measurable impact in setting global practices, resulting in cross-resource research being possible and routine.

The ELIXIR interoperability backbone will enable partners to interface their data. Many existing community efforts address sub-challenges and provide sub-solutions related to the broader area of data interoperability. Hence *ELIXIR will not dictate a single platform but rather focus on conventions that enable data interoperability*, stewardship and compliance with data and metadata standards, policies and practices. Through the implementation of community adopted and ELIXIR-endorsed standards and, importantly, a European-wide framework of experts and a credible supporting organisation, ELIXIR will drive the coordination efforts both at national and international level. ELIXIR does not “own” or ‘control’ the data resources in Europe but provide a coordinated Backbone that

enables and assists partners (e.g. other ESFRI Research Infrastructures) to make use of existing solutions and connect and interoperate their resources. Sustained infrastructure services for e.g. identifier management, data access and mappings between resources drive “standards as the community driven default” and enable long-term data management according to the FAIR principles.

Establish supporting metrics for coordination and life-cycle management of Europe’s data resources (Figure 8). European bioinformatics resources will be both more competitive on a global scale, and more attractive as a global collaborator if they are perceived to be a coordinated and regulated set of services from the European Research Area (ERA). An agreed set of metrics and quality criteria for ELIXIR resources (WP3) and transparency on how those resources are used will reinforce this message and extend excellence in European bioinformatics service provision globally. The use of defined and agreed metrics/criteria for the management of resource life-cycles will address sustainability questions of the ELIXIR resources as a whole, allowing decision making based on objective criteria. Should resources become no longer relevant to the scientific community, they may be retired, which will in turn provide opportunities for the development of new resources that have the potential to provide greater impact in supporting scientific progress. Sharing information on how to collect and report quality criteria, and how to improve these over time, will build capacity across Nodes regarding effectiveness and excellence in resource management, strengthening the European position. The clear indication that a service is an ELIXIR-affiliated resource will build confidence and trust in a broad user base, in particular for new or occasional users such as students, the general public, and SMEs.

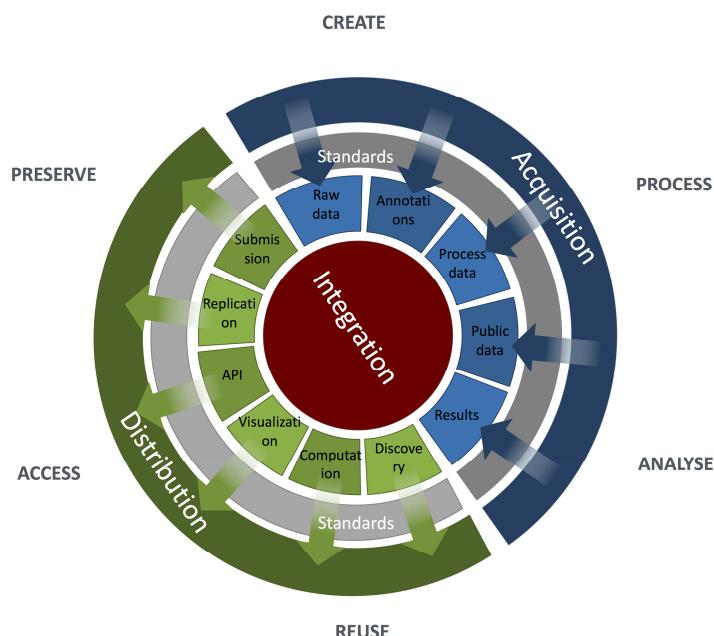


Figure 8: The Data cycle must support standards, workflows and interfaces for data acquisition (including curation and metadata) as well as those for distribution and re-use.

ELIXIR-EXCELERATE will establish effective and uniform operational processes. ELIXIR is constructed as a distributed infrastructure, built around existing centres of excellence throughout the European member states represented through national Nodes, and incorporating the internationally supported European Bioinformatics Institute (EMBL-EBI). ELIXIR is a distributed but coordinated research infrastructure; through our ELIXIR Hub-Node *Collaboration Agreements* and *Service Delivery Plans* the current European service landscape is brought into a coherent and sustained infrastructure where each Node contributes with its expertise to tackle the most urgent challenges of life-science data management. Acceleration of this integration process is the purpose and key outcome of WP12; the many national partners will be brought together into one infrastructure that interfaces future projects and partners as a single entity.

Operating a large, distributed, ‘virtual’ organisation requires strong coordination capabilities with clear roles and effective and simple processes. Implementation of this organisation and integrating existing national networks and infrastructures is a significant management challenge and the core role of the secretariat at the ELIXIR Hub. ELIXIR-EXCELERATE will drive this process through the establishment and updates of an ELIXIR-wide *ELIXIR Handbook of Operations* (WP12) that captures the overall policies and processes in an easily accessible document that together with a *Manual for ELIXIR Nodes* (WP10) serves as the basis of our organisational developments. By

establishing *Communities of Practice* for our advanced user support experts (WP10), and a dedicated training programme for bioinformatics service developers (WP11) this ensures development of our human capital and secures implementation of uniform metrics (WP3), service and tools annotations (WP1), computational services and standards infrastructure (WP4). The ‘ELIXIR All Hands’ meetings (WP12) are of particular importance as they will bring together the ELIXIR community to review progress, exchange good practice and ensure engagement and awareness of the overall ELIXIR objectives.

ELIXIR-EXCELERATE will drive deepened global partnerships and coordination for bioinformatics. The resources from ELIXIR’s national Nodes and EMBL-EBI are widely used throughout the world. The core resources and life-science data archives have long-standing global collaborations: for data infrastructures, the larger the scale of collaboration, the more datasets are available and, accordingly, the more accurate and useful are the services developed. ELIXIR’s infrastructure is of global significance and the resources it champions are widely used throughout the world. Thus, ELIXIR represents the collective European view as an infrastructure of global significance within a research community that functions thanks to agreed international standards.

2 Impact

2.1 Addressing the expected impact set out in the work programme

Despite the Grand challenges that face European society, biomedical research is a traditional strength in Europe, accounting for over 49% of the combined research output³⁶. Furthermore, with a combined public and private research investment of €60 billion annually the biomedical sector is a key component of the European knowledge economy³⁷ with significant innovation potential. Several independent recent studies demonstrate robust returns, in the order of 10%, on investment in biomedical research³⁸. The pharmaceutical industry alone directly employs over 700,000 people in the EU, generating three to four times more downstream jobs and contributing to a trade surplus of €80 billion³⁹.

In 2012, over €30 billion was invested by the pharmaceutical industry in R&D. Europe is a rich breeding ground for biotech start-ups, with 1,799 commercial healthcare biotech companies operating in Europe in 2012⁴⁰. The healthcare sector as a whole accounts for 8% of the total European workforce and for 10% of the EU’s GDP⁴¹. Against this backdrop, the economic impact of ELIXIR-EXCELERATE - although difficult to estimate quantitatively - is likely to be significant.

Investments in ELIXIR-EXCELERATE add value to the European economy and drive a range of societal benefits. For example, the databases that are part of ELIXIR act as the major repositories for the data that are generated through national and European research programmes. The Open Data Pilot in Horizon 2020 will rightly encourage more researchers to publish their data in open, accessible repositories: such is the scope and breadth of ELIXIR’s core resources, much of these Horizon 2020 data will flow into and reside within the ELIXIR infrastructure, making it available for re-use by all and driving further innovation in academia and industry.

As well as delivering real impact in meeting its objectives, ELIXIR-EXCELERATE directly addresses all of the expected impacts listed in the Call text for INFRADEV-3.

2.1.1 Contribute to the realisation of the Innovation Union flagship Commitment n. 5: "to complete or launch the construction of 60% of the ESFRI projects by 2015"

ELIXIR is considered by the ESFRI Implementation Report of 2012⁴² and the ESFRI Conclusions of May 2014 to have reached its implementation stage. ELIXIR-EXCELERATE will fast-track the implementation of ELIXIR’s crucial early operational phase. A priority will be to address all of the ESFRI recommendations and the long-term

³⁶ European Science Foundation. (2011). A Stronger Biomedical Research for a Better European Future: http://www.esf.org/fileadmin/Public_documents/Publications/emrc_wpII.pdf

³⁷ Chakma et al. NEJM, 2014. doi:10.1056/NEJMp1311068

³⁸ e.g. Glover et al. BMC Med. 2014. [doi:10.1186/1741-7015-12-99](https://doi.org/10.1186/1741-7015-12-99) and: MRC 2008. Medical Research: What’s it worth? http://www.wellcome.ac.uk/stellent/groups/corporatesite/@sitetestudioobjects/documents/web_document/wtx052110.pdf

³⁹ European Federation of Pharmaceutical Industries and Associations: <http://www.efpia.eu/facts-figures>

⁴⁰ E&Y: Beyond Borders 2014: <http://www.europabio.org/sites/default/files/facts/ey-beyond-borders-unlocking-value.pdf>

⁴¹ EC: Investing in Health 2013: http://ec.europa.eu/health/strategy/docs/swd_investing_in_health.pdf

⁴² ESFRI Implementation Report 2012:

<http://ec.europa.eu/research/infrastructures/pdf/esfriImplementationReport2012.pdf#view=fit&pageMode=none>

recommendations of the High-Level Expert Group on Assessment, thus ensuring that ELIXIR cements its position as a priority infrastructure that delivers excellence to users. We describe in Section 1.2 how ELIXIR-EXCELERATE will be used to implement these recommendations.

The implementation of ELIXIR is essential to the realisation of the full potential and functioning of the other 12 life-science infrastructures on the ESFRI Roadmap, all of which depend – in some way – on ELIXIR’s provision of access to world-class data, tools, compute, interoperability, standards, training and industry services. This is acknowledged in the role of ELIXIR as coordinator of the *BioMedBridges* cluster grant and *CORBEL*. ELIXIR-EXCELERATE will also provide many secondary benefits for the other ESFRI RIs, helping them in their own progress towards implementation. ELIXIR-EXCELERATE will provide the ELIXIR Hub and Nodes with additional capacity to share good practice with through workshops, advice on data management issues and direct collaborations and the conclusion of Memoranda of Understanding (MoU) with other ESFRI RIs.

2.1.2 Strengthen the ERA position and role in the global research environment; reinforce the partnership between the Commission, Member States, Associated Countries and relevant stakeholders in establishing pan-European research infrastructures

ELIXIR-EXCELERATE actively supports the implementation of several critical EU policies. Most directly, the project will support one of the fundamental cornerstones of the ERA, the provision of world-class infrastructure to enable science. The project does so by bringing together 19 institutes from the 11 ELIXIR member states and EMBL-EBI. In addition the project includes an additional 22 institutes from 6 ELIXIR Observer countries that will shortly join as full members (in some cases prior to the commencement of the Grant Agreement). Thus, ELIXIR represents an unprecedented European effort to establish a coordinated infrastructure for the collection, storage, annotation, validation, dissemination and utilisation of biological data. ELIXIR-EXCELERATE will add significant value to the national investments made in the infrastructure to date. At the Member State level, ELIXIR represents substantial added value. The development of pan-European standards, services and training programmes supports the integration of national capabilities and gives strong confidence in the synergy and sustainability of investments. Access to advanced life-science infrastructures is a key objective of the Smart Specialisation Strategies of many European regions⁴³. These efforts are set to directly benefit from the implementation of Europe-wide standards.

One key objective of ELIXIR-EXCELERATE will be to finalise the overall governance and coordination framework that will allow ELIXIR to participate as a single partner in future EU projects. WP10 and 12 will provide support to ensure that the Nodes and the ELIXIR Hub will establish the necessary legal agreements to allow partners to participate as one entity as soon as possible.

The shared data activities of ELIXIR-EXCELERATE also support the policy objectives of Pillar V of the Digital Agenda for Europe, where the new EC Communication on data-driven economy acknowledges that research and innovation funding on data in the EU is sub-critical and the corresponding activities are largely uncoordinated.

2.1.3 Foster capacity building and Research Infrastructure human capital development in targeted/relevant regions

For a number of countries, participation in ELIXIR-EXCELERATE 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 the project will enable organisations to improve their expertise in bioinformatics through participation in the development of best practise and access to ELIXIR’s professional Training Programme. Focusing of Node skills and capacity building (WP10) the skill and knowledge exchange between national RI’s will serve to strengthen partnerships between ELIXIR Nodes. Furthermore, establishing communities of practice (WP10) and executing training programmes (WP11) will facilitate the development of an international pan-European RI to compete on a global stage and enhance human capital development. This will ensure immense long-term benefits for ELIXIR partners, users of ELIXIR Node services and the wider ERA.

Dedicated activities within WP10 will provide targeted support to ELIXIR Nodes (Estonia, Greece, Slovenia and Czech Republic) in developing Regional Partner Facilities, aligning their efforts with regional Smart Specialisation Strategies (RIS3) and accessing EU Structural Funds (ESIF) for the operation of life science infrastructures. RIS3

⁴³ S3 Smart Specialisation Platform: <http://s3platform.jrc.ec.europa.eu>

strategies have been pushed to focus on a limited number of technology sectors. Further, regional priorities for supporting research and innovation and ESIF funding rules provide more opportunities to projects looking to construct physical infrastructures rather than operating virtual distributed resources, such as ELIXIR. Taking these issues into account, good practice in achieving practical synergies between RIS3 and ESIF within ELIXIR will be identified, assessed and selected cases pooled via the creation of a Task Force, which will disseminate this knowledge across ELIXIR-EXCELERATE partners. Many examples of good practice from within ELIXIR already exist: CEITEC - the Central European Institute of Technology⁴⁴ in the Czech Republic and partner in ELIXIR Czech - has received over EUR 200 million ERDF; ELIXIR Estonia has been funded through an Estonian ERDF programme for internationalising science; and CSC's (ELIXIR Finland) new data centre in Kajaani⁴⁵ has been funded through ERDF. A small number of ELIXIR Nodes will be provided with the capacity to understand better their local RIS3 priorities and develop concrete Business Cases for ESIF support for the operations of their Nodes.

2.1.4 Enhance the role of the Union in international organisations and multilateral forums

ELIXIR's infrastructure is not built *de novo* – it coordinates, extends and adds value to existing national and international resources. Thus, a strong partnership between the Commission, Member States, Associated Countries and other stakeholders (e.g. EIROForum organisations, e-Infrastructures, publishers and patient organisations) is a critical success factor for ELIXIR. As a coordinated European research infrastructure ELIXIR provides a common strategy and approach in discussions with global partners (e.g. the NIH BD2K initiative) and acts as Europe's voice in global discussions on life-science data, standards and sharing.

2.1.5 Support progress towards the development of global research infrastructures

Many life science disciplines have a long tradition of global data sharing and collaboration, and this is particularly true of ELIXIR. For example, efforts within the Swedish ELIXIR Node will focus on integration of the Human Protein Atlas (HPA) with other ELIXIR resources. The papers describing the HPA have received over 1,000 citations globally and it received more than 750,000 visits during 2013, of which approximately 60% was from outside of Europe.

ELIXIR is considered by the G20 as an infrastructure with great potential for global collaboration and membership and was presented as such during the 15-16 December 2014 meeting of the Group of Senior Officials. Task 4 of WP13 will work towards ensuring that countries from outside Europe join ELIXIR as members.

ELIXIR provides a single, coordinated voice for Europe in global discussions and by establishing a common approach supports the integration and interoperability of data resources towards a global research infrastructure. ELIXIR's Core resources have a strong standing as part of long-running global consortia. They are, and will continue to be, central in our collaborations with e.g. NIH BD2K and the GA4GH on global agreements for data exchange standards, data security framework and, increasingly issues such as data identifiers and federated access.

2.1.6 Enable researchers to address societal challenges with a global dimension such as climate change

Open life science data drives major societal value and truly facilitates researchers to solve the Grand Challenges. For example the identification of novel risk factors for Alzheimer's disease based on a large-scale meta-analysis⁴⁶ are founded on prior estimates on human genetic variation calculated from public datasets such as the 1000 Genomes. Another example is the development of the molecular design and docking tools that are employed on a daily basis in drug-discovery laboratories across the globe. The development and validation of drug-design tools, many of which have been successfully commercialised, has relied on carefully curated datasets extracted from publically archived data resources such as the PDB.

This integrated infrastructure will be essential for European life science research as the enhanced technical architecture will facilitate access to well-curated data, facilitate international collaboration and ultimately play an integral role in the transformation of bio-industries.

⁴⁴ CEITEC: Central European Institute of Technology: www.ceitec.eu

⁴⁵ DIGITICE Finland. <http://www.digitice.fi/web/?page=160>

⁴⁶ Jean-Charles Lambert et al., 'Meta-Analysis of 74,046 Individuals Identifies 11 New Susceptibility Loci for Alzheimer's Disease', *Nature Genetics*, 45 (2013), 1452–1458.

ELIXIR-EXCELERATE seeks to accelerate the impact of ELIXIR within the bioinformatics community and deliver excellence in innovation through wider access to scientific data⁴⁷, reduce research inefficiency through sharing high-quality reusable data⁴⁸, and encourage European Union R&D activities from foreign enterprises⁴⁹.

2.1.7 Raise the technological level of the European industry and SME's, thus improving their competitive position, through their involvement in research infrastructures development and service provision.

Industry's interest in, and usage of, Europe's bioinformatics resources is high and promoting the future innovation potential and industry impact is an important objective of ELIXIR-EXCELERATE. This clearly extends beyond the obvious industry 'users' of data and related services, and includes professional data publishers, SMEs providing data and bioinformatics services and tools as well as hardware and infrastructure providers. Indeed, ELIXIR-EXCELERATE's impact on innovation and industry can be summarised as three-dimensional:

Innovation by industry users

In the field of life science, industry usage of current bioinformatics resources is already high as demonstrated by the 110 million hits from commercial users to the EMBL-EBI website in 2013⁵⁰. Users from industry range from SMEs to large multinationals and the sectors covered include pharmaceutical, biotechnology, food and agriculture. These industries are major employers globally, generating wealth and supporting the transformation to a knowledge-based economy. Analysis of bioinformatics databases underpins the molecular design and docking tools in daily use in drug-discovery laboratories across the world^{51,52} and is imperative for commercial drug discovery. Beyond the pharmaceutical and biotech industries, an impressive volume of companies active in aquaculture, food, livestock and agriculture domains require access to ELIXIR's public data infrastructure.

Many of Europe's high-tech SMEs have business models built around developing commercial services on top of public open data platforms (i.e. Eagle Genomics, Ingenuity). Through ELIXIR's coordination and comprehensive industry support programme (WP13), vast new opportunities for commercial exploitation of these data will arise.

Supply-side innovation

It is estimated that in Europe alone, cloud computing can contribute up to €250 billion to EU GDP in 2020, which could reach a total cumulative gain of €940 billion for the period 2015-2020⁵³. The life science sector as a consumer of HPC is already high and set to increase further⁵⁴. Stimulating supply-side innovation by Europe's HPC vendors and suppliers is also a target of the project. Collectively the investments made by Member States in the data centres and cloud services run by ELIXIR partners are extensive. EMBL-EBI and ELIXIR Denmark, for example, have both recently invested significantly in data storage. Through WPs 4 and 13, ELIXIR-EXCELERATE will provide a form to engage collectively with the HPC community, especially the ETP4HPC, which can support the long-term stimulation of this industry sector further.

Industry as operators of infrastructure

Many ELIXIR Nodes have already strong industry partnerships - indeed several ELIXIR Nodes such as ELIXIR Denmark (through Novo Nordisk and Lundbeck) and ELIXIR Netherlands (through DSM, Elsevier, BaseClear, Euretos and Genetwister) exist as public-private partnerships (PPP). The ELIXIR-EXCELERATE project will also support the initial operation of several Use Cases that address user research projects in several life science

⁴⁷ A reinforced European Research Area Partnership for Excellence and Growth (COM(2012) 401- July 2012) : http://ec.europa.eu/research/science-society/document_library/pdf_06/era-communication-towards-better-access-to-scientific-information_en.pdf

⁴⁸ Al-Shahi Salman et al. Lancet. 2014 Jan 11;383(9912):176-85. doi: 10.1016/S0140-6736(13)62297-7

⁴⁹ European Commission. 2012. Internationalisation of business investments in R&D and analysis of their economic impact http://ec.europa.eu/research/innovation-union/pdf/internationalisation_business-rd_final-report.pdf

⁵¹ Keiser, Irwin, Shoichet, Biochemistry, 2010. doi: 10.1021/bi101540g

⁵² Gaulton, Ovington, Future Medicinal Chemistry, 2010. doi: 10.4155/fmc.10.191

⁵³ ETP4HPC Strategic Research Agenda:

http://www.ftp4hpc.eu/wp-content/uploads/2013/06/ETP4HPC_book_singlePage.pdf

⁵⁴ McKinsey 2011. Big Data: The next frontier for innovation, competition, and productivity:

http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

disciplines. These exemplar projects have been selected to represent a broad range of disciplines that will not only serve to address critical areas for ELIXIR operations (data deposition, coordinated annotation of complex genotypes and phenotypes for many species and effective re-use of biological data for medical and biotechnology research) but also form strong partnership with industry and SMEs. A coordinated strategy to engage with PPPs will be developed in ELIXIR under the guidance of the ELIXIR IAC. This strategy will, combined with the Use Cases to demonstrate utility, strengthen these relationships and feature as an integral part of Node Capacity Building efforts (WP10).

2.2 Measures to maximise impact

2.2.1 Dissemination and exploitation of results

Documentation, training and communication are instrumental to enable the dissemination of project results that can reach every laboratory across our 500,000 strong user base in Europe. ELIXIR has resources for communication and outreach to anchor the project outcomes within user networks and scientific communities; in ELIXIR-EXCELERATE this is strengthened by integrating the network of Communications officers across ELIXIR Nodes, provision of a shared communication platform and materials and development of a joint communication plan that uniformly address national and international stakeholders (WP13). This will also identify and reach influential users that drive changes in policy/standards adoption and increase the usage of the platforms across different user communities.

Each ELIXIR Node addresses dissemination within the national communities and need strategies adapted to its own areas of activities and services. In addition to the communications and outreach efforts in WP9 these local functions are strengthened through the Node development activities in WP6. Establishing an ELIXIR wide training platform with shared training content and training events (WP9) are a critical part of our exploitation strategy.

Teaching and sharing of new skills with industry through participation in joint workshops, the ELIXIR SME and Innovation programme (WP9) will engage with SMEs and industry users across Europe to facilitate technology transfer, facilitate uptake of open data resources and increase exchange of innovative research to the industrial users.

2.2.2 Measuring impact nationally, in Europe and globally

The progress and impact of the project will be benchmarked through quantitative key performance indicators (KPIs). Project KPIs will be established by project management (WP12) based on the agreed benchmarks for ELIXIR operations and the emerging monitoring and metrics infrastructure (WP1) and reviewed by the ELIXIR SAB and Member state representatives at the ELIXIR Board. The KPIs will include progress and spending against the measurable deliverables (Table 3.1.c) and milestones (Table 3.2a) but also track the conclusion of organisational capacity built by the secretariat and Nodes (e.g. Established ELIXIR Hub-Node agreements, implementation of national organisations, delivery of training courses to Node staff).

In addition, external impact measures will be compiled regularly such as: number of projects using the shared services; number of users taking advantage of the platforms; publications in scientific journals (citing support either directly from this project or from the common services/innovation pipelines enabled by it).

Stakeholder interactions and effectiveness of the communication strategy will be reviewed through the number of articles about ELIXIR published in general press, websites, online journals as well as social media indicators (e.g. www.impactstory.org).

2.2.3 A consolidated data management strategy for European life science

A data management strategy defines how to handle the full data lifecycle of organisation. This includes how to develop and execute architectures, policies, practices and procedures for the generation of information. In science data management plans apply more specifically to the process of controlling the information generated during a research project. Funding bodies increasingly require grant-holders to develop and implement data management and sharing plans (DMP). DMPs describe how the project data will be generated, how it will be exploited, made accessible for verification and re-use, and how it will be curated and preserved.

The default position for data in ELIXIR is Open Access; the ELIXIR Business case clearly outlined the case for maintaining and strengthening Open Access to biological research data – charging or restricting access to data seriously limit the ability of research organisations, both public and private, to exploit and create additional value

from 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.

The data management strategy for ELIXIR-EXCELERATE has an internal component that handles the data produced within this project and an external component that enables European life science to effectively plan and manage research data through robust and scalable services. Our goals for European data management, driven by the expertise of this consortium, through user services, support and training are outlined in Section 1.3.4. The capacity building, training and sharing of good practice is critical for achieving this vision, e.g. the Dutch ELIXIR node provides a national solution, including supporting software, for development of data management in close collaboration with national funders. ELIXIR is a mechanism to spread such solutions across the ERA.

Based on the well-defined exemplar Use Cases (WPs 6 to 9) and in coordination with WP3 (Data resources and services), WP5 (The ELIXIR Interoperability Backbone) and WP12 (Management and operations), ELIXIR will establish the fundamental standards, tools and policies required across the whole of the life sciences, generalising our initial Use Cases and developing long lasting solutions. These efforts will enable users to effectively manage and access data generated across the entire ERA such that data can be preserved, cited, and re-used with consideration to personal privacy, ethics and potential IP concerns. The Use Cases validate the utility of our services and compliance of our integration and interoperability; the capacity building and training scale and sustain our approach.

Furthermore, ELIXIR will work with other ESFRI biological and medical research infrastructures to integrate their data resources, services and standards; this involve jointly developing and implementing standards for research data management, publication and interoperability. Together with the other research infrastructures, partners, and large research programmes, ELIXIR will establish joint technical strategies and Memoranda of Understanding (MoU) that guide local and national shared efforts in life-science data management and resource provision. Nevertheless, there remains a risk that incorporating a large number of stakeholders into Europe-wide programmes fragment data management. Consequently, ELIXIR will play a fundamental role in ensuring harmonised data management strategies as many ELIXIR Nodes, including EMBL-EBI, have existing significant infrastructure that also includes responsibility for national research data management as well as direct consultancy links to research projects. Our partnership also include e-Infrastructures were e.g. EUDAT and OpenAIRE provide services for storage and dataset identification, WP4 aim to establish a structured collaboration with e-Infrastructures that support the development of the European data ecosystem. ELIXIR will play a particular important role through the fundamental role of our Core Data Resources as authorities for uniquely identifying entities such as ‘genes’, proteins’ and experimental findings. ELIXIR-EXCELERATE will provide further support to data management principles by ensuring Europe has the appropriate skills capacity for data management and the implementation of data standards.

At the national level the ELIXIR Nodes are 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 will enable both long-term data archiving and access but critically also enable accessibility and full data integration to make the best use of Europe’s collective and expanding capacity. This strategic objective is tightly coupled to the provision of ELIXIR Core Resources; ELIXIR will close the cycle of data annotation, deposition, provision and integration for reuse.

ELIXIR’s international collaboration with foreign infrastructures, such as the NIH BD2K programme in the US, will further serve to represent European data management standards in the pursuit of seamless international data management with interoperable legacy information.

The ELIXIR-EXCELERATE project does not request funding for generation of new data and while the use-cases are performed in close collaboration with user communities the focus, and purpose, of these efforts is on management, integration and reuse of the researchers data. Deposition into the public data resources managed by ELIXIR nodes remains the best option for long term open data management into the life sciences, recognised by funders and publishers, our strategy is centred on effectively enable this deposition through services, training and influencing policy. Nevertheless, ELIXIR-EXCELERATE will not be part of the EC Open Data pilot: while promoting and enabling open access to research results are a central tenet of this infrastructure the decision on how to manage the data, including human data consented for research, sits with the data owner and not with the infrastructure operators. ELIXIR-EXCELERATE does not generate new data and will not dictate the participation of our users.

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 a priority for ELIXIR Services (WP9 and WP12).

2.2.4 Knowledge management and innovation strategy

The fundamental principle for knowledge management and intellectual property (IP) within ELIXIR, established through the *ELIXIR Consortium Agreement* and Hub-Node *Collaboration Agreements* is that intellectual property remains with the legal entity employing staff that generates IP. This principle will also govern the work in ELIXIR-EXCELERATE.

The ELIXIR Board has established a high level working group that include representatives from member state funders that will analyse ELIXIR's policy and develop a clear Charter and recommendations for Open Access and Intellectual Property Policy. The work is currently underway and results will be presented to the ELIXIR Board in November 2015.

As described in Section 2.2.3 now new data will be generated within this project but other knowledge assets such as software, dictionaries, ontologies, training materials, processes and policies. It is important to distinguish between the ownership of Intellectual Property and what the owner elects to do with the assets. Similar to data the default position for other knowledge assets is open; software, standards, ontologies generated by ELIXIR's partners are in general deposited in community archives (e.g. GitHub, BioSharing, etc.) - often based on strong institutional policies. Independent of the ELIXIR-EXCELERATE there is a working group, including broad workshops with Nodes, planned and budgeted in 2015 that will define ELIXIR practice. ELIXIR-EXCELERATE will implement the agreed practice.

The current pan-European investment in drug-discovery translational infrastructure through IMI Knowledge Management projects has identified a significant need to establish a long-term strategy for sustainable medical knowledge management and translational infrastructure. ELIXIR will play a key role in developing this strategy and will engage intensely with other consortia to aid in the strategy development.

With the support of WP13, the ELIXIR-EXCELERATE project will develop and deliver shared services through bilateral collaborations with large research consortia and public-private partnerships, such as the IMI and the Bio-based Industries Consortium (BIC), to ensure that knowledge management efforts are aligned and sustainable in the long term. In addition, other new initiatives may emerge during the execution of the ELIXIR-EXCELERATE project, in particular through the Use Case studies (WP6 to 9), and ELIXIR will harness the strength of national Nodes to ensure that it is best placed to engage in these emerging initiatives.

The ELIXIR innovation strategy is based on open access to public research data, tools and services; we have a large constituency of industrial users and applications with many long standing partnerships. Thus the focus of our innovation efforts (WP13) is on awareness and training for life science SME across all nodes and on strengthening local partnerships. ELIXIR nodes will 'project' the capabilities and open assets of the whole infrastructure into the local innovation environment and, conversely, will be able to expose local SME partners in e.g. research proposals and large consortia.

2.2.5 Communication activities

ELIXIR exists to serve its thousands of users across the life sciences. Bioinformatics data services and tools are best developed with the full participation of users in a feedback loop and standards are most effective when widely adopted. It is therefore critical to ensure that current and future users are aware of the services, resources and developments that are being established.

The Communications, Industry and Community Engagement Work Package (WP13) will establish the channels and deliver the content to ensure that outputs from ELIXIR-EXCELERATE are disseminated effectively to all relevant stakeholder communities, including users in academia and industry, SMEs, policy-makers, related initiatives and the general public. A major objective of WP13 is to develop a coordinated and comprehensive Communications Strategy for ELIXIR, which builds on the communications presence of partner organisations. Collectively, partners represent the major institutes in the bioinformatics domain and the communication channels of many ELIXIR Nodes are already established and effective. ELIXIR-EXCELERATE will harness these existing channels and deliver new content to what is a large and heterogeneous stakeholder community.

The regular interaction of ELIXIR partners through ELIXIR’s scientific committees⁵⁵ and the operation of a new ELIXIR Communications Experts group (WP13, Task 1.2) will ensure an effective flow of information between partners. The ELIXIR Communication Expert group, including experts from across ELIXIR Nodes, provides a mechanism to plan, coordinate and deliver joined-up and effective outreach activities across Europe. The scientific and technical deliverables developed within ELIXIR-EXCELERATE will receive maximum exposure across this network and beyond.

A regular stakeholder newsletter will ensure the flow of information to stakeholders is consistent and of high quality. ELIXIR partners’ existing social media platforms will be harnessed and further developed. Social media and the project website represent the primary tool to communicate to the general public and other community stakeholder organisations.

In terms of scientific publications, ELIXIR-EXCELERATE will aim to publish Open Access articles in the appropriate journals to provide maximum transparency on new developments. Supported by the Communications Expert group, this approach will ensure the widest possible reach to all stakeholder groups.

Special attention will be paid to ELIXIR stakeholder groups as identified during the Preparatory Phase, which includes users in academia and industry, SMEs, policy-makers, HPC vendors, data providers, patient associations and the general public. ELIXIR’s Communication Strategy will ensure that the right message is presented to the right stakeholder via the correct channel. All Hands Meetings, workshops and scientific conferences will ensure that stakeholders are engaged effectively. ELIXIR-EXCELERATE will enable: annual representation at the European Conference on Computational Biology; representation at BioIT World every two years; representation at regional bioinformatics conferences such as SocBIN and BeNeLux BBC; and representation at the International Conference on Research Infrastructures.

Monitoring for efficiency

ELIXIR’s Communication Strategy (WP13, Task 1.1) will, once developed, exist as a live document, subject to annual updates. This will ensure that it can respond to new forms of communication, particularly social media channels, which may not yet exist as well as adapt to the new members that join ELIXIR. It will also allow partners to carry out reviews of the effectiveness of the strategy and introduce any necessary updates. The ELIXIR Communications Expert group will have responsibility for implementing the Communications strategy, and the ELIXIR Board, SAB and IAC will review and comment on its effectiveness for funders, users and industry respectively.

Usage of software such as Mailchimp and HootSuite will allow us to monitor the real-time effectiveness of our newsletters and social media campaigns. In addition, the ELIXIR web developer (WP12) will use Google analytics to assess the effectiveness of the ELIXIR website, monitoring traffic to and from the main website and ELIXIR Node pages and the Service Registry developed in Work Package 1.

3 Implementation

3.1 Work plan — Work Packages, deliverables and milestones

This project will accelerate and deepen the integration of national bioinformatics research infrastructure and the EMBL-EBI to transform data management for European life science. Coordinating and enhancing existing services, building capacity and human capital throughout the distributed organisation and establishing a supporting organisational framework, will achieve this.

The work package structure is designed to align a large and distributed project around the three main goals of this project (Section 2.3.1) where meeting the objectives of each Work Package would deliver the specific outcomes necessary to reach each of the goals.

There are two main mechanisms to manage coordination and dependencies between Work Packages (and, in the fully operative phase, within ELIXIR):

⁵⁵ ELIXIR Governance: <http://www.elixir-europe.org/about/groups>

- In ELIXIR-EXCELERATE our ‘use-cases’ (WP6 to 9) provide vertical integration of ELIXIR resources and services towards specific communities and challenges such as management and re-use of sensitive human data (WP9).
- The three permanent ELIXIR working groups (Data, Technical, and Training Coordinators) are tasked with infrastructure-wide coordination in their respective domain and provide a link to overall infrastructure operations.

The phased timeline of the project (Figure 9) shows that deliverables are evenly spread throughout the project and aligned with the phases and phase transitions of the ELIXIR Programme. The initiation of this project supports a ramp-up of project activities and partner staffing but given the existence of national ELIXIR nodes following the ELIXIR preparatory phase it is expected that all partners will rapidly be fully resourced for the implementation of the technical, organisational and scientific components of this project. The expected outcome is a stable and robustly validated set of services, supported by training and local capacity, offered to the research community from a unified infrastructure, for example through virtual access schemes or participation in research consortia. The final, open meeting and ‘Landmark in Bioinformatics services’ conference is a key part of this strategy and will serve to effectively showcase ELIXIR capabilities to a large user community as well as global partners and stakeholders.

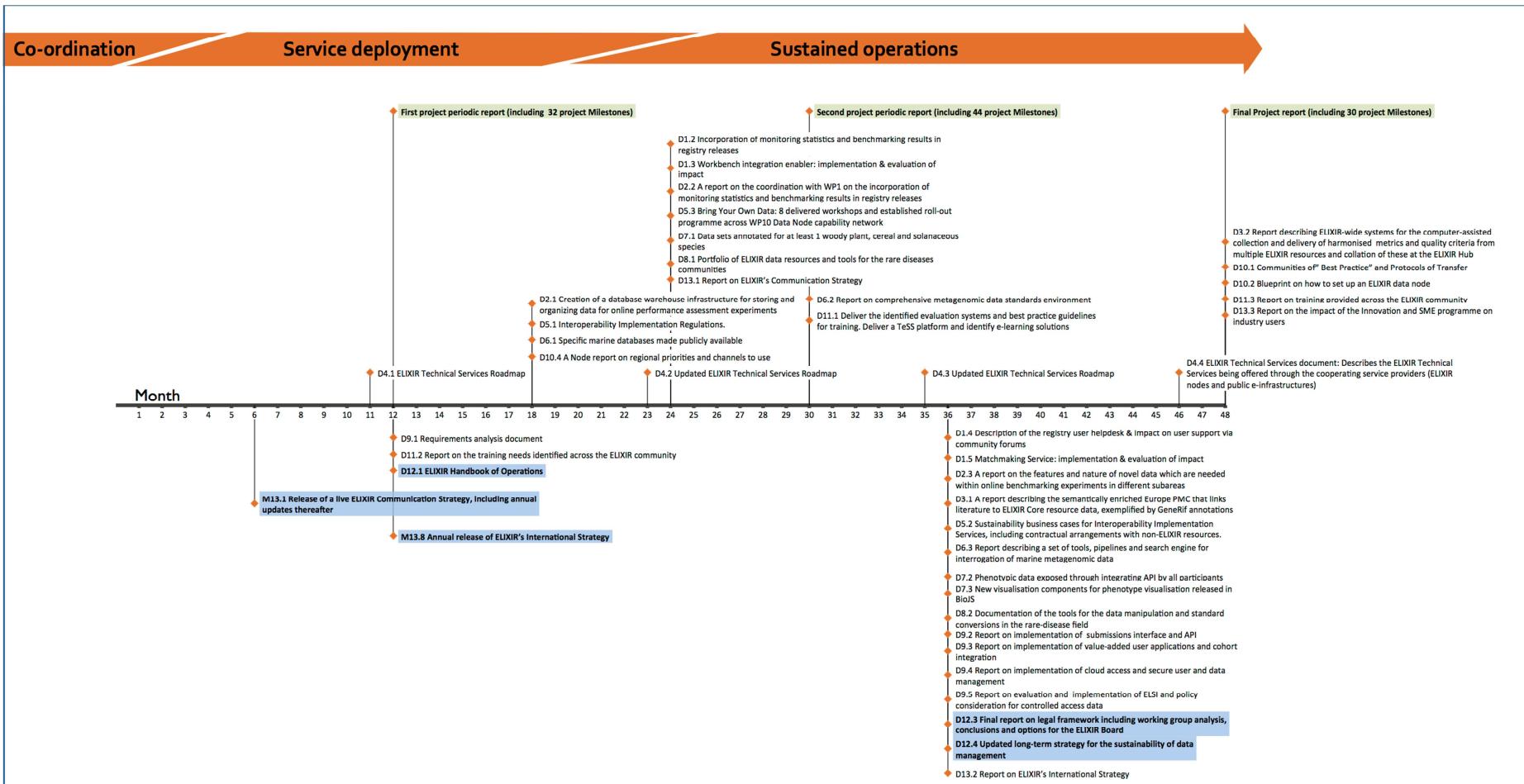


Figure 9: Overview of ELIXIR-EXCELERATE deliverables. The implementation deliverables and milestones that complete tasks recommended by ESFRI are shaded blue. The timeline from ELIXIR programme with the infrastructure implementation phases is shown for reference. To save space, the following, repeating deliverables are not included in the figure: D1.1, Registry release with comprehensive coverage of ELIXIR node resources, including resource data format curation and analysis (M12, M24, M36, M48); D8.3, Report on the ELIXIR workshop organised with the rare-disease communities (M12, M24) and D10.3, ELIXIR Advanced workshops on genome annotation (M4, M22).

3.1.1 Work Package descriptions (Tables 3.1a)

Work Package number	1	Start Date or Starting Event					M1
Work Package title	Tools Interoperability and Service Registry						
Participant number	1	2	5	7	8	10	18
Short name of participant	EBI	UOXF	UTARTU	CNIO	CRG	IRB	INESC-ID
Node	EMBL-EBI	UK	EE	ES	ES	ES	PT
Person-months per participant:	12	6	43	2	11	9	4
Participant number	22	26	27	30	36	39	
Short name of participant	UiB	SIB	CNRS	IP	MU	DTU	
Node	NO	CH	FR	FR	CZ	DK	
Person-months per participant:	18	9.5	9	12	18.2	76	

Objectives

WP1 will deliver a discovery portal built upon a federated curation of a wide range of key resources for bioinformatics resources world-wide.

It will involve service monitoring, resource integration, interoperability aspects, and community centred benchmarking efforts. All activities, including intensive user support, are focused around delivering impact for end-users across academia, health organizations, and industry. The ELIXIR Tools and Data Services Registry is the cornerstone of the WP.

WP Leads: Søren Brunak (DK) and Alfonso Valencia (ES)

Description of work

Based on its first release in January 2015, WP1 will further develop the ELIXIR registry mechanism, interfaces and content upkeep strategy. The WP contains plans for the development and extension of its functionality and scope (Tasks 1.1, 1.2 and 1.5). The federated curation of the registry will ensure comprehensive content and high quality annotations, both of which are essential for the sustainable impact of the registry in the community. Scientific and technical consistency and utility will be achieved by using the EDAM controlled vocabulary. Exposing the results of efforts addressing tool benchmarking and monitoring of the resources listed in the registry will provide the end-user with a robust, scientifically relevant measure of tool quality and performance. Furthermore, the work on workbench integration and interoperability will lower the cost to developers of integrating their resources in key workflow environments, and assist the users with establishing and updating their day-to-day workflows. Finally, WP1 contains plans for comprehensive, registry related user support, which will ensure impact for users, and a dynamic management element, including marketing and community development to build the federated organization behind the registry. The user-centric approach will thus stand as the guiding principle for the entire portal and guard its relevance to the community.

Task 1.1: Federated Registry Curation (96PM)

This task will deliver essential scientific and technical coverage in the registry and the vocabulary (EDAM) that underpins registry consistency and utility. A major community curation effort is required, including vocabulary development, resource annotation and registration. To ensure that the curation is high quality and sustainable, it must be federated across registry stakeholders, hence a major priority is building and supporting the community of federated curators. In tandem, the curation will be accompanied by focused software and other technical developments, that automate, validate and embed the curation process in relevant software systems; the essential underpinning of sustainability.

The registry has two primary purposes; to help discover tools and services and use them. Discovery means to find, understand, compare and select. It is a prerequisite to (inter)operability, which demands a precise understanding of software dependencies. Our approach is based on the acceptance that software interoperability will, for the foreseeable future, be implemented primarily by developers rather than intelligent software agents. We will therefore, once a comprehensive set of ELIXIR Node resources are described in basic detail, extend the curation of the registry to annotate, using EDAM Format URIs (unified resource identifiers), the data formats that are

supported by tools and data services. From this, we will analyse the format-usage landscape to provide a basis for targeted software developments to improve interoperability of registered resources. We foresee these developments, which might include conversion of tools to use common formats, and development of format-converter software where needed, to be facilitated via the Matchmaking Service mechanism (D1.5).

The registry scope will be: 1. Comprehensive coverage of ELIXIR Node resources, including tools, data services (APIs) and host databases, prioritising ELIXIR-badged services and new resources from the Use Cases. 2. Coverage of other biomedical science Research Infrastructures (RIs), and key resources beyond ELIXIR (European and non-European). A task force will be comprised of ontology developers, curators, scientific domain experts and relevant technical experts. It will run *Curation and Usability* hackathons with the recurrent theme of curation: resource annotation and registration, with necessary EDAM development. To facilitate networking and community build-up, two types of social event will be combined with the hackathons: 1. *Knowledge Exchange Workshops*, including representatives of relevant infrastructures, institutes and projects, on themes related to the registry suggested by the community. 2. *Cross-domain Strategy Workshops* to gather technical officers from ELIXIR Nodes, RIs, key resources, and other key initiatives, to discuss and develop common approaches for registry curation across RIs internationally.

EDAM provides the registry with a consistent vocabulary for topics (general scientific and technical disciplines), operations (tool functions), types of data, and specific data formats and data identifiers. Task 1.1 will work with the existing EDAM community, develop its open governance and contribution mechanisms and deliver essential utilities to ensure that maintenance, validation and community development is sustainable in the long term. We will assess and validate coverage by correlating EDAM concepts to terms used for curation, which will then inform and drive necessary additions and desirable clean-ups (removal of concepts). We will develop focused essential utilities for EDAM maintenance including automation of the release process, basic validation of content, reporting of changes between versions, deployment to ontology browsers such as BioPortal and OLS, technical integration of EDAM with applications including the registry and others, mapping of provider-supplied terms and phrases to EDAM, and revise annotation upon new EDAM releases.

To underpin the sustainability of the federated curation, this task will deliver focused software and other technical developments that will automate the registration and update of provider-supplied information, leveraging their own local software infrastructure where possible. We will work with providers to support them in doing this, and, where possible, adapt technically the local solutions to make them more broadly applicable to others. Further, in order to facilitate coverage, all relevant resource providers will be given smooth and convenient access to resource registration. This will be achieved by a combination of simple-to-obtain local login accounts and opening for using eduGAIN authentication to register resources.

Finally, this task will ensure that registered resource are citable, discoverable by the major search engines, and are placed in scientific context. It will also include technical mark-up to support “Semantic Web” applications, e.g. Schema.org-compatible microdata or RDFa to support Google “rich snippets” and other structured search results in the major browsers. Hence, the registry will promote the registered resources and deliver impact for developers and institutes by making resources rank higher in search results and hence more findable.

Task 1.1 partners: DK, NO, FR, CH, CZ, EMBL-EBI

Task 1.2: Benchmarking and Monitoring (15PM)

This task will support the monitoring and community benchmarking of analytical tools, in a systematic and sustainable way e.g. based on the efforts in WP2. Firstly, it will review the existing service quality and performance metrics and assess their usefulness in the context of a registry. This may require development of a light-weight controlled vocabulary capturing the concepts distilled from the preparatory activities above and those of WP2.

Task 1.2 partners: DK, ES, CZ, CH

Task 1.3: Workbench integration and interoperability (36PM)

There is general trend towards the use of workflows as a preferred environment for the convenient use of tools and data access, especially when resources must be used in combination with one another. This task will boost convenience and resource interoperability by implementing a Workbench Integration Enabler service that will develop the vision “register your software once - get it supported everywhere”. Technically, this service will translate the description of any tool or service that is registered in the Tools and Data Services Registry into the metadata format required by the existing major workbenches, including Mobyle, Galaxy and Taverna. Furthermore, we will develop a new, lightweight Service Launchpad for running tools and services which have programmatic access and which can be invoked using information available in the registry.

To develop the Enabler Service, we will align the registry software description model and the schemas used by the

workbench systems or required by the Launchpad, and subsequently revise the model and schemas to facilitate the metadata transfer. Furthermore, to prove the principle, new high priority tools and services, including those developed in the Use Cases.

Task 1.3 partners: DK, EE, FR, CH

Task 1.4: User support and derived registry development (36.7PM)

This task will provide direct and indirect user support to deliver impact for ELIXIR end-users. Direct support will be achieved primarily by leveraging the existing and highly popular user bioinformatics forums (BioStars, BioPlanet etc.). A User-support specialist will patrol such forums and respond to questions in one of four ways: 1) Where resources answering to the Users needs exist in the registry, a link to them in the registry will be provided via our API. 2) Where resources exist in the registry, but the registry API cannot be used to answer the question directly, they will request new features of the API and in so doing drive development of the Query Interface. 3) Where an appropriate resource exists but has not been registered, they will request the appropriate registry curator add it to the registry. 4) Where a registered resource exists that is close, but not quite what is required, they will forward feature requests to the appropriate developers, possibly via the Matchmaking Service (D1.5).

Indirect user support will be achieved primarily by ensuring the registry interfaces are highly usable and match very closely the needs of the user. To achieve this, we will run user experience sessions during the Curation and Usability h community. Scientific and technical consistency and utility will be achieved by using the EDAM controlled vocabulary. Exposing the results of efforts addressing tool benchmarking and monitoring of the resources listed in the registry will provide the end-user with a robust, scientifically relevant measure of tool quality and performance. Furthermore, the hackathons (see Task 1.1) in order to evaluate usability. We will develop comprehensive Good Practice Guidelines for the curation of the registry in all aspects, but in particular the annotation of common types of resources using EDAM. We will also participate in the development of an ELIXIR Experts Registry where users can discover relevant expertise within the ELIXIR network, and an ELIXIR User Helpdesk to answer general questions concerning use of the registry, forwarding specialised scientific and technical enquiries to relevant experts.

Task 1.4 partners: DK, CH

Task 1.5: Management, marketing and community build-up (46PM)

This task will build the federated organisation primarily by identifying and facilitating key collaborations between registry stakeholders. This will be achieved by organising ‘Resource Synergy Meetings’, where we will identify and encourage targeted software developments, e.g. to coordinate curation and data sharing. We will also promote resource integration and usability, e.g. by cross-linking resources and through API harmonization. As a prerequisite to these Synergy Meetings, a Resource Metadata Catalogue, listing all relevant resources, their scientific and technical scope, and information fields (schema), will be compiled and used to compare providers and identify redundancies. We will also use these meeting to cross-link the Tools & Data Services Registry with other key ELIXIR registries, for example the Training Materials Registry, the ELIXIR Events Registry, and the Experts Registry.

This task will also develop an oversight and management strategy and leverage partners within and beyond the ELIXIR organisation to implement strategy. To drive delivery, it will identify and encourage collaboration, monitor actions, identify delays, and intervene where necessary. It will raise community awareness and therefore impact by contributing to a forceful marketing campaign via all appropriate marketing channels, including popular social media. It will provide support to funders, publishers and others at the EU and national level, that policy is aligned with the aims of the registry organisation.

Task 1.5 partners: DK

Deliverables

D1.1 Registry release with comprehensive coverage of ELIXIR Node resources, including resource data format curation and analysis (Task 1). (**M12, M24, M36, M48**)

D1.2: Incorporation of monitoring statistics and benchmarking results in registry releases (Task 2, report). (**M24**)

D1.3: Workbench integration enabler: implementation & evaluation of impact (Task 3, publication). (**M24**)

D1.4: Description of the registry user helpdesk & impact on user support via community forums (Task 4, report). (**M36**)

D1.5: Matchmaking Service: implementation & evaluation of impact (report). (**M36**)

Work Package number	2	Start Date or Starting Event				M1
Work Package title	Benchmarking					
Participant number	7	12	15	26	39	
Short name of participant	CNIO	BSC	CNAG	SIB	DTU	
Node	ES	ES	ES	CH	DK	
Person-months per participant:	12	19	9.6	23	6	

Objectives

The concept of assessing bioinformatics methods in terms of quantitative performance and user friendliness is crucial to the development of the infrastructure in the general field of bioinformatics.

Accordingly, WP2 will focus on the following objectives:

- Systematically organize the relations to communities already running benchmarking exercises within biology and medicine. (Task 2.1)
- Development and maintenance of a generic infrastructure to support benchmarking exercises in different subareas. (Task 2.2)
- Develop the technology to perform online, uninterrupted methods assessment in key areas of bioinformatics. (Task 2.3)
- Development and implementation of data warehouse infrastructures to store benchmarking results and to make them accessible to benchmark participants and method developers for subsequent transfer to the ELIXIR registry. (Task 2.4)
- Development of the procedures to create standards in the different fields subject to benchmarking. (Task 2.5)
- Establish workshops, hackathons and jamborees for different user communities. (Task 2.6)

Work Package Leads: Alfonso Valencia (ES) and Søren Brunak (DK)

Description of work

World-wide, bioinformaticians already engage significantly with evaluation exercises in the form of open challenges. The role model for this type of effort is the still on-going “Critical Assessment of protein Structure Prediction, or CASP, which is a community-wide, world-wide experiment for protein structure prediction taking place every two years since 1994. This effort, as well as others, provide research groups with an opportunity to objectively test their prediction methods and delivers an independent assessment of the state of the art to the research community and software users. CASP has inspired many other similar experiments, including analysis of text mining methods (BioCreative), docking (Capri): force-field evaluation for atomistic simulations and benchmarking of small molecule docking, evaluation of multiple alignments, NGS sequencing variation analysis, gene finding and others. All these community efforts have a similar organization and similar basic infrastructure needs. A further challenge is to make these challenges not only static annual or bi-annual competitions, but to evaluate the systems in an online fashion, which would make them more sustainable. A few experiments were organized in the past (e.g. the EVA effort organized by Burkhard Rost and co-workers), but abandoned for technical reasons. The WP will reintroduce these concepts such that methods can be benchmarked based on data, which are novel to all, including the methods developers in more sustainable frameworks.

It is an essential part of the European infrastructure since:

- It provides a strong connection between the ELIXIR infrastructure and the communities carrying out benchmarking exercises within their expert knowledge domains.
- It is directly linked to the information to be disseminated in the ELIXIR tools and services registry.
- Provides direct access to information on methods and performance measures for end-users.
- Provides the benchmarking data needed for training of new methods making progress in the different subareas of field.
- Furthermore, the benchmarking activities will provide a great vehicle for developing novel standards for data and methods thus also providing useful input to other WPs.

Task 2.1: Organize the relations with communities already running benchmarking exercises (7.6PM)

Obtain agreement with existing communities on the conditions of challenges, organizes, formats, goals and other organizational issues that can lead to harmonization of efforts world-wide in addition to division of labour

decisions.

Partners: ES, DK

Task 2.2: Development and maintenance of a generic infrastructure to support benchmarking in different areas (12PM)

The emerging ELIXIR registry will be a reference for the research community. The methods to be benchmarked will be described in the registry with the proper version control and automatic access procedures. At the same time a generic infrastructure is needed in order to organize data for new and existing benchmarking efforts. WP2 will be responsible of implementing the guidelines and standards for data organization and submission of the different methods subsequently to be incorporated in the registry. We will also collect qualitative and quantitative data about the usage of these services, and different indicators about the service itself (i.e. data grow rate, uptime, etc.). These data will be stored in the data warehouse infrastructure (Task 2.4). Opinion leaders in the field will be surveyed about how useful they consider the resources are and the results will be included in the registry.

Partners: ES, DK

Task 2.3: Develop the technology to perform online, uninterrupted methods assessment in key areas of bioinformatics (18PM)

In order to make online methods performance assessment several infrastructure elements need to be in place in order to support the various challenges. These include:

- Organization of a collection of training data (validated by experts),
- Identification, collection and organization of a collection of testing data which are kept secret,
- Community agreements on the data standards, submission formats and evaluation methods (quality assessment),
- Hosting or accessing methods (e.g. by programmatic access) to obtain results from them automatically without human intervention,
- Parsing, organization and display of the results with proper statistics and comparison facilities.

Partners: ES, DK, CH

Task 2.4: Development and implementation of data warehouse infrastructures to store benchmarking results and to make them accessible to users and method developers (18PM)

In this task we will develop with each one of the communities the necessary data framework and method standards, based on the community recommendations and the experience acquired in each challenge. The standards will be essential for the operation of the benchmarking infrastructure. The standards will also facilitate the end-users interpretation of the results, and we will develop tools for the conversion of the data from different formats into the most frequent standards in collaboration with WP1. We will also develop tools to diagnose and rate the ELIXIR resources according to the level of agreement with those standards.

Partners: ES, DK, CH

Task 2.5: Development of the procedures to create standards in the different fields subject to benchmarking (7PM)

Data warehouses are key to storing and analysing the very large collection of data that will be generated by the prediction methods. Part of the WP2's mission is to store these data in a way such that they can be used for the continuous evaluation of the methods and for training of new methods. With time the ambition is that this infrastructure will be the main infrastructure of the different communities in subareas from protein structure and feature prediction to genomics and chemoinformatics.

Partners: ES, DK

Task 2.6: Establish workshops and jamborees for the different user communities (7PM)

The final goal of the infrastructure is to provide users with a continuous evaluation of bioinformatics methods and to have a positive influence on tools development. The effort requires a robust system for the provision of testing data, running methods and evaluating results. The design of the most adequate representation system for each of the areas will require additional software development efforts. In the training workshops and jamborees representatives of the scientific communities involved in the project will participate alongside new communities interested in adapting their challenges to the use of the infrastructure. The training aspects will be coordinated with the other training efforts in the project.

Partners: ES, DK

Deliverables

D2.1: Creation of a database warehouse infrastructure for storing and organizing data for online performance assessment experiments (**M18**)

D2.2: A report on the coordination with WP1 on the incorporation of monitoring statistics and benchmarking results in registry releases (**M24**)

D3.3: A report on the features and nature of novel data which are needed within online benchmarking experiments in different subareas (**M36**)

Work Package number	3	Start Date or Starting Event				M1	
Work Package title	Data Resources and Services						
Participant number	1	1	7	9	14	16	26
Short name of participant	HUB	EBI	CNIO	FVIB	UPF	IMIM	SIB
Node	EMBL-ELIXIR	EMBL-EBI	ES	ES	ES	ES	CH
Person-months per participant:	8	77	12	3.6	14.1	5.5	66

Objectives

The overall objective of this Work Package (WP) is to build a framework to inform and drive the sustainable development of Europe's core life-science data resources. The goals of WP3 are to:

- Promote excellence in resource development and operation through providing a unified framework for the identification and monitoring of key bioinformatics resources across Europe.
- Increase the sustainability of manually curated resources, which, while of high value and essential to the life-science community, are very labour-intensive to operate. This will be done by integrating the literature with data, with particular emphasis on maximizing value added by curation.

Work Package Leads: Jo McEntyre (EMBL-EBI) and Christine Durinx (SIB)

Description of work

The core mission of ELIXIR is to build a sustainable infrastructure for biological information across Europe. Data resources and services (hereinafter referred to as "resources") are a key part of this infrastructure and can vary; from submission databases that contain research data outputs such as DNA sequences (e.g. European Nucleotide Archive), to highly dynamic resources that aggregate, process and visualise research data, often adding layers of value through manual curation by highly qualified personnel. (e.g. UniProtKB/Swiss-Prot).

Task 3.1. Promote and implement good practice in data resource and service management through the formalization of metrics and quality criteria enabling the identification of ELIXIR Named and Core Resources, and informing their life-cycle management (24PM)

The first requirement for the development of a unified framework for the management of key bioinformatics resources across Europe is to identify which resources (a) meet a variety of quality criteria with respect to scientific impact and level of service, and (b) which of these are of fundamental importance to the life-sciences community. Therefore, ELIXIR resources will be identified and classified into two categories:

- *ELIXIR Named Resource* will be attributed to ELIXIR Resources from the project partners (ELIXIR Nodes) that are compliant with a set of metrics/criteria that guarantee their quality.
- *ELIXIR Core Resources* will be the subset of ELIXIR Named Resources that, based on metrics/quality criteria, are of fundamental importance to the life-science community and that are considered to be an authority in their field with respect to one or more characteristics.

Definition of clear metrics/quality criteria that measure current and projected use of ELIXIR resources as well as their scientific impact, and the reliability of the service, will underpin the identification of ELIXIR named and core resources and provide data to inform life-cycle management on an ongoing basis.

The initial set of metrics and quality criteria for ELIXIR resources will be identified based on prior resource

management experience of WP partners, on the work completed by the ELIXIR technical coordinators group, and experiences from other disciplines such as the Data Seal of Approval project⁵⁶. Formal opportunities for ELIXIR-wide review of the proposed criteria will be conducted through presentations and workshops aligned with project management meetings.

Metrics and quality criteria will evaluate both the scientific impact of the resources on the life-science community and the reliability of the service. They include, but are not limited to: uptime and download speeds, usage statistics (IPs, page views, downloads), citations in the literature, data submission rates, international collaborations, programmatic access, and curatorial effort.

In defining measures of quality it is important to recognise the context in which the service is being provided and to base categorization on a range of criteria. For example, a resource that serves a small community may not have as many page views as a large resource, yet reach 90% of the community it supports. Other may play a foundational role to derived services. It will be important to differentiate between submission databases and “added-value” databases that organize, curate, or otherwise represent submitted data, as the profile of use of these types of resource may be very different.

Equipped with an agreed set of criteria, it will be possible to effect a number of actions:

- Identify new resources for inclusion in the ELIXIR set.
- Set quality standards for emerging resources and inform their development.
- Build confidence among users through the identification of ELIXIR resources directly (such as a “badge” on the resource itself) and through a variety of portals such as the Tools and Data Services Discovery Portal (WP1).
- Monitor usage trends and manage resource life cycles effectively using objective criteria.
- Build understanding of the impact of ELIXIR resources both within the ERA and within global research infrastructures.

Resource development based on Metrics and Quality Criteria

Alongside the definition of the metrics and quality criteria, coordinated management processes will be required to review candidate resources, encourage use of ELIXIR-approved badges (or similar), and monitor resource life cycles. We expect the organizations running the resources to actively contribute to this process, and that this in itself will provide feedback mechanisms to improve and refine the criteria. This coordinated feedback model will have the added benefit of providing opportunities for peer-peer capacity building (WP10) in the areas of life-cycle management and sustainability, and metrics/quality criteria implementation as we share expertise between ELIXIR Nodes.

Partners: EMBL-EBI, CH

Task 3.2. Inform ELIXIR Resources life-cycle management and improve the ELIXIR Resource portfolio through the implementation of an active and computer-assisted infrastructure for the monitoring of ELIXIR Named and Core Resources based on the metrics and quality criteria formalized in Task 3.1 (76.1PM)

In the interest of transparency and to build excellence across resources, metrics and quality criteria for ELIXIR named and core resources will be held centrally at the ELIXIR Hub (see also WP12.3). Access to this collated data will be made available to all Nodes and resources involved, and potentially more widely as aggregated data.

In this task, technical processes will be developed to generate and collate the metrics and quality criteria agreed in Task 3.1. Operating in active mode over a period of time, the emerging trends will inform ELIXIR Resources life-cycle management and improve the ELIXIR Resource portfolio overall.

The processes developed will gather, report and upload metrics and other quality criteria in agreed formats and to an agreed timescale to the ELIXIR Hub.

The need to collate metrics/quality criteria centrally for analytical and comparative purposes raises questions regarding the technical implementation of such a system. There are a number of challenges in doing this, not least the willingness of the resource providers to share detailed metrics and quality criteria regarding their resource. Subsequent to this will be the need to provide confidence, particularly in the case of metrics, that what is being measured/reported from different resources is comparable in a fair manner; this will require sharing of methodological approaches (such as how robot traffic to websites is treated) through a shared understanding of what is considered a page view across different resources. Finally, agreement on a timetable and format for quality and metrics information will be required so that it can be easily collated in one place. These challenges may give rise to a need for technical effort in the participating resources and such requirements will be supported through the

⁵⁶ Data Seal of Approval project: <http://www.datasealofapproval.org/en/>

ELIXIR Hub core budget if required.

Partners: EMBL-EBI, EMBL-ELIXIR, CH

Task 3.3. Increase the sustainability of curated resources through literature-data integration and resource crosslinking (86.1PM)

The integration of the literature with data is critical for understanding the biological context of new results, for showing clear provenance of scientific assertions, and for discovering new information. While these are important activities for all of the scientific community using online resources, the requirement is most intense within scientific curation processes. The excellent quality of many European bioinformatics resources relies on manual curation, a process in which trained experts review experimental data reported in publications and extract relevant information for inclusion in data resources. This requires searching, reading, filtering, verifying and recording information; labour-intensive, and therefore costly, processes. However, curation saves time and adds significant value for researchers, obviating the need for potential users to individually seek out and synthesise threads of scientific information. Technological advancements in the past few years provide new opportunities to expedite the work of curators and also provide novel approaches to integrating the literature with data for the wider scientific community. For example, when a curator adds a new piece of information to a data record, the source article is cited in the record. However, it would be useful to link from that specific annotation directly to the precise point in the article that was extracted by the curator, for example, a figure legend. This will allow researchers and curators alike to understand exactly where that piece of information came from when viewing the data record, or conversely, to follow a link to see more data when reading the article - a connection that is currently not possible to traverse. Such developments will provide efficiency savings in resource and tool interfaces, reduce repetition, and when published, will provide granular deep links between the literature and data for users.

In this task, a roadmap for infrastructure that integrates the literature with data through a variety of novel approaches, including text mining, will be developed. Elements of this roadmap will be demonstrated by a collection of pilot developments that provide deep links between the literature and established or emerging ELIXIR data resources. Automated approaches, such as text mining, that identify and extract useful biological concepts will be a necessary part of this activity, from generating granular links to suggesting articles to curate in the longer term. Harnessing the expertise of the text and data mining community as a whole would maximize the impact of this aspect. This task aims to engage with existing database providers and novel Use Cases (WP6 to 9) to develop a roadmap that combines the above elements to develop an infrastructure for literature-data integration and enrichment, and furthermore to demonstrate this through a collection of pilot developments. To do this we will use known high-quality annotations such as GeneRifs (sentences extracted from articles that have been included in gene database records) and the Europe PMC database of life science research articles.

Partners: EBI, CH, ES

Deliverables

D3.1: A report describing the semantically enriched Europe PMC that links literature to ELIXIR Core resource data, exemplified by GeneRif annotations (**M36**)

D3.2: Report describing ELIXIR-wide systems for the computer-assisted collection and delivery of harmonised metrics and quality criteria from multiple ELIXIR resources and collation of these at the ELIXIR Hub (**M48**)

Work Package number	4	Start Date or Starting Event				M1
Work Package title	Technical Services					
Participant number	1	6	14	21	24	26
Short name of participant	EBI	NBIC	UPF	CSC	UiT	SIB
Node	EMBL-EBI	NL	ES	FI	NO	CH
Person-months per participant:	34	17	23.5	48	2	2
Participant number	30	32	36	37	39	
Short name of participant	IP	LIU	MU	CESNET	DTU	
Node	FR	SE	CZ	CZ	DK	
Person-months per participant:	6	12	40	24	6	

Objectives

The Technical services Work Package (WP) links the ELIXIR scientific programme 2014-2018 to the day-to-day technical service work in the distributed Nodes. The *research platform for life science* will be achieved through the following objectives:

- Develop a sustainable and supported research platform for implementing geographically and organisationally distributed Cloud, Compute, Storage and Authentication and Access infrastructure services collected in the ELIXIR registries.
- Manage external technical dependencies with e-Infrastructures and Nodes with ELIXIR technical coordinator group for services delivered as a priority for the ELIXIR-EXCELERATE Use Cases.
- Close collaboration with translational, bio-banking and imaging infrastructures at both the European and national level to ascertain that there are effective services to securely access and exchange data.

Work Package Leads: Tommi Nyrönen (FI) and Ludek Matyska (CZ)

Description of work

The role of the ELIXIR-EXCELERATE Technical Services WP is the practical integration of existing Technical Services available for ELIXIR in the Nodes and e-Infrastructure by testing and contributing to documentation and integration with small-scale programming and scripting where needed. Development is managed outside the WP. As a result of the tasks described below, WP4 will provide a generic integrated platform that can be tailored further for the ELIXIR-EXCELERATE scientific Use Cases (WP6 to 9), Training activities (WP11), and other ELIXIR pilots and projects to meet their specific needs. This includes user support, documentation and guidance to enable and promote technology adoption.

Work Package uses a mechanism of renewal of focus with the ELIXIR Heads of Nodes committee as necessary. If scientific needs change or disruptive technologies emerge that change the technical objectives heads of Nodes committee supports linking of the changed landscape of technical services implementation with the other Work Packages (e.g. ELIXIR resource governance, training, data resources, service registry). Involvement of ELIXIR heads of Nodes is used for securing physical information technology resources from Nodes, and making experts available for collaborative work.

Task 4.1: Leadership (53PM)

Subtask 4.1.1: Management and Coordination (26PM)

This task is responsible for coordinating technical work in the ELIXIR-EXCELERATE project and wider ELIXIR research infrastructure with WP12 building on the emerging community of technical experts in the ELIXIR task forces. In addition, the task establishes appropriate management and technical interfaces into the services and organisations the technical activities are dependent upon.

Partners: FI, CZ

Subtask 4.1.2: Provide a gateway to use European e-Infrastructure services for ELIXIR (GÉANT, EGI, EUDAT, PRACE) (13PM)

Regular requirements gathering from the Use Cases in WP6 to 9 and elsewhere in the ELIXIR community will define biological information service requirements and identify areas and activities that could be sourced by the European e-Infrastructures. Any planned service integration into the ELIXIR Technical Services will be identified in the regular Roadmap documents that will define a technical architecture and technology insertion roadmap. This should include defining the relevant ‘account managers’ in each public sector e-Infrastructure.

Partners: FI, EMBL-EBI

Subtask 4.1.3: ELIXIR technical community building and knowledge exchange (14PM)

Task grows the community of ELIXIR branded resource providers and sustains community of ELIXIR technical experts (i.e. ELIXIR technical coordinators and Node personnel) through engagement in major e-Infrastructure events, technical workshops, audio/video conferencing and other collaboration mechanisms. Working groups and task forces bring in relevant experts from outside ELIXIR such as e-Infrastructures. This task will be made in collaboration with WP1 (Tools) and WP12 (Management).

Partners: ELIXIR Nodes

Task 4.2: User Facing Support (71PM)

This task interacts with the individuals and projects that are users of the ELIXIR Technical Services platform through their defined Use Cases. The main consumers are the ELIXIR-EXCELERATE Use Cases (WP6 to 9) and

other ELIXIR activities (e.g. ELIXIR-EXCELERATE WPs, ELIXIR pilots, and external projects like EC funded Centres of Excellence or Virtual Research Environments).

Subtask 4.2.1: Technical requirements (28PM)

Gather and analyse the technical requirements for the Technical services platform in order to define the detailed technical specifications and interfaces of the technical service platform. One outcome is the classification of the different Use Cases with technical terms (e.g. small compute, large data input-output; large compute, data access management, etc.). This work will feed into WP12 concerning requirements requested and procured from any external service providers.

- (a) EXCELERATE Use Cases WP6 to 9.
- (b) ELIXIR tools registry, ELIXIR training events, data transfers to/from ELIXIR data resources, and authentication and authorization.

Partners: EMBL-EBI, FI, CZ, ES, NL, NO

Subtask 4.2.2: User support and integration (43PM)

Provide a support structure that can be applied to adopters of the ELIXIR Technical Services. This will be focused on the use of ELIXIR Technical platform e.g. for supporting organizing a training event. Task provides operational support for ELIXIR-EXCELERATE activities and externally funded ELIXIR activity (technical pilots and projects). This will take place through a single-point-of-contact ‘helpdesk’ function and ‘hackathons’ where users and the providers of the ELIXIR Technical Services work together to integrate functionality across AAI, cloud and data. As a result of this work a set of ‘recipes’ focused around user activities will be collected into a ‘cook book’ to enable community adoption (e.g. to run a Galaxy workflow environment on an ELIXIR-affiliated Cloud Resource with accounting if necessary).

Partners: EMBL-EBI, FI, SE, FR, NL

Task 4.3 Technical infrastructure integration (102.5PM)

This task focuses on the integration of Technical Services being delivered by individual ELIXIR Nodes and from the public e-Infrastructures in Europe to meet the requirements of the ELIXIR community (e.g. by establishing account manager relations with each e-Infrastructure). The strategy within this task is to focus on the integration of existing mature and stable services to ensure that these services are easier to uptake by bioinformatics Use Cases (WP6 to 9).

Subtask 4.3.1: ELIXIR AAI - Authentication, authorization (access) integration (28PM)

ELIXIR needs a service based on European *federated identity* that authenticates an individual is a member of a group (or has a particular role within a group) that can be managed remotely. *Group management* needs to enable delegate decision making to multiple individuals within a particular community (e.g. institutional representative within a project) and queries from other services.

- (a) Establishing an ELIXIR Identity: Federated identity technologies are fairly mature, as are many of the related tools (e.g. REMS, PERUN). This task ensures that ELIXIR research community is fully covered (including users whose home organization does not provide federated identities) and acts as a single IdP for ELIXIR branded services technical work. The task continues to integrate the existing services and ensures that they provide the interfaces needed for adoption within this Work Package, the project and externally (e.g. BMS RIs).
- (b) Providing additional AAI services: eduGAIN IdPs, Common IdPs, guest login, Proxy IdP, ELIXIR directory, Attribute self-management for users, Bona fide researcher management, Group/role management, Dataset authorisation management, Credential translation.

Partners: CZ, FI, EMBL-EBI, NL

Subtask 4.3.2: Cloud and Compute integration (42.5PM)

European and national compute centres service clusters and access to resource with open-source cloud technologies is growing. This task integrates the willing services to ELIXIR registry. The way how e.g. IaaS resources are consumed in the research community typically takes place on science-specific platforms and workflows. As a priority WP4 secures resources to support the scientific software workflows for the Use Cases WP6 to 9 and WP11 using the software environment workflows they have chosen for their data analysis framework (e.g. supporting provision of Galaxy as a service for marine metagenomics pipeline).

- (a) ELIXIR Cloud accounts: Integrate willing providers (e.g. Embassy Cloud), national level (e.g. CSC, SURFsara, Nordic Secure Cloud, MetaCentrum and CERIT-SC) and regional level (e.g. GÉANT, Helix Nebula and EGI cloud resources) and in the commercial sector (e.g. Amazon, Microsoft, Google). Mechanisms are needed to calculate virtual access costs that can be passed on to projects or funding

agencies. Key target is to make accounts to provide resources for WP6 to 9 and WP11 activities.

- (b) Enable SME access to ELIXIR cloud resources. We will support billing models such as monthly fee for service subscription or allocation-based costs when free (pre-paid) access is not available. Cost models will be developed with WP12.

Partners: NL, CZ, FR, EMBL-EBI, FI

Subtask 4.3.3: Storage and data transfer (22PM)

Data push and pull is needed in WP6 to 9 supported with commonly agreed technical tools and interfaces. Various transport mechanisms (e.g. GridFTP, http, Aspera, UDPipe, iRods) can be used to move the data to or from Data Resources (WP3). The managed access integration will be piloted in the ELIXIR-EXCELERATE Use Case WP9. Collaboration with GÉANT (e.g. bandwidth-on services) could be used to provide dedicated network links (e.g. lightpaths) for regular or large data transfer activities between the Nodes. Three common uses will drive WP4 storage and data transfer activities:

- (a) Data replication (an updated dataset being moved to multiple remote locations) and data submission (where a dataset is made available for subsequent retrieval and remote analysis). In the former the data source triggers data movement to data sink(s) (e.g. using Globus Transfer) using a replication policy around the data and updates any relevant data catalogues (e.g. B2FIND).
- (b) Service to pull relevant datasets for detailed analysis (e.g. Galaxy running on ELIXIR-affiliated cloud resource during training event). The retrieved dataset may be discarded after processing and just the results are retained based on the assumption that the original data will remain accessible for re-analysis.
- (c) Data location services will be used to manage and discover data replicas within ELIXIR sites (using technologies such as B2FIND or the EGI Data Catalogue). AAI mechanisms and workflows (e.g. REMS) are needed for gaining approved access entitlements in collaboration with the responsible granting bodies such as data access committees (e.g. EGA).

Partners: SE, EMBL-EBI, ES, CZ, FI

Subtask 4.3.4: Service Registry (10PM)

Integrate with WP1 and WP3 service registry and existing e-Infrastructure registries to enable a wide range of ELIXIR services and resources (e.g. cloud, storage, datasets) so that they become discoverable entities. The service registry provides a ‘gateway’ by which service providers can advertise and the users consume services. The service registry needs to provide a ‘service discovery’ function for consumers, but also written advice and requirements on how service providers can advertise their services.

Partners: EMBL-EBI, CZ

Deliverables

D4.1: ELIXIR Technical Services Roadmap: Incorporates the technical requirements coming from the Use Cases, details of the service offerings coming from the ELIXIR Nodes and the public e-Infrastructures, and provides a roadmap detailing when new services will appear within the ELIXIR Technical Services registry and the work being undertaken by the service providers to meet ELIXIR’s requirements (**M11**)

D4.2: Updated ELIXIR Technical Services Roadmap: Incorporates the technical requirements coming from the Use Cases, details of the service offerings coming from the ELIXIR Nodes and the public e-Infrastructures, and provides a roadmap detailing when new services will appear within the ELIXIR Technical Services portfolio and the work being undertaken by the service providers to meet ELIXIR’s requirements. (**M23**)

D4.3: ELIXIR Technical Services Roadmap: Incorporates the technical requirements coming from the Use Cases, details of the service offerings coming from the ELIXIR Nodes and the public e-Infrastructures, and provides a roadmap detailing when new services will appear within the ELIXIR Technical Services portfolio and the work being undertaken by the service providers to meet ELIXIR’s requirements. (**M35**)

D4.4: ELIXIR Technical Services document: Describes the ELIXIR Technical Services being offered through the cooperating service providers (ELIXIR Nodes and public e-infrastructures). (**M46**)

Work Package number	5	Start Date or Starting Event	M1
Work Package title	The ELIXIR Interoperability Backbone		

Participant number	1	2	3	4	6	10	12	18
Short name of participant	EBI	UOXF	TGAC	UNIMAN	NBIC	IRB	BSC	INESC-ID
Node	EMBL-EBI	UK	UK	UK	NL	ES	ES	PT
Person-months per participant:	12	51.4	10.6	30	38	21	19	3
Participant number	25	27	31	32	33	35	39	
Short name of participant	UiO	CNRS	CNR	LIU	UL	UOCHB	DTU	
Node	NO	FR	IT	SE	SI	CZ	DK	
Person-months per participant:	4	10	2	32	6	3.2	10	

Objectives

The objectives for WP5 are:

- Define agreements on identifiers and machine processable (meta)data descriptions with data providers (WP3, WP6 to 9), for (i) data repositories and (ii) knowledge bases, using established technical and domain standards, working with global initiatives to ensure broader interoperability.
- Consolidate existing Interoperability Services to support: (i) machine processable identity, data formats, experimental reporting guidelines, knowledge representations, and (ii) resource operational practices for transparent releases, versioning, provenance, updates.
- Implement data interoperability between prioritized resources, in priority areas.
- Run a programme of Bring Your Own Data (BYOD) “bootcamps” and coordinate with WP11. Build capacity with WP10: Data Nodes Network.

Work Package Leads: Barend Mons (NL) and Carole Goble (UK)

Description of work

Optimal Interoperability is attained when data access and use can be completely automated: programming and interfaces conform to standards that specify consistent syntax and formats; and data are associated with metadata and terminology identifiers and codes that support computational aggregation and comparison of information that reside in separate resources. As an exemplar implementation of this principle, The Interoperability Backbone in WP5 is data and use case driven, implementing FAIR principles⁵⁷, working in partnership with the custodians of the datasets. We will support cross-resource questions, for instance bridging genomic and phenotypic data for variant identification, using machine processable (RDF/XML) representations of the metadata. We use this as a reference for building Node capacity in skills and knowledge for data interoperability and access to technical infrastructure that supports data interoperability. For example: Rare disease data described with the ORDO (Orphanet Rare Disease ontology) to integrate rare and common disease (WP8); Different plant phenotypic data described with a variety of specialist and overlapping terms (e.g. Plant Ontology, Crop Ontology, Plant Trait Ontology) with sample environment descriptors drawn from the Environment Ontology (EnvO) and the eXtensible Experiment Markup Language (XEML) (WP7); Marine metagenomics data described using EnvO, XEML, plus descriptions to screen environmental metagenomic sequence datasets (e.g. FOAM), and the Genomic Standards Consortium’s reporting checklists and provenance aspects (such as environment from which samples originate).

For the data resource and service management of Named and Core Resources (WP3) we emphasize self-described datasets and explicitly described and published life-cycle metadata, using machine processable representations and common APIs for accessing it. The approach taken here will be an iterative one, working on tasks in parallel, to ensure that continuous but stepwise improvements for interoperability are yielded, and forged in practice against expressed need. We will not attempt comprehensive perfection: instead we aim for a principled “Just Enough and Just in Time” interoperability “on demand”, while raising the bar on general data service quality along with WP1. No standard is committed to without an example implementation.

Task 5.1: FAIR Principle Interoperability Implementation Agreements (62PM)

⁵⁷ FORCE11 - The FAIR data Guiding Principles: <https://www.force11.org/group/fairgroup/fairprinciples>

Many of the necessary services already exist across ELIXIR but need alignment and support. This task is aimed at identifying critical services, defining the interfaces between them, and developing a long-term integration strategy. By doing so, this will survey national (Node) practices and needs. Work is based on existing ELIXIR infrastructure and community emergent conventions and the forging of partners in and outside Europe. The WP partners have established track records of metadata management for biology datasets or metadata mechanisms, notably using Linked Data (e.g. Semantic Web) technologies for publishing self-described data (e.g. RDF), resolving identifiers (e.g. URIs) and defining ontologies.

For datasets we will agree:

- Practices of data management and data publishing; Managed APIs and message formats, with agreed APIs for access to dataset descriptors. We tackle data repositories and biological knowledge bases differently, reflecting their different content and their content lifecycles.
- Common exchange formats: such as RDF and XML data schemes.
- Common reporting guidelines: submission, curation and validation tools using data templates (example: ISATAB and tools), focusing on interoperability of standards via common data element mappings.
- Common terminologies: general (e.g. VOID, PROV for provenance); data type/community specific (examples: ORDO, Plant Trait Ontology); cross-community common elements (example: EnvO environment descriptors).
- Common APIs: for common data types.
- Best practices for publishing data as Linked Data, leveraging the EMBL-EBI's RDF Platform, LinkedISA, and resources of other platforms (example: IMI Open PHACTS Discovery Platform), as a semantic interoperability platform in addition to the use of APIs.

For biological knowledge bases the task will concentrate on common conventions for:

- Descriptions using common terminology, standard data formats, and mappings between common data elements and standard ontologies.
- Descriptions of the dependencies, curation and computational processes used to generate the current record for the biological entity, where appropriate.
- Good practices for publishing data as Linked Data, leveraging the EMBL-EBI's RDF Platform⁵⁸ and resources of other platforms (example: IMI Open PHACTS Discovery Platform⁵⁹), as a semantic interoperability platform in addition to the use of APIs.

Partners: all partners will be involved in sprints, focussing on specific dataset combinations to achieve WP6 to 9 questions. Managers/Core: UK, NL, EMBL-EBI, SE

Subtask 5.1.1: Use Cases (WP6 to 9) (26PM)

Outcome: Interoperability, common APIs and descriptors workable in the field. First on WP7 (Genomic and Phenotypic Data for Crop and Forest Plants) and WP8 (Rare Disease) followed by WP6 (Marine metagenomics). Jointly identify data-driven interoperability examples, cross-dataset questions, bottlenecks and common descriptions. Survey national (Node) practices and needs.

Subtask 5.1.2 Core and Named Resources (WP3) (18PM)

Outcome: Common APIs and dataset descriptors workable in the field. Focus on interoperability at the dataset level. Workshops to co-produce common APIs and dataset descriptors leveraging proposed standards (example: W3C HCLS Dataset descriptor). Contribute to dataset metrics/quality criteria for data service life-cycle management in WP3. Survey national (Node) practices and needs.

Subtask 5.1.3 Global engagement: international organisations and multilateral forums (18PM)

Outcome: Establish a leading role for ELIXIR internationally on this aspect and compatibility with other international interoperability initiatives. Engagement with European initiatives (examples: IMIs, RIIs, EUDAT EUON), global initiatives (examples: NIH BD2K, GA4GH, Force11), commodity standards (examples: W3C,

⁵⁸ EMBL-EBI RDF Platform: <http://www.ebi.ac.uk/rdf/platform>

⁵⁹ Open PHACTS: <http://www.openphacts.org/>

DataCite, ORCID, VIVO, ORE) and community standards (example: RDA). Feedback between external forces and ELIXIR resources, maintaining ELIXIR visibility in key meetings with interoperability initiatives.

Task 5.2 FAIR Interoperability Implementation Services (65PM)

Integrating complex datasets requires services to handle identifiers for data and biological concepts (phenotypes, diseases), and tools that allow users to map data between different sources and help users find and apply standards. Many of these tools exist. This task brings these services together.

Partners: UK, NL, EMBL-EBI, SE, FR

Subtask 5.2.1: Identity Management, Mapping and Tracking services (14PM)

Outcome: Make explicit the scope and limitations of identifiers, the mappings between identifiers for entities (example: Ensembl gene identifiers can be mapped to Uniprot identifiers) and provide identifier services used by data resources in the field. Work includes: identity authorities for specific data types and concept categories; identity resolution, identity mapping, and entity resolution.

Subtask 5.2.2: Reporting Guidelines, Formats, Controlled Vocabulary Services (10PM)

Outcome: The best of breed services assembled, organized into a coherent tool suite, and used in practice in WP3 and WP6 to 9. Workshops of service providers and users to “bake-off” alternatives and integrate complementary services.

Subtask 5.2.3: Dataset publishing for API interoperability (13PM)

Outcome: Common and standardised practices to dataset lifecycle management and release management, including: Distributed revision control (example: GIT), dependency management; and well described, validated and maintained APIs registered in catalogues (example: BioCatalogue, BioSharing).

Subtask 5.2.4: Biological knowledgebase publishing for Linked Data interoperability (14PM)

Outcome: Data published as Interoperable Linked Data for some biological knowledge bases. We will develop services for the creation and management of mappings, as first class artefacts, between data entities to describe the curation and computational processes used to generate the current record for the biological entity (leveraging work at EBI/SIB, DTL, SciLifeLabs and the UK).

Subtask 5.2.5: Sustainability of Interoperability Implementation Services (14PM)

Outcome: A strategy for sustaining key services.

Task 5.3: “Bring Your Own Data” (BYOD) & Capacity Building Workshops (125.2PM)

Implement data interoperability between resources for WP6 to 9 and data publishing of resources for WP3. Provide practical support and guidance through a programme of Bring Your Own Data bootcamps. BYODs are a mix of tutorials and hands on, practical “hackathons” with specific datasets: database custodians’ work with experts in semantic web and linked data technologies to make their data available in a FAIR way using machine processable (meta) data and APIs. During the BYOD Task 5.1’s specifications, standards and data templates are refined in practice and Task 5.2’s tools and services are exercised.

Partners: NL, UK, ,FR, PT, SI, SE, CZ, IT, ES

Subtask 5.3.1: Manage and Run BYOD Workshops (65PM)

Outcome: Organise and manage workshops synergistically with WP3 and WP6 to 9 work plans, as well as aligned with WP11 training activities. Monitor workshops and feedback to Tasks 5.1 and 5.2. After the first 6 months of the project we anticipate a BYOD every 3 months.

Subtask 5.3.2: Create and manage BYOD training materials (24.2PM)

Outcome: Materials for BYOD, updating in step with Tasks 5.1 and 5.2. Deploy BYOD materials on the WP11 TeSS Portal. Develop BYOD materials for Data Carpentry training (WP11).

Subtask 5.3.3: Data Node Capacity Building (36PM)

Outcome: Build capacity in the Data Nodes (WP10, Task 10.2) using 5.3 as a reference example for interoperability practice and the BYOD methodology rolled out across the Nodes. Have capability to independently run BYOD's for new or national datasets in one third of the Nodes.

Deliverables

D5.1: *Interoperability Implementation Regulations.* A set of standards, rules, controlled vocabularies, authorized unique identifiers and interoperable service APIs for the data repositories and biological knowledge bases agreed and implemented in the field with WP6 to 9, and WP3; resources registered in BioSharing; BioCatalogue and BioRegistry (WP1); ontologies accessible via Ontology Lookup Services. (**M18**)

D5.2: *Sustainability business cases* for Interoperability Implementation Services, including contractual arrangements with non-ELIXIR resources. Data stewardship agreements with major data interoperability activities in Europe [EGA, EUDAT etc.] USA [DataOne, BD2K] and globally [RDA, GA4GH, Force11]. (**M36**)

D5.3: *Bring Your Own Data* 8 delivered workshops and established roll-out programme across WP10 Data Node capability network; materials deposited on WP11 TeSS. (**M24**)

Work Package number	6	Start Date or Starting Event		M1	
Work Package title	Use Case A: Marine metagenomic infrastructure as driver for research and industrial innovation				
Participant number	1	17	20	24	27
Short name of participant	EBI	FCG	CCMAR	UiT	CNRS
Node	EMBL-EBI	PT	PT	NO	FR
Person-months per participant:	28	2	11	36	10
					10

Objectives

The main objective for this Use Case is to develop a sustainable metagenomics infrastructure to enhance research and industrial innovation within the marine domain before M36 of the ELIXIR-EXCELERATE project. The main objective will be achieved by the following specific objectives:

- Development and implementation of selected standards for the marine domain. (Task 6.1)
- Development and implementation of databases specific for the marine metagenomics. (Task 6.2)
- Evaluation and implementation of tools and pipelines for metagenomics analysis. (Task 6.3)
- Development of a search engine for interrogation of marine metagenomics datasets and establish training workshops for end users. (Task 6.4)

Work Package Leads: Nils P Willassen (NO) and Rob Finn (EBI)

Description of work

Metagenomics has the potential to provide unprecedented insight into the structure and function of heterogeneous communities of microorganisms and their vast biodiversity. Microbial communities affect human and animal health and are critical components of all terrestrial and aquatic ecosystems. They can be exploited *e.g.* to identify novel biocatalysts for production of fuels or chemicals (bioprospecting), make functional feed for aquaculture species, and for environmental monitoring. However, in order to expand the potential further for the research community and biotech industry, especially within the marine domain, the metagenomics methodologies need to overcome a number of challenges related to standardization, development of relevant databases and bioinformatics tools. New and emerging sequencing technologies, integration of metadata gives an extra burden to the development of future databases and tools. The Use Case “*Marine metagenomic infrastructure as driver for research and industrial innovation*” will contribute to the overall objectives of the ELIXIR-EXCELERATE project by developing research infrastructure and service provision specific for the marine domain in order to enable metagenomic approaches responding to societal and industrial needs. The outcome of the proposed Use Case will meet the major needs expressed by the marine domain (*e.g.* ESF Marine board Position Paper 17 “*Marine Microbial Diversity and its*

role in Ecosystem Functioning and Environmental Change” and Position Paper 15 “Marine Biotechnology: A New Vision and Strategy for Europe”).

Task 6.1: Development and implementation of a comprehensive metagenomics data standards environment for the marine domain (12 PM)

To maximise the impact and long term utility and discoverability of metagenomics datasets, it is essential the experimental methods and data acquisition/storage protocols be established. In Task 6.1, we will bring together a comprehensive metagenomics data standards environment in collaboration with marine experimental scientists, data providers, end users and the existing communities involved in marine standards development. The environment will bring together three components:

- **Data format conventions and standards** will address the various data types for which sharing is required, that will include contextual data (e.g. sample information, expedition-related data), metadata (e.g. provenance and tracking information, descriptions of experimental configurations and bioinformatics tools in use) and data (e.g. raw sequence data, aligned reads, taxonomic identifications, gene calls).
- **Reporting standards** will address community-accepted thresholds for richness/precision that are required to make data useful, including depth of raw machine data, such as resolution of sequence quality scoring, conventions for references to reference assemblies and minimal reporting requirements for contextual data.
- **Validation tools** will address the automated validation of compliance with conventions and standards and the meeting of minimal reporting expectations for given datasets in preparation by the marine research community. In this task, we will bring together components that exist already – in particular the contextual data and metadata reporting standards we have developed under the Micro B3 project (EU FP7), data standards and conventions developed around our European Nucleotide Archive (ENA) programme, such as CRAM, FASTQ conventions, work existing in the biodiversity and molecular ecology domains (such as tabular data conventions and BIOM matrices) – and construct new components as required. The major output of this work will be a set of well described and navigable elements to aid the marine community in the preparation, sharing, dissemination and publication of highly interoperable and comprehensive metagenomics datasets.

Partners: EMBL-EBI, NO

Task 6.2. Establishment of marine specific data resources (20PM)

Due to the data biases of existing reference databases, only about one quarter of sequences are annotated, and this fraction diminishes further when more diverse samples such as soil and marine are analysed. To improve the characterisation of marine metagenomic samples, this task involves the construction of sustainable public data resources for the marine microbial domain. Task 6.2 will be achieved by establishing marine microbial databases including reference genomes, nucleotide and protein databases. The established databases, based on the standards developed in Task 6.1, will enhance the precision and accuracy of biodiversity and function analysis. The reference databases will be non-redundant datasets generated from sequences acquired from ENA (as part of the International Nucleotide Sequence Database Collaboration), UniProt and other publicly available datasets. In particularly, we will use some of the higher-coverage and higher quality sequence outputs from the TaraOceans and Ocean Sampling Day metagenomic projects, to build high quality marine specific reference databases. All datasets will be checked with respect to quality, consistency, and interoperability, and in compliance with standards developed in Task 6.1. The respective knowledge-enhanced databases will be the cornerstone for sustainable analysis of marine metagenomics sequence data. The databases will be developed in collaboration with members of the ESFRI infrastructures EMBRC and MIRRI and made publicly available through ELIXIR.

Partners: NO, EMBL-EBI, IT

Task 6.3: Gold-standards for metagenomics analysis (58PM)

The majority of existing metagenomics analysis platforms, while providing insights into the prokaryotic taxonomic diversity and functional potential for individual samples, but lack the tools that enable discoverability across samples and industrial innovation. This task will focus on the evaluation and implementation of new tools and pipelines in order to accelerate research, discoverability and innovation, reducing time to market for new products. In combination with new standards and databases developed in Task 6.1 and Task 6.2, respectively, new tools for community structure (microbial biodiversity), genetic and functional potential will be evaluated and implemented for environmental applications. For industrial application tools and pipelines for the identification of gene products (e.g. enzymes and drug targets) and pathways will be implemented and made publicly available. The evaluation and implementation will be performed in near collaboration with end-users (research groups, environmental centres, biotech companies) to ensure usability for the end user community in order to improve

quality, productivity and functionality, as well as reduction of costs for the end-users. New tools and pipelines will be made publicly available through the e.g. META-pipe (ELIXIR-NO), EBI Metagenomics Portal (EMBL-ELIXIR) and/or EMBL Embassy cloud technology. Technical requirements will be mapped by WP3 and implemented to meet the requirements of the ELIXIR community.

The continued advancement of sequencing technologies and the growing number of public marine metagenomics projects means that it is becoming increasingly difficult to mine these vast datasets. In this task, initially a web-based search engine will be developed for the interrogation of marine metagenomics results available from the EBI Metagenomics Portal, based on combinations of queries to our web services (already in existence, or to be built as part of existing projects outside ELIXIR-EXCELERATE) for the discovery of data through metadata, taxonomic and functional fields. This will extend the back-end search functionality that is to be developed as part of on-going efforts. In addition to being downloadable, we will enable search results to flow into an expanded comparison tool (currently limited to gene ontology terms from samples in the same project), to allow more in-depth analysis of a user selected datasets, allowing functional and taxonomic comparisons.

In the second phase of this task, the search engine will build upon the data exchange formats in Task 6.1, and federate the search across different pipeline results sets (e.g. META-pipe), so that different results based on the same underlying dataset, can be amalgamated into a single search. This will dramatically enhance the discoverability across different marine datasets, allowing the identification of common trends and/or differences. These tools will be developed using user-experience testing and in collaboration with end users to ensure they are fit for purpose.

Partners: NO, EMBL-EBI, IT, FR, PT

Task 6.4: Training workshops for end users (7PM)

In this task training workshops will be established, in collaboration with WP11 “ELIXIR Training Programme”, for end-users with the aim to facilitate accessibility, by training European researchers and industry to more effectively exploit the data, tools and pipelines, and compute infrastructure provided by the ELIXIR marine metagenomics infrastructure. These training workshops and materials will be converted to online training resources, extending the reach of the workshop.

Partners: PT, NO

Deliverables

D6.1: Specific marine databases made publicly available. (**M18**)

D6.2: Report on comprehensive metagenomic data standards environment. (**M30**)

D6.3: Report describing a set of tools, pipelines and search engine for interrogation of marine metagenomic. (**M36**)

Work Package number	7	Start Date or Starting Event					M1
Work Package title	Use Case B: Integrating Genomic and Phenotypic Data for Crop and Forest Plants						
Participant number	1	3	6	17	19	21	28
Short name of participant	EBI	TGAC	NBIC	FCG	IBET	CSC	INRA
Node	EMBL-EBI	UK	NL	PT	PT	FI	FR
Person-months per participant:	12	21.2	9	2	48	11	24.7
Participant number	29	34	38				
Short name of participant	CIRAD	NIB	VIB				
Node	FR	SI	BE				
Person-months per participant:	12	30	11				

Objectives

The main objective of WP7 is to design and test an infrastructure to allow genotype-phenotype analysis for crop plants based on the widest available public datasets. To achieve this, the specific objectives for WP7 are to:

- Make data interoperable (in accordance with the ‘FAIR’ principles specified in WP5) through the development of controlled vocabularies and standardised APIs, proving the concept of a common phenotypic API through which any participant in an open network can advertise the availability of their data in a common

domain.

- Annotate and submit key exemplar datasets to relevant public archives.
- Engage industry in defining priorities in genotype/phenotype annotations, and collaborate with WP13 in showcasing the developed resources to the agroforestry industry.
- Collaborate with WP11 in delivering specific training for the use of developed resources.

Work Package Leads: Paul Kersey (UK) and José Leal (PT)

Description of work

This work will facilitate the analysis of many of phenotypes against large panels of crop accessions through the aggregation of locally held data; and thereby, enable more powerful association analysis, opening the way to understanding of function, candidate gene prioritisation, and improved crop breeding. Working on exemplar species, we will establish a sustainable model for the interaction of distributed phenotypic repositories with defined genomic and sample reference data, in which organisations can expose data to the system through conformity with standards for annotation and interface, allowing the subsequent expansion of the approach to other species and domains. It will also provide resources (in the form of standards, ontologies and models for annotation and collaboration) for use within ongoing species-centric (e.g. the Wheat Initiative) and/or national endeavours.

Massive sequencing and genotyping of crop and forest plants (and their pathogens and pests) is generating large quantities of genomic variation data. These efforts are likely to accelerate in the near future, with further expected reductions in the cost of sequencing and international efforts (such as the DivSeek Initiative) aiming to catalogue all genetic diversity present in global germplasm resources. Such data could serve as a powerful panel in association screens and facilitate precision breeding of increasingly complex traits. But structural variation in most crop plants is enormous (more so than in humans), and phenotypic characterisation data is (i) often inaccessible (ii) diverse and non-standard (iii) lacks any route of unified access. Indeed, “phenotype” is a broad concept, covering many data types (descriptive data, molecular data, image data) at many scales (laboratory, phenotyping centre, field data) on many species; and moreover, a phenotype exists in the specific concept of an experiment (in contrast to a genotype, which is assumed to be constant in a given sample). Both technical and sociological progress in data definition and sharing are lagging experimental progress.

To address this problem, we will harness the domain-specific expertise and data held in a distributed fashion across many national Nodes with interests in agriculture. Seven ELIXIR Nodes will jointly establish a technical infrastructure and associated social practices to define an open model for the publication and sharing of plant genotype-phenotype data, working on a minimum of 3 exemplar species from different domains of plant life to establish a model applicable in all species. We will establish a scalable, distributed model, transparently integrated through the development and use of common vocabularies and search technologies, adhering to the FAIR principles⁶⁰ (WP5), and using established repositories for genomic data and sample meta data. Domain-specific training will be coordinated with WP11 and will include training users and data curators. The expected impact is accelerated research and plant breeding through the exploitation of an interoperable commons of public data.

Task 7.1: Development/adoption of appropriate controlled vocabularies for annotating plant phenotypic data (45PM)

The use of controlled vocabularies, to define the material assayed, the form of the experiment and the observed phenotype are critical to enabling diverse datasets to be interrogated and compared. A number of initiatives have developed controlled vocabularies that can be used for the annotation of plant phenotypic data, including the Plant Ontology (<http://www.plantontology.org>), Crop Ontology <http://www.cropontology.org>), Plant Trait Ontology (http://www.obofoundry.org/cgi-bin/detail.cgi?id=plant_trait), Environment Ontology (<http://environmentontology.org>), XEML (<http://xeml.codeplex.com>). Different vocabularies apply in different species, with some specific and some overlapping features; in many countries, national lists of traits found in specific crop varieties are in use by breeders (distinct from the terms in use in academia). Slow-growing forest species have overlapping needs to annual food crops.

Together with representatives from the appropriate infrastructure resources, communities and ontology developers, and industrial/SME partners, we will work on establishing common guidelines for ontology usage when annotating crop and forest species. We will hold a workshop within the first 4 months of the project bringing experts together from all participating Nodes and key industrial participants, working on the target species to agree on a common set

⁶⁰ FORCE11 - The FAIR data Guiding Principles: <https://www.force11.org/group/fairgroup>

of vocabularies to be used in the project (by the end of the month 6). Existing ontologies will be extended where necessary, and cross-references established between corresponding high-level terms in the recommended vocabularies (e.g. between common anatomical concepts in different species-specific ontologies). Outputs will be regularly showcased to industry and list reconsidered at 6 monthly intervals.

Partners: EMBL-EBI, BE, FR, FI, NL, PT, SI, UK

Task 7.2: Annotation of key plant phenotypic datasets with agreed controlled vocabularies (51 PM)

At least 3 exemplar species will be chosen, including one cereal species, one woody species, and one other crop species, each of which is of interest to at least 2 ELIXIR partners (maize, oak and potato have been identified as initial candidates). Phenotypic data is very varied and we will work on a variety of data types, including quantitative trait loci, association genetics (laboratory and field data), and biotic stress. Participating Nodes will collect and curate sample and experimental meta data and phenotypic description data to agreed standards using the vocabularies identified and extended in Task 7.1.

Partners: BE, FR, NL, PT, SI, UK

Task 7.3: Submission of exemplar genomic and phenotypic datasets to appropriate public repositories (33.9PM)

Annotated datasets will be submitted to appropriate repositories wherever possible, using existing platforms designed for such scope; for example, the European Nucleotide Archive (developed by EMBL-EBI) (for genomic and transcriptomic data), or phenotypic archives e.g. PIPPA (the PSB Interface for Plant Phenotypic Analysis, in development at VIB), BreeDB (in development at NBIC), GoMapMan (in development at NIB), and Ephesis (in development at INRA). Sample identification will be handled through the BioSample DB at EMBL-EBI, or, where the sample is an accession from a public gene bank, by cross-references to EURISCO, the European catalogue of plant collection data (<http://eurisco.ipk-gatersleben.de>). We will work closely with EURISCO and the gene banks to ensure that appropriate meta data is reliably, non-redundantly associated with samples, and that genomic and phenotypic data held in different resources but derived from the same biological material can be reliably identified.

Partners: EMBL-EBI, BE, FR, NL, PT, SI, UK

Task 7.4: Development and implementation of agreed public APIs for access to data in participating repositories and exposure via public computational infrastructures (51PM)

We will develop a common API for data query and retrieval, in close collaboration with WP5. We will build from the experiences already gained by partners in developing repositories and interfaces and will utilise established standards for programmatic data access (e.g. REST, RDF, etc.). The API will be implemented at each partner holding a genotypic, phenotypic or sample repository, allowing users to query a single end point that will return results meeting a common standard from dispersed resources. The API will be exposed to users via the ELIXIR computational infrastructure and other computational infrastructures in the plant sciences (for example, on the iPLANT infrastructure in the United States).

The first version of the API will support identification and query of datasets held in distributed repositories matching specified experimental and sample meta data. The API will be later be extended to encompass the querying of resources based on the phenotypic descriptions.

We would also like to enhance the interpretability of results and lower the barrier of computer competency required by users to access data queries from this platform. We will use the data served by the API to develop intuitive visualisation components to allow researchers to mine phenotypic data accessible through the API. These components will be developed within existing frameworks e.g., PIPPA or BioJS (a JavaScript library of open source components for biological visualisation), allowing their re-use in other contexts.

Partners: EMBL-EBI, BE, FR, NL, PT, SI, UK

Relation to other WPs

Propose phenotype resources to service registry (WP3)

Exposure on the ELIXIR cloud resources (WP4)

Adhering to and develop vocabularies and API compatible to the FAIR principles (WP5)

Training in the development of plant ontology development and in the use of resources (WP11)

Deliverables

D7.1: Datasets annotated for at least 1 woody plant, cereal and solanaceous species; genotype, phenotype and

sample meta data submitted to appropriate public archives. (M24)

D7.2: Phenotypic data exposed through integrating API by all participants. (M36)

D7.3: New visualisation components for phenotype visualisation released in BioJS. (M36)

Work Package number	8	Start Date or Starting Event					M1
Work Package title	Use Case C: ELIXIR infrastructure for Rare Disease research						
Participant number	4	6	7	10	12	15	23
Short name of participant	UNIMAN	NBIC	CNIO	IRB	BSC	CNAG	NTNU
Node	UK	NL	ES	ES	ES	ES	NO
Person-months per participant:	6	6	10	13	10	38.4	12
Participant number	27	31	33	39			
Short name of participant	CNRS	CNR	UL	DTU			
Node	FR	IT	SI	DK			
Person-months per participant:	12	3	15	12			

Objectives

WP8 aims to empower actors involved in the development of new rare diseases therapies through the execution of the following specific objectives:

- Build the ELIXIR registry of data resources and analysis tools critical for the development of the rare disease research. (Task 8.1)
 - Continuous monitoring of resources and tools in Rare-diseases.
 - Implementation of a system for the generation of datasets adequate for the assessment of methods in the area of rare-diseases.
 - Implementation of the ELIXIR rare-disease portfolio in the ELIXIR registry.
- Implementation of a technical framework for the comparison and standardization of services useful for the rare-disease communities. (Task 8.2)
- Collaboration with the rare-disease communities for the organization of training courses, workshops and jamborees. (Task 8.3)

Work Package Lead: Ivo Gut (ES) and Marco Roos (NL)

Description of work

The International Rare Diseases Research Consortium (<http://www.irdirc.org>) established the ambitious goal of developing 200 new therapies by 2020. ELIXIR as a whole and in particular this Work Package is aligned with this effort. The overall objective of this Work Package (WP) is to address the needs of the rare diseases community through the instantiation of the ELIXIR resources described in WP1-5. These resources do not constitute a replacement of the current research projects organized around the rare diseases area. Indeed the aim is to empower them and to help in the sustainability of the resources created by these projects in the long term.

This WP is organised around the actors that play a major role on the development of these new therapies. These actors are the main users of the ELIXIR infrastructure: data generators and curators (usually personnel working in hospitals, genomics-based companies, and members of large research consortia), researchers (bioinformaticians, geneticists, and clinical doctors), diagnosis companies, CROs (usually SMEs), and the pharmaceutical industry among others

Task 8.1: The ELIXIR portfolio of data resources developed in collaboration with the rare diseases communities. (69.4PM)

Subtask 8.1.1 Monitoring of resources and tools. (25.4PM)

There is a wide range of data resources and analysis methods used in the rare-disease area. Many of those resources are provided by ELIXIR Nodes, for example the European Genome-Phenome archive (EGA) currently stores data from major research initiatives in rare diseases like the RD-connect project. In this subtask we will review the

current data resources and evaluate their usability and potential impact on the rare disease community. An important aspect of the evaluation will be the security of the data that is a key aspect in rare disease domain given the low frequency of the associated genomic variants in the population.

One critical aspect of the development of the registry is to engage the different communities in the submission and rating of the tools. In this task we will work together with representatives of the major projects in the field of rare-diseases to create a customized portfolio of ELIXIR tools and services devoted to assist them in the development of these new therapies. As an example we will ask for proposals of tools that serve to interpret the effect of genomics variants on a group of patients that belong to the same family. We strongly believe that this link between the end-users and the tools developers will help ELIXIR to understand better the problems that are actually facing the main actors in the rare diseases research and hence to better solutions.

The final outcome of this task will be the ELIXIR data resources and analysis tools useful to the rare disease communities.

Partners: NO, ES, SI, IT, NL

Subtask 8.1.2: Creation of reference datasets adequate for the specific assessment of methods and standards in the area of rare-diseases. (30PM)

While the creation of these tools should stay as a priority for researchers, large scale projects, SMEs and the industry increasingly need access to benchmarked methods on which to build their analysis strategies.

The evaluation of the methods requires the adequate selection of the datasets and benchmarking strategies. The systems for the selection of the datasets for the benchmarking have to be fast and effective to enable the continuous evaluation of the methods, as described in WP2. We will collaborate with the ELIXIR benchmarking strategy (WP2) to build the appropriate strategies for the selection of the datasets (subtask 8.1.1 above) and with the rare-disease communities to implement the adequate quality reporting standards. Moreover we will integrate these pipelines in the ELIXIR benchmarking framework (WP2) to continuously monitor the selected methods with the newly generated datasets.

Partners: ES, DK, IT, FR, SI, UK

Subtask 8.1.3 Implementation of the ELIXIR rare-disease portfolio in the ELIXIR registry. (14PM)

The ELIXIR registry will be a reference for the research community (WP1), as it will reflect the quality and the real-time status of the services included on it. This registry will act as a one-stop shop for services provided by ELIXIR. The goal is to allow users from the different countries, communities and projects to discover which are the tools available at a given time, with the associated information about the community based rating (see WP2), instructions for correct use and associated examples. We will encourage tools developers to adopt the EDAM standard to describe their tools and to share several metrics about the performance and usage of these of the tools (see description in WP1).

Those services promoted as relevant by the end-users will be listed in a special section in the ELIXIR registry.

Partners: DK, ES, FR.

Task 8.2: Standardisation of rare disease services in collaboration with the RD communities. (36PM)

The ecosystem of RD services will inevitably be a combination of distributed and centralized resources, because of the sheer number of rare diseases and rare disease organisations, as well as legal and ethical constraints between countries and communities. At the same time, because of the low frequency in the population, combining data across patient registries, biobanks, and -omics databases is the single most important way of getting new insights towards new treatments.

One of the most recurrent issues when attempting to perform research across resources is the lack of standards or the poor adoption of existing standards by RD stakeholders. Rare disease standards concern different types of data including genomic and phenotypic characteristics, causative genetic variation status, quality criteria, analysis protocols, supporting evidence and follow-up indicators. These problems will be analysed in workshops including experts in semantic web, linked data technologies and rare-disease experts (see previous experiences and proposal in “Bring Your Own Data (BYOD) bootcamps”, in WP5). The initial experience with this methodology (see⁶¹) is

⁶¹ Linked Data BYOD for Rare Disease registries, biobanks, and Orphanet, organised by RD-Connect in collaboration with

that a critical bottleneck is the identification of the most appropriate terms and identifiers to annotate data for cross-resource questions. Based on this experience we aim to address two major 'white spots' in the available infrastructure for Rare-diseases: (i) the current infrastructure of the rare disease platform: RD-Connect, does not contain backbone services for functional interlinking, (ii) a majority of RD sources are not equipped to provide data, metadata, and data updates using appropriate standard procedures. To address these needs we will work together with WP5, the rare-disease communities and the RD-Connect project to (i) deploy and test the services and guidelines for standardization 'at the source', (ii) provide standardized interfaces that Rare-disease communities can work with from a central location, (iii) build capacity in the RD community by enabling them to work with these services themselves.

Partners: FR, ES, DK, NL.

Task 8.3: Training workshops targeting different user communities. (32PM)

In this task training workshops and courses will be delivered, in partnership with WP11 "EXCELERATE Training Programme". The training will be approached from two sides. First, in collaboration with the Train the Researcher task in WP11 we will train rare diseases' researchers in the use of relevant tools, standards and infrastructure produced by ELIXIR. Second, we will run "feedback workshops" in which those who are developing the methods will be exposed directly to problems faced by the rare disease community. These userthons will help to shape the ELIXIR portfolio. The direct collaboration with WP11 Train the Researcher will ensure that researchers are trained to a high standard in state-of-the-art analysis techniques for rare disease data and that innovative training approaches developed in this task are applied elsewhere in ELIXIR.

Partners: UK, SI, NL.

Deliverables

D8.1: Portfolio of ELIXIR data resources and tools for the rare diseases communities. (**M24**)

D8.2: Documentation of the tools for the data manipulation and standard conversions in the rare-disease field. (**M36**)

D8.3. Report on the ELIXIR workshop organized with the rare-disease communities. (**M12, M24**)

Work Package number	9	Start Date or Starting Event					M6
Work Package title	Use Case D: An ELIXIR framework for secure archiving, dissemination and analysis of human access-controlled data; enabling biobanks, cohorts and local resource services to leverage the EGA						
Participant number	1	5	6	8	21	25	32
Short name of participant	EBI	UTARTU	NBIC	CRG	CSC	UiO	LIU
Node	EMBL-EBI	EE	NL	ES	FI	NO	SE
Person-months per participant:	34	16	5	33	18	6	6

Objectives

This Work Package has three main objectives:

- **To upgrade and make more portable -omics data collection and submission tools utilizing the European Genome-phenome Archive (EGA)** as the core of an ELIXIR community secure data sharing network for -omics data. Tools developed here will support submission of all types of -omics data from human samples consented for biomedical research from disease consortia such as International Cancer Genome Consortium (ICGC), Rare Diseases (Rd-Connect), national cohorts, and biobanks. Emphasis is given to supporting investigator and locally driven research projects with human data consented for biomedical research. To enable these projects, the data submission tool chain will be made more portable and user-friendly with the goal of distributing a common toolset "in-a-box" to enable local and national groups to collect -omics data and meta data in a distributed manner which is consistent across European groups through ELIXIR coordination.

- **To enable value-added services at project specific, regional, or national resources** by establishing ELIXIR-wide community facing tools that allow local resource owners and developers to add value to their systems through data and metadata services from the EGA. For example, local research projects would be enabled to make their data discoverable and searchable, and linked with available -omics data from various sources, by leveraging stable unique EGA identifiers. Further, locally developed project specific data portals will be enabled through defined standard APIs using real time secure data links which allow -omics big data archived in the EGA to be presented in combination with biobanks or cohort data.
- **To extend and generalise the system of access authorization management and high volume secure data transfer developed in the EGA project** to address the secure data access needs across ELIXIR resources and open new modes of secure data access such as through public and private clouds. For example, a trusted ELIXIR Cloud service can receive local copies of selected datasets through a secure data mirroring system and provide access to data and compute to those users that already have data access permissions available from appropriate Data Access Committees stored in the EGA system.

The WP will partner first with 2-4 large resource owners to gain the required expertise, document the process in multiple ELIXIR member states and finally to propose a way to scale up these services to match wider European requirements. This WP will also be used to drive creation of the ELSI framework that supports the workflow (WP12).

Work Package Leads: Justin Paschall (EBI) and Arcadi Navarro (ES)

Description of work

This WP delivers the core ELIXIR workflow for long term archive and re-use of human data consented for biomedical research requiring access-control based on a data access agreement and approval process. The workflow supports data submitters and ELIXIR Node coordination on data deposition into the EGA archive in a manner that will maintain data ownership in the hands of the original research data owner, enable data release to authorised individual users from the archive and to partner with downstream secure ELIXIR data analysis platforms. This workflow ad supporting infrastructure will allow the data owners to focus on their unique areas of data generation and analysis expertise while being able to rely on EGA and the ELIXIR infrastructure for their common –omics big data storage, coordination and distribution needs under appropriate legal frameworks. The work described here will leverage the work of other ELIXIR-EXCELERATE Work Packages, for example WP10 to scale each service structure to cover all ELIXIR Nodes and with WP4 for technical service support, and relies on WP12 to establish the necessary legal framework that supports workflows.

The Workflow can be summarized as:

1. Data preparation, validation, and submission to the EGA making use of common supporting tools and data models (e.g. through Node data Network, WP10). Focus on providing software tools and remote APIs enabling local leadership and customisation within context of specific projects, supported by common ELIXIR coordinated tools and data models.
2. Bidirectional linking and secure data streaming between -omics data archived in EGA and local repositories or data portals that hold further information about the project and samples.
3. Management of user access-rights for release of archived data to authorized researchers under Data Access Agreements using ELIXIR tools, such as the REMS (WP4), that allow *resource* owners to manage data access rights.
4. Expanded access through ELIXIR partner secure clouds that can host EGA datasets, requiring the provision of metadata and authorization APIs
5. Data synchronization between the main EGA archive and authorized project specific resources and access points, such as compute clouds.

Task 9.1: Enhanced secure data submission tools. (40PM)

This task will update the existing EGA submission tools and documentation to facilitate large-scale data submissions operations, emphasizing local leadership and customization within a common framework.

Partners: ES, EMBL-EBI, FI

Subtask 9.1.1: Support for large scale submission of -omics data and sample metadata to the EGA. (25PM)

Support for large-scale submission of -omics data and sample metadata to the EGA through improved online tools, automated verification, and tools for the application of standard vocabularies to phenotype collection. These tools will make use of table “spread-sheet” based views of data for submitters less comfortable with technologies such as

XML. Further tools and reports supporting global EGA stable identifier mappings will allow easier integration with local identifiers, in support of federated global tracking of submitted samples and their derived -omics data.

Subtask 9.1.2: Portable submission toolkit. (15PM)

This task is composed of data format definitions and software components, a “mini-EGA in-a-box” will allow increased local control and coordination of data collection, and allow early validation of standardized data and metadata formats. This implementation provides the practical means for distributed projects to collect access-controlled human biomedical data in a manner that maintains a coordinated data model and dataset registry, enabling federated and a centralized single-point of discovery and access.

Task 9.2: Integrating centralized and distributed projects through transparent access to secure data: enabling local projects within a European wide framework. (40PM)

This task will enable local projects, such as study-specific data portals, local cohort resources, and national bioinformatics hubs by providing developer level APIs and services such that local efforts can efficiently build customized project branded solutions which make use of underlying ELIXIR and EGA tools and data archives.

Partners: ES, EMBL-EBI, FI, EE, NL, NO, SE

Subtask 9.2.1: Support secure integration of EGA data to downstream project client websites. (10PM)

Support secure integration of EGA data and metadata to downstream project client websites by providing new EGA programmatic interfaces that support standardized REST calls and provides results in ELIXIR endorsed formats (WP3 and WP6).

Subtask 9.2.2: Access management workflow support. (10PM)

Support access management workflows by data access committees through ELIXIR for EGA and other projects through developing applications of the Resource Entitlement Management Systems (REMS) expanding on an existing pilot project. This effort is focused on providing tools to delegate management to local projects and ELIXIR Nodes through new administrative roles.

Subtask 9.2.3: ELIXIR and EGA access integration. (20PM)

Specific efforts supporting controlled access -omics data infrastructure for use of partner national cohort studies in terms of submission, permissions management, and local and customized presentation of data under the cohort branding. Services will be tailored to respect the unique policy and data protection requirements of national cohorts, allowing single point of request and download from cohort branded web-pages. Support will be provided for distributed local hosting of datasets, within a common ELIXIR framework, where restrictions exist on the movement or hosting of data based on national borders.

Task 9.3: Federated authentication, large scale data management, and secure clouds in practice. (38PM)

This task is closely linked to the technically focused WP4 that provide the technical solutions required to deliver the outcomes of Task 9.1 and 9.2. In this task, technical components, including high volume secure data transfer and authentication and authorization management, are brought together to make -omics data from EGA and phenotypic data from cohort studies available for secure download, remote API access or from within public or private Cloud-based secure analysis environments. Cloud-based access to the EGA ecosystem provides a new access mode meeting a significant user need from research groups with limited local resources for compute and large-scale reference data storage.

Partners: EMBL-EBI, ES, FI, EE

Subtask 9.3.1: Large scale data mirroring support. (12PM)

Support for automated large scale data mirroring from the EGA archive to the authorized ELIXIR partner local services and cloud compute or HPC providers. This process instantiates concrete data flows based on data transfer technologies in WP4 to track domain specific files, versions of files, confirms transfer success, and tracks files available in different locations. Generic interfaces should provide transparent access to multiple underlying transfer and storage modules (e.g. gridFTP/irods/object store etc.)

Subtask 9.3.2: EGA data access authorization integration. (12PM)

Integrate EGA data access authorizations to local project data portals and Cloud access providers. This is a new service that allows authorized third-party services to programmatically check compliance with the current user data access authorizations from the ELIXIR coordinated repositories such as the EGA database each time user accesses a file in the cloud or other remote service. A first planned project using EGA data within the private, secure, cloud at CSC in Finland will provide our reference implementation.

Subtask 9.3.3: Data access APIs. (14PM)

Develop and implement standard **data access APIs to be to** used for inter and intra cloud communication and for secure remote REST API access in coordination with the Global Alliance for Genomics and Health (GA4GH).

For tasks 1-3 we expect to list a number of updates to the submission tools while we work with the first 2-4 chosen

resources. These updates will be prioritised in the scope of this WP.

WP4 will provide AAI support for WP 9, and vice versa WP9 will work with WP4 to set the requirements for ELIXIR AAI services. WP9 needs to information on service component availability and this information is expected to be available from technical services registry such as cloud resource allocation, valid EGA data access authorizations, and file mirroring status if data are not yet ready to be used in the cloud. WP12 will Create a set of Legal Frameworks for ELIXIR-related operations that will be integrated within WP9 with the technical solutions devised for particular EGA needs.

Deliverables

D9.1: Requirements analysis document. (**M12**)

D9.2: Report on implementation of submissions interface and API. (**M36**)

D9.3: Report on implementation of value-added user applications and cohort integration. (**M36**)

D9.4: Report on implementation of cloud access and secure user and data management. (**M36**)

D9.5: Report on implementation of ELSI and policy consideration for controlled access data. (**M36**)

Work Package number	10	Start Date or Starting Event					M1
Work Package title	ELIXIR Node Capacity Building and Communities of Practice						
Participant number	1	2	5	7	9	13	17
Short name of participant	EBI	UOXF	UTARTU	CNIO	FVIB	CSIC	FCG
Node	EMBL-EBI	UK	EE	ES	ES	ES	PT
Person-months per participant:	6	4	20	1	3.6	2	2
Participant number	18	21	22	24	26	27	32
Short name of participant	INESC-ID	CSC	UiB	UiT	SIB	CNRS	LIU
Node	PT	FI	NO	NO	CH	FR	SE
Person-months per participant:	10	4	4	4	6	5	24
Participant number	33	35	36	38	40	41	
Short name of participant	UL	UOCHB	MU	VIB	BSRCAF	HUJ	
Node	SI	CZ	CZ	BE	EL	IL	
Person-months per participant:	30	8	26.6	10	12	8	

Objectives

WP10 is focused on strengthening the ELIXIR infrastructure by supporting coordination of Node activities and increasing the organisational capacities of ELIXIR Nodes. ELIXIR Nodes are at very different levels of maturity, ranging from national infrastructures that have existed for over a decade to newly formed consortia. Activities will focus on spreading the knowledge and bioinformatics best practice that exists within ELIXIR's larger and more established Nodes, with newer or smaller ELIXIR Nodes in less research-intensive areas of the EU. This will help to create a stairway to excellence for partners involved, and support the creation of a true European Research Area. One of the deliverables will be a set of "*Good practices*" for setting up and running an ELIXIR Node, which will be of substantial value for both current and future Nodes.

Its two Objectives are:

- Implement a programme of organisational capacity building in newly formed ELIXIR Nodes, including sharing of best practice between partners in accessing EU Structural Funds (ESIF) for operating infrastructure.
- Construct and coordinate ELIXIR-wide 'communities of practice' that support and develop the professionals who deliver advanced data and bioinformatics support and services in ELIXIR Nodes.

Work Package Leads: Jiří Vondrášek (CZ) and Bengt Persson (SE)

Description of work

This WP will address the issue of how to get people in Nodes coming together in capacity building, as detailed in the tasks below. There will be accompanying training needs in this capacity building and those training needs will be addressed in WP11. The training needs are in advanced training of the staff handling data and performing genome annotation and assembly. Other training needs for Use Cases will be in general addressed in WP11, but not specific to every Node. For Node capacity building, advanced training will be needed also in management and know-how on operating Nodes, performed in close collaboration with Task 10.1.

Specialized Capacity Building

A *Community of practice* is a group of people who share a craft or a profession, created to coordinate efforts to solve defined tasks and/or with the goal of gaining knowledge related to their field. ELIXIR is looking to establish such *Communities of Practice* of bioinformatics experts involved in advanced bioinformatics user support across the Nodes to effectively interact with bioinformatics infrastructure users at interfaces of different research fields. ELIXIR *Communities of Practice* would be the primary mechanism for ELIXIR to establish domain specific services, for example, forming a community of genome annotators across Nodes to meet the need from national researchers of ready access to genome annotation resources. Other examples could be to meet the needs of Rare Disease or Medical genomics research, agricultural or marine bioinformatics and chemical compounds for biology. ELIXIR will start to build these *Communities of Practice* to enable coordination and knowledge exchange in selected areas in tasks 10.2 and 10.3. Task 10.2 is directed to create *Good Practices* in setting up data Nodes, of importance to create a sustainable and scalable data flow from laboratories to national Nodes and further to European or global databases. Task 10.3 is directed to coordinate and exchange expertise in the field of genome annotation and assembly and to create *Good Practice* in for this field. In the future, further *communities of practice* are envisioned, arising from needs identified by the Use Cases (WP6 to 9) and identified through community workshops and surveys (Task 10.4). The creation of a sustainable mechanism for establishment of communities of practice is also addressed in Task 10.4.

Task 10.1: ELIXIR Node Capacity Building (46PM)

This task will support the formation of an ELIXIR community. There are significant differences between existing ELIXIR Nodes in their capacity, level of expertise and maturity of services/tools/data. We will increase the joint competence and capacity for Nodes lacking a large national user community, large-scale projects and big data or having a limited record of offered tools and services. These Nodes will benefit from mutual collaboration and connection with well-established and more advanced Nodes they can utilize their know-how for a more rapid Node development. Altogether, this will help shape ELIXIR as an efficient pan-European infrastructure.

The major aim of this task is to provide management knowledge transfer among Nodes to create a set of well-balanced, well-functioning and compatible Nodes.

Support in coordinating national Nodes, including Skills and Knowledge exchange between ELIXIR Nodes. Nodes with different experiences will help to provide knowledge regarding good practice in different situations and providing direct support to implementation of national infrastructures (e.g. by national / regional workshops with external experts, support to national community building efforts). The heterogeneity of Nodes established will help providing multiple effective ways for coordination and to get funds from national providers and their commitments to the infrastructure. Knowledge exchange will be catalysed by workshops, staff exchange programme and visits. This activity is based on the ELIXIR community practice experience but it is more general and should cover some features brought by larger staff community.

Identify and apply technical solutions at/between Nodes. The reason for particular technical solution must be explicitly formulated and the solution must be applicable on more than 2 Nodes. The capacity building deliverables would be primarily workshops based on Technical Services and/or Training WP deliverables.

Partners: CH, CZ, EE, NO, PT, SE, SI, UK, ES, EL, IL, EMBL-EBI

Task 10.2: Capacity Building in Data Nodes Network (34PM)

One of the aims of ELIXIR is to establish a network of data Nodes (Nodes with large data collections and databases with established way of data deposition and curation) to enable scalable data storage and their transferability by means of standardised formats. In this task, we will focus on establishing guidelines and good practices to facilitate efficient data collection into core data resources (cf. WP3), primarily focusing on data needed for selected Use Cases (WP6 to 9). This is tightly linked with IT solution by means of storage, dedicated networks and connections (cf. WP4). A distributed network following the same standards will also simplify international

sharing of datasets for which this is ethically permitted.

This task both includes creation of routes for data publishing in a uniform manner across ELIXIR with data Nodes in each country and includes data repositories for replication of reference data allowing for fast access. The setting up of a data Nodes network has been identified by the technical experts within ELIXIR as a prioritised area.

Task 10.2 also includes development of *Good Practices* in setting up data Nodes enabling secure storage of sensitive data, such as sequence data related to patients. The task is interfacing with WP4 regarding technical developments on AAI and data transfer. Furthermore, there are connections with WP4 on data interoperability and the Use Case in WP9 on sensitive data.

Partners: SE, FI, CZ, EMBL-EBI, SI, PT, ES, EE. In due time, all ELIXIR Nodes are expected to have an ELIXIR data Node.

Task 10.3 – Capacity Building in Genome Assembly and Annotation (44PM)

Specialised expert platforms for genome assembly and annotation are already available in several ELIXIR countries. They provide critical support to complex genome projects and deliver annotations that serve as the basis for scientific inquiry into the genomics of newly sequenced organisms. The specialised expertise at multiple ELIXIR Nodes would benefit from capacity building through competence-spreading advanced workshops and staff exchange.

The capacity-building efforts will benefit the Use Cases in WP6 on marine organisms and in WP8 on plant Use Cases. The genome annotation groups will contribute with domain-specific knowledge about different species, e.g. marine organisms (SE, NO), woody plants (PT) and crop plants (SI).

Furthermore, in order to facilitate access to genome annotation to the users, we propose a deployment of web services to enable genome projects in the scientific community to efficiently interact with the data. The development of such web services is intended together with the EnsEMBL team to create a pan-European collaboration on genomics resources to provide researchers with a unified analysis platform carried by multiple partners.

Partners: SE, NO, FR, PT, EBI, SI, BE, CZ, ES.

Task 10.4 – Sustainability of capacity building (30PM)

The main goal of Task 10.4 is periodical and long-term discovery of users with specific capacity needs at ELIXIR Nodes and/or research groups within Nodes. This knowledge of capacity needs/gaps will be gathered through surveys and face-to-face meetings. With capacity needs identified the Task 10.4 team will connect users with WP11 groups that have at their disposal training infrastructure, learning materials and knowledge needed to implement the capacity building. In order to ensure the sustainable flow of knowledge and stable capacity maintenance we need to provide long-term networking of capacity seekers and providers. They will be focused to the great extent to the *Good Practices* from Task 10.2 and 10.3 (and WP6 to 9). With well-formed ELIXIR Communities of practice, the Task 10.4 will be able to lead the reuse or even suggest the adaptation of WP11 courses and training materials for specific capacity building needs.

It is of great importance that capacity needs will be periodically (but in long-term perspective) tested through surveys, which will also contribute to the sustainability of training infrastructure and learning materials provided by WP11.

Task 10.4 will monitor the implementation of capacity building in Tasks 10.1, 10.2 and 10.3 in order to extract good practices and compile good practice recommendations and guidelines which can be used in other capacity building contexts.

Partners: SE, SI, CZ, BE, EE, EL, IL, EMBL-EBI

Task 10.5: Supporting ELIXIR Nodes in understanding Smart Specialisation Strategies and accessing EU Structural and Investment Funds (ESIF). (36.2PM)

The potential for exploiting funding synergies between EU Research programmes and ESIF are well known⁶². Those ELIXIR Nodes eligible for ESIF are therefore presented with a real opportunity for local funding of their Node, particularly in light of the proposed focus on ESIF and ESFRI that many Member States are making within their national plans to the Junker Investment Plan. However, understanding the local priorities for funding, rules,

⁶² CoPoRI - Workshop about Research Infrastructures and Structural Funds:
http://www.elixir-europe.org/system/files/copori_report_final.pdf

and application procedures presents is complex and time consuming and securing ESIF for operational costs of life science infrastructures is a real challenge. For ELIXIR Nodes to access ESIF in any meaningful way, support needs to be targeted at the local level, allowing scientists to build up an understanding of their local Smart Specialisation Strategy, which dictates the funding opportunities for that region, and then develop a strong business case that can be used for subsequent funding applications.

Partners: CZ, SI, EE, EL

ELIXIR ESIF Task Force (Months 1-12)

ELIXIR Structural Funds Task Force grouping funding specialists across ELIXIR Nodes will be established to share best practice in ESIF use for research infrastructures. The Task Force would also engage external experts such as ones from national managing authorities for ESIF, DG REGIO, DG EMPLOY, DG Enterprise and Industry and Jaspers and would make use of existing reports such as the ESPON KIT report (www.espon.eu).

An ELIXIR-wide Workshop early at start of the project to pool good practice on using Structural Funds to support research infrastructures and facilitate personal interactions. Meeting will be hosted and organised by CEITEC, who leads this task.

This would include talks from ELIXIR Nodes with experience of accessing Structural Funds (Estonia, Czech Rep, Slovenia), as well as other ESFRIs such as ELI that have done this successfully in other disciplines

Local priorities and their overlaps identification towards Business Case (Months 6-24)

As all regional priorities are different, and as the application process for funding is done in the local language and following local rules, target Nodes will work with their regional partners to understand the priorities. This task will support Nodes in understanding their local Smart Specialisation Strategy and the regional priorities relating to research and life sciences. Access support from Jaspers following the connections built up within Months 1-12.

Supporting Nodes in actually developing the Business Cases and applications for Structural funds to support the construction and/or operation of the Node. The timing of this work will depend on when the calls will be opened for each region.

Partners: CZ, SI, EE, EL

Deliverables

D10.1: Communities of "Good Practice" and Protocols of Transfer. (**M48**)

D10.2: Blueprint on how to set up an ELIXIR data Node (Report). (**M48**)

D10.3: ELIXIR Advanced workshops on genome annotation. (**M4, M22**)

D10.4: A Node report on regional priorities and channels to use. Available to all relevant partners in the Node network in that country. (**M18**)

Work Package number	11	Start Date or Starting Event				M1	
Work Package title	ELIXIR-EXCELERATE Training Programme						
Participant number	1	2	4	5	6	7	11
Short name of participant	EBI	UOXF	UNIMAN	UTARTU	NBIC	CNIO	UMA
Node	EMBL-EBI	UK	UK	EE	NL	ES	ES
Person-months per participant:	14	44	30.8	42	11.5	1	2
Participant number	13	17	25	26	27	31	32
Short name of participant	CSIC	FCG	UiO	SIB	CNRS	CNR	LIU
Node	ES	PT	NO	CH	FR	IT	SE
Person-months per participant:	0.8	12	2	14	4	12	4
Participant number	33	36	40	41			
Short name of participant	UL	MU	BSRCAF	HUJ			
Node	SI	CZ	EL	IL			
Person-months per participant:	42	7.2	2	4			

Objectives

WP11 has two principal objectives:

- To build a sustainable training infrastructure for ELIXIR's community - encompassing both a technical infrastructure and training expertise, as well as mechanisms for guaranteeing quality of training. (Task 11.1)
- To develop and deliver training in topics selected as training gaps within the ELIXIR community in selected application areas (namely WP6 to 9 Use Cases). (Task 11.2)

Work Package Leads: Chris Ponting (UK) and Patricia Palagi (CH)

Description of work

As medicine and the life sciences become increasingly centred on the generation, analysis and interpretation of big data, most science professionals will need to become more proficient in exploiting bioinformatics data and systems. The ELIXIR Community is estimated at 500,000 people, drawn from across the spectrum of individuals ranging from life scientists and bioinformaticians, to tool developers and infrastructure operators. This WP will up-skill European researchers focused on the WP6 to 9 Use Cases, who will be empowered to more effectively exploit the data, tools, standards and compute infrastructure provided by ELIXIR, and on ELIXIR-EXCELERATE developers and infrastructure operators.

This Training Programme will be the foundation upon which international bridges will be built, in order to harmonise efforts, to share resources, to avoid duplication/redundancy and to maximise effectiveness. Specific organisations and initiatives with which to collaborate may be (but are not limited to): GOBLET (the Global Organisation for Bioinformatics Learning, Education and Training), ISCB (the International Society for Computational Biology), and Software Carpentry Foundation, and also the training programmes of other ESFRIs such as BBMRI, EATRIS, and ISBE. Collaboration with GOBLET essentially provides a ready-made global gateway to ELIXIR's training resources, and will ensure that the foundations built by GOBLET are not duplicated – the leader of task 11.1.2 also leads GOBLET, assuring that their distinct roles will evolve in harmony. Close engagement with industry will be sought such as through the ELIXIR Industry Programme.

A robust, high quality training infrastructure supporting ELIXIR resources and services will increase the impact and visibility of ELIXIR as a whole, ensuring that ELIXIR resources are introduced into industry and academia. It will catalyse and support the (self-) training of researchers, increase the impact of ELIXIR services, and facilitate scientific excellence.

All activities in this Training Programme will build on existing national training experience and expertise in the Nodes, and will complement, enhance, and disseminate these activities. Training activities will be closely aligned with the ongoing development of data infrastructure, standards, tools and services in each Node. This work paves the way for a later comprehensive, high quality and sustainable training programme to be rolled out over the entire ELIXIR Community.

The ELIXIR Training Programme targets the following distinct stakeholder groups:

- The ELIXIR community encompassing its partners, associated industry and SMEs (Task: 11.1)
- Developers and Infrastructure operators of ELIXIR's services (Task: 11.2.1)
- Users of ELIXIR's services (Task: 11.2.2)
- ELIXIR's trainers (Task: 11.2.3)

Task 11.1: Building an ELIXIR Training Infrastructure (143.3PM)

Subtask 11.1.1: Assessing training quality, good practice and impact (45PM)

ELIXIR Training needs to be timely, impactful, of high quality, and at scale. It will draw upon existing expertise in member countries that has yet to be combined under a unified structure. A Good Practice Coordinator will identify and deliver a framework of good practices throughout ELIXIR and ELIXIR-EXCELERATE. An initial workshop involving the ELIXIR Training Coordination Group (TrCG), industry/SMEs and other stakeholders will focus on defining specifications, metrics and key performance indicators to ELIXIR training. This subtask will build confidence for those seeking training (whether from academia, industry or other sectors) that ELIXIR training is being delivered to high standards, ensuring “best in class” training provision. It will also provide ELIXIR with a mechanism to capture and report on the impact of its training programme on the European and international level, and will harmonise across Nodes in standards, metrics and sharing of training materials. Development of well-structured training routes – through the use of workflows – embedded in online resources (WP11.1.2 and WP11.1.3) will specifically address urgent needs of WP6 to 9 Users and industry for cost-effective, time-effective,

impactful training.

Partners: UK, EMBL-EBI

Additional resources required: Workshops EUR 20,000

Subtask 11.1.2: TeSS Training Portal (51.3PM)

A training portal, TeSS, will be developed in this task to be an active forum for aggregating, disseminating and coordinating information on ELIXIR-EXCELERATE's training activities/materials, including those relating to ELIXIR's Core Resources. Building on the TeSS prototype being piloted by ELIXIR-UK, this task will 'harden' the prototype, harness the outputs from WP6 to 9's Use Cases and 3rd-party content providers, and synergise globally with GOBLET. Importantly, training information from all Nodes will be pulled into the TeSS, ensuring ELIXIR-wide coverage. The TeSS will enable registration and discovery of training activities/materials through multi-centre information aggregation, it will allow users to collect ('package') sets of materials/tools/data required for training, and offer workflows that allow related resources to be identified and harvested from source. The roles of GOBLET and TeSS are complementary: the former (a materials/course repository and trainer directory) acts as a feed to the latter (a resource aggregator and dissemination hub), obviating the need for ELIXIR-EXCELERATE to build its own repository. No mechanism currently exists via which users may either readily discover ELIXIR-EXCELERATE training events/resources (many of which are dispersed on websites across ELIXIR member states and beyond) and/or determine their relevance (e.g., what is the audience of this course (beginner, advanced, etc.), what is its duration, where is it being held, which course should I take next, etc.?). The added value of this task is therefore in coordinating and making discoverable ELIXIR-EXCELERATE training activities/materials, and surfacing information in ways that support user decisions and choices. We will ensure wide uptake by all stakeholders (from trainers and trainees, to resource providers and developers, including those across industry/SMEs) via community-building events. This subtask will result in the release of the TeSS as an open resource, contributed to by the community, and shaped both by the community and by the outputs of ELIXIR-EXCELERATE's Use Cases (WP6 to 9).

Partners: UK, CH, PT, NL

Subtask 11.1.3: eLearning (47PM)

Distance training has become essential to reach large audiences spread over many countries, as is the case for the ELIXIR community. This task will survey the current technological and pedagogical options, the existing e-learning expertise and technology in the Nodes, and possible international partners (Coursera, edX, Udacity, GOBLET etc.). In concert with the ELIXIR training community, we will decide for the best and suitable e-learning strategy for ELIXIR and then proceed into implementing this strategy, which will complement the TeSS training portal. During this process we will develop ELIXIR e-learning expertise, which will be exploited to create guidelines and recommendations for future trainers. A major part of this task will be devoted to derive scalable training materials related to the WP6 to 9 Use Cases and ELIXIR resources, which will all be made available to the ELIXIR community of trainers, developers and users. The trainers will benefit when delivering their courses; developers and users will benefit from the availability of materials.

Partners: SI, NL, CH, PT, IL

Additional resources required: Workshops EUR 20,000

Task 11.2: Delivering training to the ELIXIR-EXCELERATE community (106PM)

Subtask 11.2.1: Train the Developer and Infrastructure Operator (30PM)

This subtask targets all developers and infrastructure operators who are developing and maintaining ELIXIR services, distributed over all Nodes and ELIXIR WPs (most notably WP3, WP4, WP6 to 9, and WP10). The main activities are: a) mapping the competences across the Nodes and analysing collected training needs from the technical people through internet queries, discussion rounds at the annual ELIXIR All Hands Meeting and within the technical coordinators group; b) providing targeted training based on the mapped needs as face-to-face workshops or webinars where applicable; c) implement a dedicated virtual community space to ease the communication between the developers and infrastructure operators across the Nodes. The aim is to have a virtual coffee room where issues regarding resource development and ELIXIR services can be shared and re-occurring problems can be easily noticed and solved. The virtual community will allow buddying up individual developers across the Nodes who are working on similar tasks but otherwise would not meet nor have discussion possibilities. On top of all the information gathered, specific training can be quickly delivered in response to identified needs. Improving developers' skills and enabling easy knowledge exchange across the community will have immediate impacts on the quality of tools provided by the ELIXIR community. With substantial competition among analysis tools and databases in the life sciences, better designed and executed services obtain greater visibility not only in

Europe but worldwide thus further raising the awareness of ELIXIR services.

Partners: EE, EMBL-EBI, SE, IL

Additional resources required: Training Procurement EUR 60,000

Subtask 11.2.2: Train the Researcher (43PM)

It is essential that ELIXIR-EXCELERATE Users (e.g. in WP6 to 9) acquire skills that empower them to transform big data into meaningful knowledge. This Subtask will plug the training gaps using HPC facilities and expertise, specific cloud computing workflow training needs (WP4) and existing solutions, and the provision of training workshops. Life scientists, computational biologists, and bioinformaticians engaged in Marine Metagenomics, Plant Genotype-Phenotype interactions, Rare Disease and Biological Sample Collection research will be trained in domain-specific knowledge. To become independent and savvy users of the ELIXIR resources training will be given in: scientific programming, statistics, HPC, ontologies, workflows, data curation and annotation (WP3), as well as in the effective use of standards and FAIR data. Reproducibility in the life sciences is a rapidly growing problem. Targeted training is anticipated to improve experimental reproducibility, to empower scientists to address new avenues of research requiring big data analysis, and to cross-pollinate diverse disciplines with innovative analytical approaches. Training opportunities will be prioritised, in consultation with Industry/SMEs, and training delivery will be monitored (WP11.1).

Partners: ES, CH, UK, NL, SI, NO, PT, EL

Additional resources required: Workshops EUR 140,000

Subtask 11.2.3: Train the trainer (33PM)

For the ELIXIR training to achieve maximum impact, adequate capacity to train in all Nodes is required. This task will train the trainer pool that already exists – albeit in a fragmented manner – across ELIXIR Nodes and to identify and train new trainers. It will build a highly skilled and coherent community of training instructors by initially surveying available training capacity and developing an appropriate train the trainer (TtT) framework. Using targeted surveys and drawing upon knowledge and skills from Nodes with greater experience (and with WP11.1), expertise will be developed in Nodes, for ELIXIR Core Resources and Use Cases, where training opportunities have hitherto been scarce. The task will deliver a framework and associated material and guidelines for training new trainers across ELIXIR, from both industry and academia. To enhance delivery, training will take advantage of ready-to-run virtual machines (VMs, developed within ELIXIR and/or in collaboration for instance with BioImg.org) and clouds, assisted by a workshop that gathers both trainers and infrastructure specialists from WP4.2. TtT sessions will accompany training courses in Marine Metagenomics (WP6), Plant Genotype-Phenotype (WP7) and Rare diseases (WP8) providing good practice, and guidance for prospective new trainers. TtT will build capacity whilst stimulating collaborations spanning ELIXIR Nodes and other stakeholders, including industry/SMEs which represent a currently unexplored resource.

Partners: IT, EMBL-EBI, SE, CH

Additional resources required: Workshops EUR 154,000

Deliverables

D11.1: Deliver the identified evaluation systems and good practice guidelines for training. Deliver a TeSS platform and identify e-learning solutions. (**M30**)

D11.2: Report on the training needs identified across the ELIXIR community. (**M12**)

D11.3: Report on training provided across the ELIXIR community. (**M48**)

Work Package number	12	Start Date or Starting Event	M1
Work Package title	Excellence in ELIXIR Management and Operations		
Participant number	1		
Short name of participant	HUB		
Node	EMBL-ELIXIR		
Person-months per participant:	126		

Objectives

The objectives are:

- Coordinate ELIXIR-EXCELERATE project management with risk assessment and quality control to ensure

successful and on-time completion of the project deliverables. (Task 12.1)

- Conclude the implementation of ELIXIR coordination and operational structure and collect defined procedures into ELIXIR Handbook of Operations to be available online for all partners. (Task 12.2)
- Review the suitability of ELIXIR's legal framework, as recommended by ESFRI, and provide conclusions and options to the ELIXIR Board. (Task 12.3)
- Analyse emerging technical data and experiences gained from Nodes to update the long-term strategy for the sustainability of data management, which will advise future funding strategy and the ELIXIR Programme. (Task 12.4)

Work Package Lead: Niklas Blomberg (ELIXIR Hub)

Description of work

The purpose of this Work Package is to:

- Run the overall project management function ELIXIR-EXCELERATE; and
- Deliver value to Member States by completing the ELIXIR Management and Operational processes so that ELIXIR Hub and Nodes - operates as a world-class distributed infrastructure to its users.

The WP addresses quality management, risk management, and service sustainability and aims to ensure the delivery of long-term, mission-critical services of high quality. It also includes activities that ensure ELIXIR takes part in future EU grants as a single entity. It delivers the activities that directly addresses the ESFRI recommendations for ELIXIR: 1) strengthen the central coordination role of the ELIXIR Hub, 2) evaluate the appropriateness of the ELIXIR legal framework, 3) enhance ELIXIR operational processes to support the full deployment of the service streams, 4) develop data access and ELSI procedures, and 5) develop common procurement and recruitment procedures. The two remaining recommendations on industry engagement and expansion of ELIXIR's membership are addressed in WP13. This WP further addresses the long-term bottleneck in implementation identified in the Assessment Expert Group review: sustainable funding models for critical, core, data resources that bring together national and international stakeholders

Task 12.1: Coordinate ELIXIR-EXCELERATE project management (78PM) (M1-M48)

Subtask 12.1.1: Day-to-day management, reporting, quality control and risk management (40PM)

The ELIXIR Hub will lead the project management and coordination of ELIXIR-EXCELERATE. This will, as far as possible, make use of established ELIXIR structures and will be complemented by additional resources (where relevant) to ensure appropriate project management practice and completion of all reporting requirements. Time in regular ELIXIR advisory body meetings will be dedicated to discuss ELIXIR-EXCELERATE; however, due to the size and complexity of the implementation project, additional meetings for Heads of Nodes and other permanent ELIXIR working groups will be required.

This task covers all the necessary Project Management, KPI tracking, reporting and risk management activities necessary for a large, distributed project and embeds these activities in the ELIXIR Hub project management unit. Further, in this task the quality control procedures on milestones, deliverables and other project results will be established (e.g. internal peer review, HoN review and acceptance for milestones). This will ensure that deliverables adhere to some quality principles (such as completeness, relevance, uniformity in presentation, etc.), which is important both for internal and external communication and reporting to the EC.

In order to bring together ELIXIR's users and operators, an international conference: "Landmark in Bioinformatics Services" will be organised during the second year of the project. This meeting includes breakout sessions and hackathons for ELIXIR users, in addition to keynote talks by external experts. The meeting invitation will be extended to related infrastructures with the expectation that participants pay for the meeting registration and for their own travel.

Subtask 12.1.2: Develop and maintain a high-class intranet for project partners, committees and governance bodies. (38 PM)

Information flow within the project is ensured through an intranet that will be developed and maintained by the ELIXIR Hub. This is a critical task for a large, distributed project with multiple partners. It will help the ELIXIR Hub to manage documents, assign tasks, report project outcomes, follow up KPIs, manage risks and find people involved in ELIXIR. The intranet will be aligned to the ELIXIR website so information can easily flow from the private area to the public area. The ELIXIR Hub will also develop and maintain system for exposing ELIXIR events, news, vacancies and shared information across the ELIXIR Hub and Nodes website. The ELIXIR Hub will adapt the emerging iAnn standard to federate this content and provide a solution to facilitate exchange of

announcements like news and events across ELIXIR Nodes, ELIXIR participants' institutions as well as other relevant organisations in the life sciences domain.

Partners: EMBL-ELIXIR

Task 12.2: Conclude the implementation of ELIXIR coordination and operational structure (28PM) (M1-M48)

Subtask 12.2.1: Develop ELIXIR operations framework and processes. (23PM)

In order to ensure pan-ELIXIR coherency in the delivery of high quality and mission-critical services, ELIXIR has established permanent working groups on technical and training services: Technical Coordinators Group (TCG) and Training Coordinator Group (TrCG) were formed in 2014. In 2015, upon recruitment of ELIXIR's Data Coordinator, a Data Coordinator Group (DCG) will be established. The coordinator groups constituted by Node Technical, Training and Data coordinators and led by the respective ELIXIR coordinator, coordinate service objectives and action lines within ELIXIR. In this project they have a key role to bring coherence across the Work Packages, in particular the four Use Cases (WP6 to 9) and ensure that solutions are made available nationally and to other communities. They are also responsible for maintaining a long-term view of their respective area. The aim of this task is to drive the technical, data and people coordination within ELIXIR by strengthening the ELIXIR permanent working groups (DCG, TCG, TrCG).

To strengthen the harmonisation of processes across the whole infrastructure (the ELIXIR Hub and the Nodes, as well as external interactions), ELIXIR procedures, recommendations and guidelines will be collected into a living document: the ELIXIR Handbook of Operations. This will be made available through the ELIXIR intranet. The handbook will encompass all ELIXIR operations, e.g. the ELSI guidelines, communication plan, international strategy, diversity and equal opportunities, common procurement and recruitment procedures, and will also link to the Handbook for Nodes (delivered by WP10). The handbook is updated annually and final version released at the end of this project. According to the ELIXIR governance, any infrastructure-wide policies will need to be decided by the ELIXIR Board and therefore, the *Handbook of Operations* will feed into the development of the next *ELIXIR Scientific Programme for 2019-2023*, which the Board will approve in the spring 2018. Progress on the Handbook development will be regularly reported to the ELIXIR Board.

Subtask 12.2.2: Develop the ELIXIR equal opportunities policy. (1PM)

Gender balance and equal opportunities have already been considered within ELIXIR when setting up the ELIXIR governance structure with requirements on e.g. advisory boards. In addition, each ELIXIR Node has their own employment policies in place. However, there is not yet a developed ELIXIR-wide equal opportunities policy. Within this project, the equal opportunity recommendations in access to training and services will be reviewed throughout ELIXIR, along with developing the Charter and Code of Access to research infrastructures. This task will draw on the diversity work across BMS research infrastructures planned within the *CORBEL* cluster project. In addition, collaborations with the emerging NIH BD2K Biomedical Data Science Training Coordination Centre will be sought, not only on bioinformatics training and shared infrastructure but also on gender/diversity issues. Based on these a good practise recommendation for ELIXIR on diversity issues will be completed to be included into the *ELIXIR Handbook of Operations*.

Subtask 12.2.3: Investigate and recommend options for common procurement across Nodes. (2PM)

ESFRI recommends ELIXIR to develop common procurement and recruitment procedures. To gain understanding in the processes, we will investigate how commercial partners work with data intensive partners in the life science. A feasibility analysis will assess options on compute and storage services with ELIXIR datasets and explore joint procurement for commercial cloud services, e.g. AWS, Google Cloud, and Azure. This analysis will include access for SMEs and will be performed together with WP4 and partners from the EMBL-EBI, Finnish and Czech Nodes, who are experts in technical services. The ELIXIR Hub will implement the strategy and include it into the ELIXIR Handbook of Operations.

Subtask 12.2.4: Establish ELIXIR's internal processes for ELSI. (1PM)

Within Europe, BBMRI-ERIC has already established Common Service ELSI (<http://bbmri-eric.eu/common-services>). Although their work is focussing on the biobanking community, the Common Service ELSI provides a very useful framework for ELIXIR. We are currently developing a Memorandum of Understanding with BBMRI-ERIC with the intent of establishing a long-term relationship for ELSI Services that maximise infrastructure synergies. The aim of this task is to establish ELIXIR's internal processes for ELSI by bringing together the planned work within *CORBEL* (see section 1.3.2) and ELIXIR Nodes' national ELSI processes and ensure that ELIXIR's roles, policies and responsibilities are clear and transparent – the foundation for effective partnership. Once established, the ELSI processes will be included into the *ELIXIR Handbook of Operations*.

Subtask 12.2.5: Review ELIXIR Node Collaboration Agreements to streamline participation on grants. (1PM)

There is a clear need to review and assess the suitability of ELIXIR Nodes' legal structures in the context of EU grant applications, where the intention is for ELIXIR to participate in future grants with as fewer numbers of separate beneficiaries as possible. ELIXIR-EXCELERATE shows that ELIXIR Nodes - and their various legal forms - are very heterogeneous and in many cases don't yet allow for participation by a single entity. Options such as Joint Research Units and Linked Third Parties will be encouraged.

Partners: EMBL-ELIXIR

Task 12.3: Assess the suitability of ELIXIR's legal framework as recommended by ESFRI (14PM) (M18-M36)

The legal framework of ELIXIR is based on the ELIXIR Consortium Agreement (ECA), which has been concluded among the Member States and EMBL and officially entered into force on 12 January 2014. The ECA covers ELIXIR's mission, membership, obligations of the Members and the ELIXIR Hub, the governance structure of the ELIXIR Hub and relationship to the ELIXIR Nodes. Based on the ECA, EMBL carries out activities on behalf of and as mandated by the ELIXIR Consortium that require EMBL's legal personality. ESFRI recommends that under the ELIXIR-EXCELERATE project, ELIXIR would "consider the long term sufficiency of the present legal framework as EMBL special project". Therefore, this task will: 1) Arrange a workshop with ELIXIR members to agree on a scope of the legal framework review and develop a request for tender document, 2) tender for an external expert to review and compare different legal frameworks, and 3) arrange a workshop to conclude recommendations for ELIXIR's future legal framework based on the consultant review. This task will require the ELIXIR Board administrative delegates and legal experts from ELIXIR Member States. As the Board members are not from partner institutes, their travel costs are budgeted within this task. The tender evaluation committee will be appointed by the ELIXIR Board.

Partners: EMBL-ELIXIR

Additional resources required: Procurement (external consultant) EUR 60,000

Task 12.4: Ensuring long-term strategy for the sustainability of data management (6PM) (M1-M36)

The long-term sustainability of Europe's public data resources is of paramount importance. Indeed, ELIXIR is recognized as Europe's attempt to coordinate, integrate and facilitate the long-term sustainability of these critical resources. The aim of this task is to evaluate the relevance of the previous work in this area and present an updated, forward-looking strategy to feed into the long-term funding strategy and ELIXIR Programme for 2019-2023. This task is a direct response to the recommendations of both ESFRI and the Assessment Expert Group, and will bring together senior policy makers, technical experts and resource owners from ELIXIR Nodes.

This task will build on other activities within ELIXIR-EXCELERATE. A permanent working group, already created within ELIXIR, will work within ELIXIR-EXCELERATE and monitor the emerging technical data from WP3 (development of metrics around core resources) and WP5 (interoperability requirements) and the experiences from WP10 (increased capability from Nodes). The WG will consolidate the collected data into a high-level strategy that informs the Programme. In addition, institutional good practises will be defined and rolled out via the Node Handbook (WP10). The WG members are ELIXIR Board administrative delegates and scientific experts from ELIXIR Member States. As the WG members are not from partner institutes, their travel costs are budgeted within this task.

Partners: EMBL-ELIXIR

Deliverables

D12.1: ELIXIR Handbook of Operations. (**M12**)

D12.2: Project periodic and final scientific and financial reporting accomplished. (**M12, M30, M48**)

D12.3: Final report on legal framework including working group analysis, conclusions and options for the ELIXIR Board. (**M36**)

D12.4: Updated long-term strategy for the sustainability of data management. (**M36**)

Work Package number	13	Start Date or Starting Event	M1
Work Package title	Communications, Industry and Community Engagement		
Participant number	1		
Short name of participant	HUB		
Node	EMBL-ELIXIR		

Objectives

The overall aim of WP13 is to ensure that a maximum number of scientists and companies are aware of ELIXIR's services, developments, opportunities and events. Reaching out to specific communities, including industry and SMEs, will allow ELIXIR to understand user's needs better, ensuring that the most effective services are developed. This will be achieved by carrying out tasks that address four, complementary objectives:

- Develop a coordinated and user-centric Communications presence for ELIXIR. (Task 13.1)
- Support innovation and usage of ELIXIR from Europe's life-science industries and SMEs. (Task 13.2)
- Create specific outreach channels to key user communities. (Task 13.3)
- Engage potential members across the globe. (Task 13.4)

WP13 will build upon many existing communications resources and channels. The main ELIXIR-Europe website, for example, will house the project website and existing ELIXIR stakeholder mailing lists will help ensure that outcomes from the project are disseminated widely. WP13 will focus on enhancing these existing platforms and optimising them to ensure maximum impact from a relatively modest WP budget.

Work Package Lead: Andrew Smith (ELIXIR Hub)

Description of work

It is estimated by the EC that there are around 500,000 life scientists in the EU alone⁶³. The fundamental importance of data-driven research in the life sciences means that many of these scientists are potential users of ELIXIR's services. ELIXIR likely has largest user community of any ESFRI Research Infrastructure. Furthermore, ELIXIR's Preparatory Phase report⁶⁴ highlighted the range of users supported, including: bioinformaticians and computational biologists; bench scientists; biologists; geneticists; biochemists; clinical specialists; and plant, environmental and marine scientists. In addition, industry is a major user of ELIXIR resources from Pharmaceutical and Biotech industries through to crop science and aquaculture companies and SMEs.

Unlike many physical or single-sited infrastructures, where operators can build face to face dialogue with a small number of users, ELIXIR is an Open Data infrastructure where millions users access services via web interfaces. It is vital to create the effective communications and outreach mechanisms to interact with these users in a manner that is effective for users and efficient for ELIXIR partners.

Task 13.1: Develop a coordinated and user-centric Communications presence for ELIXIR (16PM)

The scale of users across the globe requires that ELIXIR uses modern, effective tools to communicate its scientific and operational messages. The distributed nature of ELIXIR, with over 100 institutes involved, also requires that there is a consistency and clarity of messaging between partners. Task 1 comprises three Sub-tasks that will collectively ensure that the internal processes are in place so that ELIXIR can present a single, coordinated interface to its users and that effective communications materials are developed.

Partners: EMBL-ELIXIR

Subtask 13.1.1: Develop and implement an ELIXIR-wide Communications Strategy (8PM)

The ELIXIR Communications Strategy will set the overall strategic direction for ELIXIR's dissemination activities to users in academia, industry policy-makers and funders. The strategy will include scientific communication, social media, web-based resources, newsletters, video and printed materials. It will be reviewed and updated annually to take into account new technologies and developments. Owned by the ELIXIR Hub, it will be developed jointly by the emerging Communication Expert group (Task 13.1.2) and other key strategic bodies, most notably the ELIXIR Heads of Nodes committee and the Work Package leads of ELIXIR-EXCELERATE. It will identify the key scientific conferences that ELIXIR should attend and the suite of materials that will be produced to reach out to user groups (Task 13.1.3). It will also harness the existing social media channels run by ELIXIR Nodes, and build on the distribution channels created during ELIXIR's Preparatory and interim phases.

Subtask 13.1.2: Create and operate an ELIXIR Communications Expert group (4PM)

In order to implement the most effective Communications Strategy, a tightly-coordinated network of communications and outreach experts from ELIXIR Nodes is required. Harnessing the local communication

⁶³ EC: Enabling Science, 2013: <http://ec.europa.eu/research/infrastructures/pdf/enabling-science.pdf>

⁶⁴ ELIXIR Preparatory Phase User report: <http://www.elixir-europe.org/documents/user-communities-report>

channels that ELIXIR Nodes already have with their own respective national bioinformatics communities will provide efficiencies of scale at the European-level. We will create an information exchange network of Node experts that adopts the latest software and web-based applications. Annual face-to-face meetings of all Communications experts and bilateral meetings between members will be critical to cement relationships within the network, and will be supplemented by video/telephone conferencing and use of the ELIXIR intranet (WP12).

Subtask 13.1.3 Develop suite of high-quality communications materials (4PM)

To effectively communicate ELIXIR's services to various stakeholder communities, a suite of high-quality promotional materials will be produced. A short ELIXIR video will feature interviews with scientists and industry users and will set out the structure of ELIXIR, the range of services offered and how these can be accessed. Printed materials, posters and banners will be printed and distributed to all ELIXIR Nodes to support their local dissemination activities. Annual scientific reports will capture the scientific and operational output of ELIXIR-EXCELERATE.

Additional resources required: Video and printed materials subcontracting costs EUR 45,000

Task 13.2: Support innovation and usage of ELIXIR by Europe's life-science industries and SMEs (8PM)

Industrial use of Europe's bioinformatics resources is high. These industries are major employers globally, generating wealth and supporting transformation to a knowledge-based economy.

Industry sees added value in ELIXIR⁶⁵ in terms of reduced costs and decreased duplication of effort for them, higher scalability, access to common data and interface standards and much better public-private data integration. Many of these requirements will be addressed by the general implementation of ELIXIR and are as relevant for academia as for industry. Additionally, Task 13.2 will ensure that industry and SMEs are offered direct, bespoke support and engagement.

Subtask 13.2.1: Build lasting partnerships with industry initiatives, bodies and SME associations (4PM)

In order to understand the needs of Europe's industry and SME sector, it is necessary to forge close links with industry and SME association bodies, particularly EFPIA, local EENs, the European Biotechnology Network, the BioBasedIndustries Consortium and regional bioclusters. A scoping exercise and face-to-face meetings will map these organisations and establish dissemination channels. This will ensure that a maximum number of companies are aware of ELIXIR's services and that the content of ELIXIR's Innovation and SME programme (Task 13.2.2) is relevant to the needs of Europe's SMEs. The Innovative Medicine Initiative (IMI) - and many of its current projects - is considered as a priority for closer engagement. ELIXIR will seek to conclude formal collaboration agreements with key IMI projects such as OpenPHACTS.

Subtask 13.2.2: Expand the reach and impact of ELIXIR's Innovation and SME programme (4PM)

Expanding ELIXIR's emerging Innovation and SME programme⁶⁶ will allow partners to disseminate the results, services, tools and outcomes to a large number of companies across Europe. Building on a successful start to this programme, we will fund 8 dedicated events across the duration of the project. With around 50 delegates attending each event, hundreds of Europe's SMEs will be directly supported across the duration of the grant. Individual events will be themed to address the outcomes from ELIXIR-EXCELERATE Use Cases (WPs 6 to 9) so will include health and rare diseases, marine sciences and plant sciences. We will work through the key bioclusters and industry organisations identified within Task 13.2.1.

Additional resources required: EUR 120,000 for Innovation and SME events (EUR 15,000 per event)

Task 13.3: Create specific outreach channels across the globe to key user communities and potential member states (6PM)

No infrastructure is effective over the long-term unless it offers services that meet the needs of users. The Preparatory Phase showed that ELIXIR is perhaps unique in the size and variety of its respective user communities. Task 13.3.1 will ensure that ELIXIR engages effectively with users through attendance at scientific conferences. Task 13.3.2 will allow ELIXIR to forge lasting relations with the key bioinformatics community-driven initiatives such as GOBLET for training and proteomics standards initiatives.

In addition, for ELIXIR to realise its potential on the world stage and drive global bioinformatics collaboration, regular interaction, and ultimately direct membership from countries outside Europe, is required. Equally, collaboration with other major global data initiatives is necessary. Task 13.3.3 will ensure that ELIXIR works strategically with key countries outside Europe and the major international infrastructure initiatives that are of

⁶⁵ ELIXIR: Understanding industry's needs: http://www.elixir-europe.org/industry/understanding_industrys_needs

⁶⁶ ELIXIR – Supporting Industry's needs: <http://www.elixir-europe.org/industry/supporting-industry-users>

relevance.

Subtask 13.3.1. ELIXIR representation at key scientific events and conferences (3PM)

Scientific conferences provide an effective way of communicating ELIXIR's services to its large and diverse user base. ELIXIR will attend, display booths and run tech track sessions at a number of key scientific conferences within Europe and beyond. Conference participation will focus primarily on showcasing ELIXIR to scientific conferences at which ELIXIR's users attend. However, some conferences will have policy-makers and funders as the target audience. The exact conferences to attend will be defined in the ELIXIR Communication's Strategy but major events are summarised in Part B section 2.2.5.

Additional resources required: Conference costs of EUR 60,000

Subtask 13.3.2: Forge lasting relations with bioinformatics community initiatives (3PM)

In the life sciences, many users, developers and operators that share a common interest are organised into 'communities'. Many of these are very well established taking the form of initiatives, networks, societies or consortia. I.e. the "HUPO Proteomics Standards Initiative" for the development of standards in proteomics. Such communities provide fora for knowledge exchange and collaboration and platforms for driving common agreements and community development. ELIXIR recognises the contribution of these and will interact with them strategically. We will not duplicate their efforts but rather engage with them to support and promote their, adding value at the European level. This Task will ensure that ELIXIR identifies and collaborates with the key initiatives.

Task 13.4: Develop ELIXIR's International strategy (6PM)

ELIXIR is considered by the G20 countries to be of global significance and with great potential for membership by countries outside of Europe. Task 13.4 will facilitate the process of countries joining ELIXIR as members, thus addressing the Call requirement of increasing membership.

The International Strategy will have a two-fold purpose: 1) support the membership of countries outside of Europe, particularly G20 countries; 2) and foster collaboration between ELIXIR and global science initiatives, such as the Global Alliance for Genomics and Health and INCF. The International Strategy will primarily serve as an internal document and will be subject to annual updates following review by ELIXIR Board and the SAB.

Deliverables

D13.1: Report on ELIXIR's Communication Strategy. (M24)

Making use of user surveys, this Deliverable will assess the effectiveness of the Strategy. Timed for completion in Month 24, it will feed into the development ELIXIR Programme for 2019-2023.

D13.2: Report on ELIXIR's International Strategy. (M36)

The Deliverable is a final report that will assess the effectiveness of the International Strategy. The outcomes and findings of this will be used to feed into the ELIXIR Programme 2019-2023.

D13.3: Report on the impact of the Innovation and SME programme on industry users. (M48)

All participants at events will be polled and asked to complete a detailed user survey, which will form the evidence-base for the Deliverable. ELIXIR's IAC will provide input into the report.

3.1.2 List of Work Packages (Table 3.1b)

WP No	Work Package	Lead Participant No	Lead Participant Short Name	Person Months	Start Month	End Month
1	Tools Interoperability and Service Registry	39, 7	DTU, CNIO	229.7	M1	M48
2	Benchmarking	7, 39	CNIO, DTU	69.6	M1	M48
3	Data Resources and Services	1, 26	EBI, SIB	186.2	M1	M48
4	Technical Services	21, 36	CSC, MU	226.5	M1	M48
5	The ELIXIR Interoperability Implementation	6, 4	NBIC, UNIMAN	252.2	M1	M48
6	Use Case A: Marine metagenomic infrastructure as a driver for research and industrial innovation	22, 24	UiB,UiT	97.0	M1	M48
7	Use Case B: Integrating Genomic and Phenotypic Data for Crop and Forest Plants	1, 17	EBI, FCG	180.9	M1	M48
8	Use Case C: ELIXIR infrastructure for Rare Disease research	7	CNIO	137.4	M1	M48
9	Use Case D: An ELIXIR framework for access-controlled secure archive, dissemination and analysis of human data; enabling biobanks, cohorts and local resource services and integration through the European Genome-phenome Archive (EGA)	1, 14	EBI, UPF	118.0	M1	M48
10	ELIXIR Node Capacity Building and Communities of Practice	35, 32	UOCHB, LIU	190.2	M1	M48
11	ELIXIR-EXCELERATE Training Programme	2, 26	UOXF, SIB	249.3	M1	M48
12	Excellence in ELIXIR Management and Operations	1	HUB	126.0	M1	M48
13	Communications, Industry and Community Engagement	1	HUB	36.0	M1	M48
Total Person Months				2099.0		

3.1.3 List of deliverables (Table 3.1c)

Deliverable Number	Deliverable Name	Work Package Number	Short name of lead partner	Type	Dissemination Level	Delivery Date
D1.1	Registry release with comprehensive coverage of ELIXIR Node resources, including resource data format curation and analysis (Task 1)	WP1	DTU	R	PU	M12, M24, M36, M48
D1.2	Incorporation of monitoring statistics and benchmarking results in registry releases (Task 2)	WP1	DTU	R	PU	M24
D1.3	Workbench integration enabler: implementation & evaluation of impact (Task 3)	WP1	DTU	R	PU	M24
D1.4	Description of the registry user helpdesk & impact on user support via community forums (Task 4)	WP1	DTU	R	PU	M36
D1.5	Matchmaking Service: implementation & evaluation of impact	WP1	DTU	R	PU	M36
D2.1	Creation of a database warehouse infrastructure for storing and organizing data for online performance assessment experiments	WP2	CNIO	R	PU	M18
D2.2	A report on the coordination with WP1 on the incorporation of monitoring statistics and benchmarking results in registry releases	WP2	CNIO	R	PU	M24
D2.3	A report on the features and nature of novel data which are needed within online benchmarking experiments in different subareas	WP2	CNIO	R	PU	M36
D3.1	A report describing the semantically enriched Europe PMC that links literature to ELIXIR Core resource data, exemplified by GeneRif annotations	WP3	EMBL-EBI	R	PU	M36
D3.2	Report describing ELIXIR-wide systems for the computer-assisted collection and delivery of harmonised metrics and quality criteria from multiple ELIXIR resources and collation of these at the ELIXIR Hub	WP3	EMBL-EBI	R	PU	M48
D4.1	ELIXIR Technical Services Roadmap: Incorporates the technical requirements coming from the Use Cases, details of the service offerings coming from the ELIXIR Nodes and the public e-Infrastructures, and provides a roadmap detailing when new services will appear within the ELIXIR Technical Services registry and the work being undertaken by the service providers to meet ELIXIR's requirements	WP4	CSC	R	PU	M11
D4.2	Updated ELIXIR Technical Services Roadmap: Incorporates the technical requirements coming from the Use Cases, details of the service offerings coming from the ELIXIR Nodes and the public e-Infrastructures, and provides a roadmap detailing when new services will appear within the ELIXIR Technical Services portfolio and the work being undertaken by the service providers to meet ELIXIR's requirements	WP4	CSC	R	PU	M23

D4.3	ELIXIR Technical Services Roadmap: Incorporates the technical requirements coming from the Use Cases, details of the service offerings coming from the ELIXIR Nodes and the public e-Infrastructures, and provides a roadmap detailing when new services will appear within the ELIXIR Technical Services portfolio and the work being undertaken by the service providers to meet ELIXIR's requirements	WP4	CSC	R	PU	M35
D4.4	ELIXIR Technical Services document: Describes the ELIXIR Technical Services being offered through the cooperating service providers (ELIXIR Nodes and public e-infrastructures)	WP4	CSC	R	PU	M46
D5.1	<i>Interoperability Implementation Regulations.</i> A set of standards, rules, controlled vocabularies, authorized unique identifiers and interoperable service APIs for the data repositories and biological knowledge bases agreed and implemented in the field with WP6 to 9, and WP3; resources registered in BioSharing; BioCatalogue and BioRegistry (WP1); ontologies accessible via Ontology Lookup Services	WP5	NBIC	R	PU	M18
D5.2	<i>Sustainability business cases</i> for Interoperability Implementation Services, including contractual arrangements with non-ELIXIR resources. Data stewardship agreements with major data interoperability activities in Europe [EGA, EUDAT etc.] USA [DataOne, BD2K] and globally [RDA, GA4GH, Force11]	WP5	NBIC	R	CO (Public Press Release and report will be provided)	M36
D5.3	<i>Bring Your Own Data</i> 8 delivered workshops and established roll-out programme across WP10 Data Node capability network; materials deposited on WP11 TeSS	WP5	NBIC	R	PU	M24
D6.1	Specific marine databases made publicly available	WP6	UiB	R	PU	M18
D6.2	Report on comprehensive metagenomic data standards environment	WP6	UiB	R	PU	M30
D6.3	Report describing a set of tools, pipelines and search engine for interrogation of marine metagenomic data	WP6	UiB	R	PU	M36
D7.1	Datasets annotated for at least 1 woody plant, cereal and solanaceous species; genotype, phenotype and sample meta data submitted to appropriate public archives	WP7	EMBL-EBI	R	PU	M24
D7.2	Phenotypic data exposed through integrating API by all participants	WP7	EMBL-EBI	R	PU	M36
D7.3	New visualisation components for phenotype visualisation released in BioJS	WP7	EMBL-EBI	R	PU	M36
D8.1	Portfolio of ELIXIR data resources and tools for the rare diseases communities	WP8	CNIO	R	PU	M24
D8.2	Documentation of the tools for the data manipulation and standard conversions in the rare-disease field	WP8	CNIO	R	PU	M36
D8.3	Report on the ELIXIR workshop organized with the rare-disease communities	WP8	CNIO	R	PU	M12, M24
D9.1	Requirements analysis document.	WP9	EMBL-EBI	R	PU	M12

D9.2	Report on implementation of submissions interface and API	WP9	CRG	R	PU	M36
D9.3	Report on implementation of value-added user applications and cohort integration	WP9	CRG	R	PU	M36
D9.4	Report on implementation of cloud access and secure user and data management	WP9	EMBL-EBI	R	PU	M36
D9.5	Report on implementation of ELSI and policy consideration for controlled access data	WP9	CSC	R	PU	M36
D10.1	Communities of "Good Practice" and Protocols of Transfer	WP10	UOCHB	R	PU	M48
D10.2	Blueprint on how to set up an ELIXIR data Node	WP10	UOCHB	R	PU	M48
D10.3	ELIXIR Advanced workshops on genome annotation	WP10	UOCHB	R	PU	M4, M22
D10.4	A Node report on regional priorities and channels to use. Available to all relevant partners in the Node network in that country	WP10	UOCHB	R	PU	M18
D11.1	Deliver the identified evaluation systems and good practice guidelines for training. Deliver a TeSS platform and identify e-learning solutions	WP11	UOXF	R	PU	M30
D11.2	Report on the training needs identified across the ELIXIR community	WP11	UOXF	R	PU	M12
D11.3	Report on training provided across the ELIXIR community	WP11	UOXF	R	PU	M48
D12.1	ELIXIR Handbook of Operations	WP12	HUB	R	PU	M12
D12.2	Project periodic and final scientific and financial reporting accomplished	WP12	HUB	R	PU	M12, M30, M48
D12.3	Final report on legal framework including working group analysis, conclusions and options for the ELIXIR Board	WP12	HUB	R	PU	M36
D12.4	Updated long-term strategy for the sustainability of data management	WP12	HUB	R	PU	M36
D13.1	Report on ELIXIR's Communication Strategy	WP13	HUB	R	PU	M24
D13.2	Report on ELIXIR's International Strategy	WP13	HUB	R	PU	M36
D13.3	Report on the impact of the Innovation and SME programme on industry users	WP13	HUB	R	PU	M48

3.2 Management structure and procedures

The ELIXIR-EXCELERATE project will make use of existing ELIXIR governance structures and permanent working groups. INFRADEV-3 is a Call to support the implementation and initial operation of ESFRI research infrastructures and the ELIXIR-EXCELERATE will be used to strengthen ELIXIR's governance and established working groups and drive operational effectiveness, for example, by funding meetings and other supporting activities. No new governance bodies will be created through this project, e.g. the scientific activities will be reviewed annually by ELIXIR's existing SAB in the context of the evaluation of the overall infrastructure and progress reported to the ELIXIR Board as the ultimate decision making body for all ELIXIR activities.

Consortium Agreement – The governance of the project, IP management and the roles and responsibilities of partners and various bodies will be set out in a Consortium Agreement following good practice for large project consortia as set out e.g. in the DESCA 2020 templates⁶⁷ with amendments building on the established ELIXIR governance structure. The CA will state the objectives of the collaboration within parties and their corresponding expected deliveries. The CA will define the project management structure between the partners including roles and responsibilities (as outlined below), budget allocation, grant transfer process, access rights to pre-existing know-how and novel knowledge from the project, IPR, exploitation, dissemination activities, publication rights and processes, liabilities, knowledge disclosure and confidentiality.

3.2.1 Overview of the ELIXIR Governance framework relevant for the ELIXIR-EXCELERATE project

The formal governance structure of ELIXIR is set out by the *ELIXIR Consortium Agreement (ECA)* with the Member State representatives in the ELIXIR Board as the highest level of authority. The ELIXIR Board is ELIXIR's strategic decision-making body representing the ELIXIR Member States and ELIXIR Member organisations (such as EMBL). The ELIXIR Board appoints and dismisses the ELIXIR Director who is responsible for executing Board decisions and the operational implementation of ELIXIR. The ELIXIR Board also appoints the members of the ELIXIR SAB, which reviews overall ELIXIR activities as well as Nodes annually to ensure scientific impact and excellence. The ELIXIR SAB reports to the ELIXIR Board and the ELIXIR Director.

The Heads of Nodes Committee (HoN) is composed of the Heads of each ELIXIR Node and is chaired by the ELIXIR Director. The Heads of Nodes Committee is the senior scientific decision making body and management team within ELIXIR: together with the Director the HoN Committee develops the ELIXIR Scientific Programme, decides on grant application strategies and oversees ELIXIR technical activities. The HoN Committee takes the leading role in developing the strategy for ELIXIR services, monitoring of performance as well as identification of service gaps. As the senior scientific management team for ELIXIR, the HoN Committee will also maintain the overview of the technical and scientific activities in the infrastructure to ensure scientific excellence and impact.

The *ELIXIR Programme* drives all scientific and technical activities in ELIXIR on a five-yearly basis. It is developed jointly by the ELIXIR Director and HoNs, reviewed by the SAB and agreed with the Member States through the ELIXIR Board. The ELIXIR Director is accountable to the Board for this implementation. The ELIXIR Programme sets out the strategic framework and objectives for ELIXIR: all activities within the infrastructure should align with the programme regardless of whether they are funded nationally, through the ELIXIR core budget or through external grants (e.g. ELIXIR-EXCELERATE).

In addition to the ELIXIR governing bodies identified through the ECA there are additional permanent working groups within the infrastructure:

- Technical Coordinators Group
- Training Coordinators Group
- Data Coordinators group (to be established as part of ELIXIR-EXCELERATE).

The purpose of the ELIXIR Permanent working groups is to identify good practice, national initiatives and challenges and propose implementation projects that support and strengthen the national activities within each Node. Permanent ELIXIR working groups will all establish annual objectives, based on the *ELIXIR Scientific Programme*, that are agreed with the ELIXIR HoN group. The permanent groups report on progress, outcomes and

⁶⁷ DESCA 2020 Model Consortium Agreement: <http://www.desca-2020.eu/>

achievements to the HoN and ELIXIR SAB and in the ELIXIR annual report. The permanent working groups will have their regular meetings funded through ELIXIR-EXCELERATE (WP12). Each of the groups is expected to meet twice annually, in conjunction with other meetings as far as possible to minimise travel.

3.2.2 *ELIXIR-EXCELERATE Governance*

The ELIXIR-EXCELERATE governance will follow the general principles for large projects set out in the DECSA templates with some important adjustments described below (Figure 10).

Project Coordinator - The Project Coordinator (Niklas Blomberg, Director EMBL-ELIXIR), is the legal entity acting as the intermediary between the Parties and the Funding Authority. The Project Coordinator (PC) shall, in addition to its responsibilities as a Party, perform the tasks assigned to it as described in the Grant Agreement and the Consortium Agreement such as overall reporting and financial management of the project.

Heads-of-Nodes Committee – The Heads-of-Nodes (HoN) Committee is comprised of one senior representative per Node and is the legal decision-making body for the project. The HoN Committee will be responsible for decisions on overall financial management and resourcing, changes to project structure and approval of additional partners. The Project Coordinator chairs the Committee. The Committee shall be quorate if two-thirds or more of its members are present or represented. Where necessary, decisions can be passed by written procedure. The HoN committee will meet quarterly with two face-to-face meetings annually organised alongside other meetings. Thus the Head-of-Nodes group, chaired by the ELIXIR Director, will act as the project Executive Board, the supervisory body for the execution of the Project. In the context of ELIXIR-EXCELERATE it will report to the General Assembly and is accountable for overall delivery to the ELIXIR Board through reports and annual SAB review.

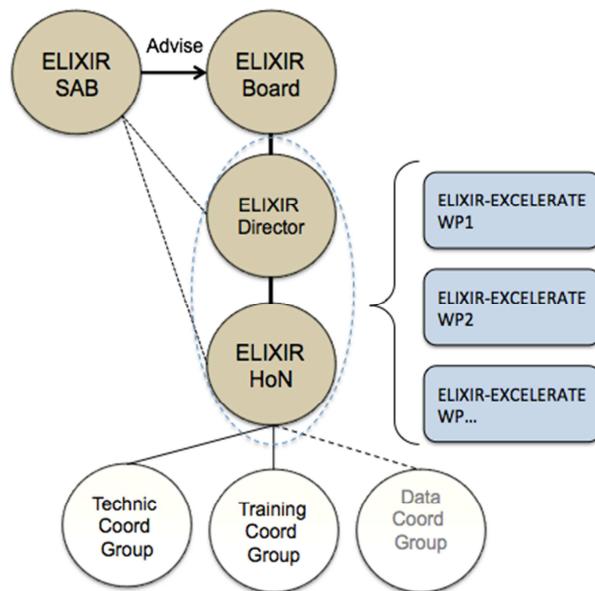


Figure 10: ELIXIR-EXCELERATE governance

ELIXIR Board - The ELIXIR Board is ELIXIR's ultimate decision-making body representing the ELIXIR Member States and ELIXIR Member organisations. The ELIXIR Board consists of administrative and scientific delegates and oversees the financial and scientific execution and impact of ELIXIR considering, as appropriate, advice from the ELIXIR SAB and IAC.

3.2.3 *Advisory bodies*

ELIXIR-EXCELERATE will make use of two pre-existing advisory groups.

ELIXIR Scientific Advisory Board (SAB) – The fully-independent SAB combines the expertise of world-renowned experts to provide scientific strategic feedback to the ELIXIR Board. The SAB has played an integral [ELIXIR-EXCELERATE]

role in the application process for each of participating Nodes and will provide recommendations for the scientific strategy of the ELIXIR-EXCELERATE project.

The ELIXIR SAB includes experts in the biological and biomedical life science area whose backgrounds cover science and industry and are elected by the ELIXIR Board. The SAB meet in person once a year in December and hold one teleconference in summer. They provide direct feedback to the ELIXIR Board.

The ELIXIR SAB currently consists of:

Robert Gentleman, *Genentech, USA* (Chair)

Gunnar von Heijne, *Stockholm University, Sweden* (Vice-Chair)

Ségolène Aymé, *French Medical Research Council (INSERM), France*

Dr Janet Kelso, *Max Planck Institute for Evolutionary Anthropology, Germany*

Edward Marcotte, *University of Texas at Austin, USA*

Nicola Mulder, *UCT Computational Biology Group (NBN), South Africa*

Jérôme Wojcik, *Quartz Bio, Switzerland*

Alan Archibald, *University of Edinburgh, UK*

Elina Ikonen, *University of Helsinki, Finland*

ELIXIR Industry Advisory Committee (IAC) - To promote open innovation and cross-company research partnerships the IAC will provide high-level strategic advice on global industry initiatives. The IAC has been formed from experts from industry, SMEs, suppliers, and publishers and meet once per year.

Philippe Sanseau, *GlaxoSmithKline, UK*

Wendy Filsell, *Unilever R&D, UK*

Angel Pizarro, *Amazon Web Services, USA*

Jakob de Vlieg, *Bayer CropSciences Innovation Center, Belgium*

Martin Ebeling, *Hoffmann-La Roche, Switzerland*

Anita Eliasson, *Biocomputing Platforms Ltd, Finland*

Mark Forster, *Syngenta, UK*

Montserrat Vendrell, *Biocat, Spain*

Natalia Jiménez Lozano, *Bull, Spain*

Iain Hrynaszkiewicz, *Nature Publishing Group - Macmillan, UK*

Claus Stie Kallesøe, *Gritsystems A/S, Denmark*

3.2.4 Operational management of project delivery

The operational management of the ELIXIR-EXCELERATE project is one of the three defined themes outlined in the project's approach (section 1.2.3). Briefly, the ELIXIR Hub will be responsible for the overall management and coordination of the project and will be supported by Work Package 10 that will deliver member value through the deep integration of ELIXIR Nodes. The three permanent ELIXIR working groups will serve to achieve translational objectives between individual Nodes and Work Packages.

Head-of-Nodes / Work Package leads group - The project will operate a Work Package leads group that will consist of the ELIXIR Head-of-Nodes together with Work Package leaders, complemented by invited technical coordinators and expert staff on an agenda-driven basis. The role of this group is to oversee the progress of the project and to ensure interoperability and alignment of co-dependent tasks across project Work Packages and the total ELIXIR programme. It is expected to identify issues, risks and opportunities and take appropriate actions. Risks or opportunities that cut across more than one Work Package will, together with a suggested action, be agreed and addressed in this group by the senior scientific management team within ELIXIR. Matters that cannot

be resolved, or where the Elixir Consortium Agreement requires a Board decision, will be elevated to the ELIXIR Board.

The Group will meet face-to-face for a full day twice per year to allow in-depth discussions into content and progress (AHM + one additional meeting), complemented by quarterly conference calls. In addition, connections, cross-fertilisation and interfaces between Work Packages will be supported by e.g. cross-WP task forces that are focused on specific issues and that have a start, end and clear expected outcome (e.g. deliverable), or ad hoc meetings or phone conferences on specific project areas and topics.

Table 5: ELIXIR-EXCELERATE Work Package Leads

WP Number	Work Package	Lead(s) Organisation and function
1	Tools Interoperability and Service Registry	Søren Brunak , DTU; Professor Alfonso Valencia , CNIO; Director
2	Benchmarking	Alfonso Valencia , CNIO; Director Søren Brunak , DTU; Professor
3	Data Resources and Services	Jo McEntyre , EBI; Team Leader Christine Durinx , SIB; Associate Director
4	Technical Services	Tommi Nyrönen , CSC; Project Director Ludek Matyska , MU; Professor
5	The ELIXIR Interoperability Implementation	Barend Mons , NBIC; Professor Carole Goble , UNIMAN; Professor
6	Use Case A: Marine metagenomic infrastructure as a driver for research and industrial innovation	Nils P Willassen , UiT, Professor Rob Finn , EBI, Team Leader
7	Use Case B: Integrating Genomic and Phenotypic Data for Crop and Forest Plants	Paul Kersey , EBI; Team Leader José Leal , FCG, Principle Investigator
8	Use Case C: ELIXIR infrastructure for Rare Disease research	Ivo Gut , CNAG; Director Marco Roos , NBIC, Investigator
9	Use Case D: An ELIXIR framework for access-controlled secure archive, dissemination and analysis of human data; enabling biobanks, cohorts and local resource services and integration through the European Genome-phenome Archive (EGA)	Justin Paschall , EBI; Team Leader Arcadi Navarro , UPF; Professor
10	ELIXIR Node Capacity Building and Communities of Practice	Jiří Vondrášek , UOCHB; Group Leader Bengt Persson , LIU; Professor
11	ELIXIR-EXCELERATE Training Programme	Chris Ponting , UOXF; Professor Patricia Palagi , SIB; Unit Head
12	Excellence in ELIXIR Management and Operations	Niklas Blomberg , ELIXIR Hub, Director; Project Coordinator
13	Communications, Industry and Community Engagement	Andrew Smith , ELIXIR Hub, Senior External Relations Officer

Work Packages - the Work Package Lead (WPL) and co-leads (as appropriate) will manage individual Work Packages with the WPLs being responsible for reporting and accounting. Work Package leads are jointly responsible for ensuring the integrity of the WP and also that all tasks are completed including the deliverables. They will also prepare the annual report for the project for their own WP.

Work Package Leads will hold regular meetings or conference calls and technical interchanges, as well as staff exchanges as appropriate. The Leads of the common services Work Packages will be responsible for interacting with their related use cases so that a good flow of information is achieved. Similarly, the Leads of the Use Cases - supported by the technical coordinators - will be responsible for ensuring good interactions with the appropriate common services WPs. This communication will be facilitated through the HoN group and the three ELIXIR permanent working groups.

3.2.5 Project meetings

An annual ELIXIR “All Hands Meeting” in spring will bring together all individuals who are active within the infrastructure, and will also function as the project’s annual general meeting. This is a key annual event within the ELIXIR infrastructure that reports on implementation progress, user-projects and impact as well as providing a forum for forward planning.

The purpose of the permanent ELIXIR working groups (Data, Technical, and Training) is to foster collaborative working across ELIXIR Nodes and to identify and spread good practice across the infrastructure. The permanent working groups will have their regular meetings (venue and accommodation) funded. Each of the groups is expected to meet twice annually, in conjunction with other meetings as far as possible to minimise travel.

Kick-off meeting: The project will start with a kick-off meeting held within 3 months after the Grant Agreement is signed in which all project partners and members of the advisory bodies will attend.

All Hands Meetings: Within one year after the Grant Agreement is signed, the first annual All Hands Meeting including all members of ELIXIR and the ELIXIR-EXCELERATE project will be held. The All Hands Meeting will also serve as a meeting point for several other key meetings including that of the HoN Committee and the Coordinators Groups (Table 6).

Table 6: Standing ELIXIR Meetings used in ELIXIR-EXCELERATE

	Attendees	Frequency	Chair	Objective
Heads of Nodes (HoN) Committee	HoN leaders, contributors	Annually at the AHM and quarterly teleconferences	ELIXIR Director	Oversee technical progress of the project and ensure interoperability and alignment of co-dependent tasks across Work Packages
Scientific Advisory Board (SAB)	SAB members, ELIXIR Hub	Biannual, once by TC (summer), once in person (December)	Robert Gentleman	Scientific follow-up and scientific advice in technical meetings
Industry Advisory Committee (IAC)	IAC members, ELIXIR Hub	Annually (first meeting early 2015)	TBC, appointed at first meeting	Strategic advice and industry input
Technical Coordinators (TCG) Group	TCGs from Nodes	Biannual, in conjunction with other meetings e.g. AHM. Monthly TC	Rafael Jimenez, ELIXIR CTO	Foster technological collaboration between ELIXIR Nodes, implement good practice
Training Coordinators (TrCG) Group	TrCGs from Nodes	Biannual, in conjunction with other meetings e.g. AHM. Monthly TC	ELIXIR Training Coordinator (TBC)	Foster training collaboration between ELIXIR Nodes, implement good practice
Data coordinators (DCG) Group	DCGs from Nodes	Biannual, in conjunction with other meetings e.g. AHM. Monthly TC	ELIXIR Data Coordinator (TBC)	Foster data collaboration between ELIXIR Nodes, implement good practice

3.2.6 List of milestones (Table 3.2a)

Milestone (M) Number	Milestone Name	Related Work Packages	Estimated Date	Means of verification
M1.1	EDAM release with coverage of different resource categories and RIs. Implementation of tooling for sustainable community development	WP1	M12, M24, M36, M48	Report to HoN
M1.2	Implementation & evaluation of impact of Resource Pages on resource discoverability	WP1	M12	Report to HoN
M1.3	Implementation of registry-literature integration	WP1	M24	Report to HoN
M1.4	Implementation of support for “close to source” resource annotation in key documentation generators, and software development frameworks	WP1	M24	Report to HoN
M1.5	Good Practice Guidelines	WP1	M36	Report to HoN
M1.6	Implementation of resource metadata catalogue & evaluation of impact of Resource Synergy Meeting series	WP1	M36	Report to HoN
M1.7	Implementation of novel highly usable interfaces from analysis of user experience and usability requirements	WP1	M24, M36	Report to HoN
M2.1	A report on standards needed to organize online performance assessment experiments	WP2	M18	Report to HoN
M2.2	A position paper on online performance assessment experiments	WP2	M24	Report to HoN
M2.3	A report on workshops, hackathons and jamborees held for different user communities	WP2	M48	Report to HoN
M3.1	List of metrics/quality criteria measuring the scientific impact, service usage, service delivery, and eligibility for “ELIXIR Named Resource” and “ELIXIR Core Resource” labels, to allow construction and extension of the ELIXIR Resource portfolio	WP3	M12	Report to HoN
M3.2	List of metrics/quality criteria measuring the scientific impact, service usage, service delivery, which allow ongoing monitoring of ELIXIR Resources as well as their life cycle management	WP3	M12	Report to HoN
M3.3	Report containing formalized metrics and quality criteria to be used in the process of identification and monitoring of usage and scientific impact of ELIXIR Core and Named Resources - recommendation and good practice for ELIXIR	WP3	M12	Report to HoN
M3.4	Plan for collation of metrics and quality data at the ELIXIR Hub	WP3	M24	Report to HoN
M3.5	Computer-assisted service-monitoring tools that deliver metrics data to the ELIXIR Hub	WP3	M48	Report to HoN
M3.6	Report on requirements for an infrastructure to support data enrichment from the literature for researchers and database curators	WP3	M24	Report to HoN
M3.7	Publicly available dataset of GeneRifs with granular article locations detected through the application of text-mining and reported in standard format(s)	WP3	M18	Report to HoN
M3.8	An API to anchor GeneRifs in articles	WP3	M30	Report to HoN
M4.1	Demonstrator I. Some ELIXIR Technical Services are operating and are usable to support aspects of the ELIXIR-EXCELERATE Use Cases. It is not expected that these services will be operating 365/7/24 but resources will have been identified (usually outside of ELIXIR-EXCELERATE) to improve the stability of the services. Change management will be activated with ELIXIR HoN as needed	WP4	M12	Report to HoN

M4.2	Demonstrator II. Most ELIXIR Technical Services operating and in more robust and sustained manner and capable of supporting most aspects of the ELIXIR-EXCELERATE Use Cases. Some services will be capable of operating 365/7/24 and the overall reliability and end-to-end experience will have improved	WP4	M24	Report to HoN
M4.3	Demonstrator III. The ELIXIR Technical Services will be able to support all aspects of the Use Cases (e.g. WP9) and will be able to demonstrate reliability and performance that is applicable across other uses	WP4	M36	Report to HoN
M4.4	Demonstrator IV. Majority of ELIXIR Technical Services are operating providing robust services to multiple research communities. ELIXIR is one of the key European ICT technology drivers and benefits from e-Infrastructure technology advances	WP4	M48	Report to HoN
M5.1	<i>Roadmap</i> of bottlenecks, stewardship standards, dictionaries good practices, services	WP5	M12	Report to HoN
M5.2	<i>First cohort</i> of data templates and data description/identity specifications	WP5	M18	Report to HoN
M5.3	<i>Reference implementation</i> of dataset descriptors, service management and API for three WP3 datasets, with first dataset level BYOD	WP5	M18	Report to HoN
M5.4	<i>Reference implementation of the Interoperability Implementation</i> for testing with WP8 interoperability with first data content descriptors BYOD	WP5	M24	Report to HoN
M5.5	<i>Reference implementation of the Interoperability Implementation</i> for testing with WP7 interoperability with data content descriptors BYOD for Linked Data	WP5	M30	Report to HoN
M5.6	<i>Reference implementation of the Interoperability Implementation</i> for testing with WP6 (Marine) interoperability	WP5	M36	Report to HoN
M6.1	Report from metagenomics data standards environment stakeholder requirements workshop	WP6	M12	Report to HoN
M6.2	Report from metagenomics data standards environment stakeholder review workshop	WP6	M24	Report to HoN
M6.3	Completion of four workshops for end users	WP6	M30	Report to HoN
M7.1	Community workshop held to agree of common ontologies	WP7	M4	Report to HoN
M7.2	First set of agreed ontologies for annotation of phenotype in crop and forest plants	WP7	M6	Report to HoN
M7.3	Agreement and implementation of a common system for sample identification across crop and non-crop species, leveraging existing systems for sample identity management (EURISCO, BioSampleDB)	WP7	M12	Report to HoN
M7.4	Specification of common API for distributed data identification and retrieval	WP7	M18	Report to HoN
M8.1	Publication of the catalogue of resources, data sources and methods assessed in collaboration with the rare-diseases communities	WP8	M12	Report to HoN
M8.2	Open the system for the fast configuration of validation datasets	WP8	M24	Report to HoN
M8.3	Open Access to the ELIXIR rare-disease catalogue in the ELIXIR registry following the appropriate technical standards	WP8	M24	Report to HoN
M8.4	Distribution of the specifications of the technical framework for the standardization rare-disease services	WP8	M36	Report to HoN
M9.1	REST API specifications for submission and data access	WP9	M18	Report to HoN

M9.2	Data submission tools prototype	WP9	M18	Report to HoN
M9.3	Prototype of cloud and remote access systems	WP9	M18	Report to HoN
M9.4	Community access online tutorial and documentation	WP9	M36	Report to HoN
M10.1	Knowledge exchange workshops - Turning Capability into Capacity	WP10	M12	Report to HoN
M10.2	Technical solution for capacity building	WP10	M18	Report to HoN
M10.3	Staff exchange and pilot for technical solution to capacity building	WP10	M24	Report to HoN
M10.4	Community of practice for selected BMS Infrastructures	WP10	M36	Report to HoN
M10.5	Pilot set up of at least one ELIXIR data Node	WP10	M24	Report to HoN
M10.6	At least three ELIXIR data Nodes in action	WP10	M48	Report to HoN
M10.7	1st ELIXIR Advanced workshop on genome annotation	WP10	M4	Report to HoN
M10.8	2nd ELIXIR Advanced workshop on genome annotation	WP10	M22	Report to HoN
M10.9	Network of communities of practices operative with communication facilities	WP10	M24	Report to HoN
M10.10	Creation of a Structural Funds Task Force	WP10	M3	Report to HoN
M11.1	Collect evaluation methods and metrics. Deliver developer and infrastructure operator workshops	WP11	M12	Report to HoN
M11.2	Release TeSS version 1 with a selection of content and first feeds	WP11	M12	Report to HoN
M11.3	Release TeSS version 2 with improvements	WP11	M30	Report to HoN
M11.4	Hold community workshops to survey and discuss e-learning specifications, selection of Use Cases, guidelines and evaluation methods	WP11	M12, M30, M48	Report to HoN
M11.5	Community testing (i.e. running courses) of the e-learning solution and contents. Explore modules for connecting eLearning platform with bioinformatics tools, computing infrastructure and other LMSs. Test the eLearning platform with a selection of face-to-face courses and open educational resources (OER) for the ELIXIR-EXCELERATE Use Cases	WP11	M30	Report to HoN
M11.6	Apply evaluation and benchmarking to training delivery, analyse performance data and reviews	WP11	M48	Report to HoN
M11.7	Review the first year TrT courses and their impact on TtT trainees post course	WP11	M12	Report to HoN
M11.8	Hold a programming workshop and produce the accompanying core materials, building on the strong foundation established in ELIXIR Pilot. Train on ELIXIR infrastructures–HPC, cloud computing and programming– targeting the Use Cases WP6 to 9 users. Expand the virtual community space for developers and infrastructure operators to other training communities	WP11	M30	Report to HoN
M11.9	Hold workshops and TtT courses in universities and industry/SMEs, focusing on: VMs and clouds in training, bioinformatics, Marine Metagenomics (WP6), Plant Genotype-Phenotype (WP7) and Rare diseases (WP8)	WP11	M48	Report to HoN
M11.10	Provide a framework and associated materials that trainers can use to run their own TtT initiatives	WP11	M48	Report to HoN
M12.1	Consortium agreement adopted	WP12	M3	Report to Board
M12.2	Develop and implement a risk management plan to be included into ELIXIR risk register to ensure project objectives are achieved	WP12	M3	Report to Board
M12.3	First version of the ELIXIR intranet ready for testing	WP12	M6	Report to HoN
M12.4	Define permanent working group roles, responsibilities and membership with annual objectives	WP12	M12	Report to Board

M12.5	Develop guidance Chapter in ELIXIR Handbook of Operations on steps Nodes need to follow to participate as single entities	WP12	M12	Report to Board
M12.6	Release of ELIXIR intranet	WP12	M12	Report to Board
M12.7	Meeting: The annual All Hands Meeting	WP12	M12, M24, M36, M48	Report to Board
M12.8	Annual release of the ELIXIR Handbook of Operations	WP12	M24, M36, M48	Report to Board
M12.9	Define performance indicators of permanent working groups to assess effectiveness and impact	WP12	M18	Report to Board
M12.10	Review of equal opportunities within ELIXIR	WP12	M20	Report to Board
M12.11	1 st Workshop on sustainability of data management to collate data from WP3, WP4 and WP10	WP12	M22	Report to Board
M12.12	Workshop to agree scope on the legal framework review	WP12	M22	Report to Board
M12.13	Meeting: Landmark in Bioinformatics Services	WP12	M24	Report to Board
M12.14	Establish ELIXIR's internal processes for ELSI	WP12	M24	Report to Board
M12.15	External expert's report on legal framework review	WP12	M30	Report to Board
M12.16	2 nd Workshop on sustainability of data management to consolidate the data into a high-level strategy to inform ELIXIR Programme	WP12	M34	Report to Board
M12.17	Release of ELIXIR Jobs, events and news portal	WP12	M36	Report to HoN
M12.18	Joint workshop with Helix Nebula on common procurement	WP12	M36	Report to HoN
M13.1	Release of a live ELIXIR Communication Strategy, including annual updates thereafter	WP13	M6, M24, M36, M48	Report to Board
M13.2	Report capturing the social media channels, distribution lists, press contacts to be fed into the ELIXIR Communications Strategy	WP13	M3	Report to Board
M13.3	Develop tender specification for video production by external company	WP13	M4	Report to Board
M13.4	Delivery of video by external company	WP13	M12	Report to Board
M13.5	Mapping report on the key European, national and regional industry associations and the most effective communication channels	WP13	M24	Report to Board and IAC
M13.6	Report on the Key Performance Indicators and metrics that will be applied to understand what constitutes success for the ELIXIR Innovation and SME programme	WP13	M36	Report to Board and IAC
M13.7	Report on community initiatives with which ELIXIR should interact	WP13	M6	Report to Board
M13.8	Annual release of ELIXIR's International Strategy	WP13	M12, M24, M36, M48	Report to Board

3.2.7 Critical risks for implementation (Table 3.2b)

The ELIXIR-EXCELERATE project will see 41 partner Institutions running 13 Work Packages during 4 years. A project of this size with so many components has a number of critical risks relating to project implementation. The overarching management of the project by the already running ELIXIR structure supported by regular meetings and ample opportunities for discussion about progress and potential technical or implementation issues, will ensure that critical risks are identified in an early state and can be mitigated. Per the recommendations set out by the Assessment Expert Group evaluation, ELIXIR already created a risk register, which was included into the ELIXIR Programme and is regularly reviewed and updated. In addition to this, ELIXIR has developed a project specific register for ELIXIR-EXCELERATE, as described in table 3.2b.

A significant advantage of this consortium is that the organisation of the partners is around the national ELIXIR Nodes, themselves bound to the ELIXIR Hub. By coordinating the flow of activities at the national level, the Nodes will ensure the proper execution of the work plan. The Nodes have been working closely together already since the constitution of the ELIXIR network and they are bound through various projects to the success of the ELIXIR project. They have shown to be reliable, accountable, and involve the most creative investigators in Europe working in close collaboration. Moreover, a large part of the resources and infrastructure that will be used in this project are already in place, since ELIXIR-EXCELERATE's objectives are to enhance the implementation of the ELIXIR project. Therefore risks regarding unavailability of resources and infrastructure are considered low. Since the project will apply a number of steps that go beyond the state of the art, there are a number of potential risks in that area, as described in table 3.2b. However, for each of these problems ELIXIR has identified mitigating actions that are already in place or can be initiated if needed.

In addition, management risks must be acknowledged in such a complex and large project. We have identified two major areas in the project that need special attention: **management/operations, communications/industry and community engagement**.

Risks associated with these issues that are confined to two specific Work Packages will be identified, monitored and, if necessary, mitigated directly within the Work Package in question. The project management will identify risks spanning different Work Packages and/or project-wide tasks in close collaboration with the Work Package leaders. Such risks will then be presented to the Heads-of-Nodes committee, which will agree on an action plan. Any risks that cannot be resolved at project management level will be escalated to the ELIXIR Board.

1. Management and operations: A suitable management plan is included in the WP12 (management) description. The coordinator will implement an online platform for decentralized uploading of deliverables, budget information and reports, and to support internal, project-related communication. Given the size of the consortium, for this to work efficiently, the platform to be used will need to be simple and versatile, and all project partners must be comfortable with using it. Valuable experience with such a platform has already been gained during the BioMedBridges project⁶⁸. The different meetings (across Work Packages, Tasks, Project Management Group, etc.) will be scheduled to coincide as much as possible to facilitate communication across the entire project and all Nodes and partners, and save costs. A standard template for meeting notes will be made available by the Project Coordinator. The template will include a list of decisions; identify responsible persons for any necessary follow-up action, and how they will be communicated to the rest of the consortium. After each meeting, a report will be prepared and shared with all other project partners through the internal website, access-restricted to partners. Online communication tools will be used extensively for day-to-day communication especially for smaller meetings at the Task or WP level, which will also reduce cost. A list of Skype names and emails will be created and made available to the participants and published in the internal website.

2. Communications/industry and community engagement: Within the project, optimal knowledge sharing will be supported through scientific meetings, and training workshops organised by WP10-WP11 and scientific exchanges around the scientific Use Cases (WP6 to 9). Dissemination of project results to external users will be ensured through several channels: (a) direct engagement with user groups by the Use Cases within Work Packages 6 to 9, (b) through the activities of the communication Work Package 13, (c) through interactions with related initiatives, and (d) by utilising existing communication channels and industry contacts of the ELIXIR Nodes participating in the project.

⁶⁸ see e.g. <http://www.biomedbridges.eu/sites/biomedbridges.eu/files/documents/deliverables/REPORT%20D1-1%20Website.pdf>

Table 3.2b: Critical risks for implementation

Description of risk	Work Package(s) involved	Proposed risk mitigation measures	Level
Delay in or failure of recruiting suitably qualified personnel for the assigned effort	All WPs	Monitoring of recruitment by the project manager	Low
Failure of effective management of such a large consortium of 41 partners	WP12	The ELIXIR Heads of Nodes will oversee the activities in each country and report to the coordinator and the ELIXIR board. Ensure that the Heads of Nodes and data, training and technical coordinator groups function effectively. Ensure that roles and responsibilities are clearly defined. Together with WP13 (communication strategy) ensure communication across the consortium. One objective of ELIXIR-EXCELERATE is to move towards participation in grants as a single entity	High
External communication failure with stakeholders including users, policy makers, etc.	WP13	Communication is supported by WP13 that develop and implement the communication strategy for ELIXIR. The project will run a large portfolio of representative use cases (WP6 to 9) and industry engagement	Low
Failure to meet users' needs in services	WP1-9, WP11	Close engagement with users to collect frequent user feedback to improve use, value, understanding, access and delivery. Develop metrics to measure usage and impact, regular SAB reviews of services.	Medium
Users loose trust in service delivery, including training, due to data security and technical problems	WP3, WP4, WP11	Build up process for service monitoring and quality control. The ELIXIR core budget is being used to develop ELIXIR data security policy.	Medium
Failure to establish effective interfaces between common services and use cases	All WPs	The 4 Use Cases have been chosen for their capacity to use the entire set of services. Overall integration will be monitored by the Heads of Nodes Committee	Low
Failure to utilise innovation potential	WP13	Work Package dedicated to supporting innovation and engagement towards industry; utilise established ELIXIR Nodes industry connections	Low
Risk on third party loss/damage and IP infringement	WP12	These issues will be set out in a Consortium Agreement following good practice for large project consortia as set out e.g. in the DESCA 2020 templates	Low
Ethical and reputational risks due to services on personal data (ELSI)	WP12	Based on AEG recommendations, ELIXIR is using core funding to develop ELIXIR Ethical Framework	Low

3.2.8 Minimising the carbon footprint

The ELIXIR-EXCELERATE project will follow a carbon emission reduction policy to minimise the environmental impact of its operation via extensive use of telepresence technologies. The project will benefit from the expertise in remote coordination and management of the distributed research infrastructures participating in the project. The majority of project governance activities are expected to perform remotely and significant efforts will be made to hold physical meetings back-to-back in order to minimise overall travel.

3.3 Consortium as a whole

The partners involved in ELIXIR have significant relevant experience with complex projects of a similar scope. For example, most of the HoN group worked together as a steering committee during the ELIXIR preparatory phase project where they have established efficient communication channels on both the management and operational level. In addition this group have worked extensively together through the ELIXIR implementation phase and in establishing the ELIXIR programme. The several of the National Nodes have been in place as national infrastructures for over 10 years: e.g. SIB (CH) was established in 1998, the Norwegian Node in 2002. Based on our in-depth experience with such similar projects, established good practice and in collaboration with a project management consultancy, ELIXIR has developed and refined an organisational and project management structure that will be able to handle the complexity of the ELIXIR-EXCELERATE project.

A key notion is the need to anchor the emerging shared service platform within the organisation and strategy of the established and developing ELIXIR Nodes. Hence the ELIXIR HoNs will lead each Work Package in the implementation project; the ELIXIR HoN group will act as the senior management team as well as Work Package leaders group for the grant and have oversight of both ELIXIR-EXCELERATE deliverables as well as the core ELIXIR activities. The project also makes use of existing strong advisory bodies with world-leading experts to guide projects and impact as well as providing a forum for forward planning.

The main strengths of this consortium are:

- Full coverage of required expertise with recognised expertise in data management and security, ethics, and innovation management as well as the core disciplines
- A strong track record of organising large distributed BMS infrastructure at the European level based on strong local links and national investment
- Strong capabilities and proven ability to deliver and implement training curricula at European level
- International partnership: through the participating Nodes, ELIXIR has strong links to national Nodes of other BMS RI, key international groups and organizations

3.3.1 Gender aspects and promotion of equal opportunities

ELIXIR-EXCELERATE acknowledges that gender equality includes: gender balance in research teams; gender balance in decision-making; and integrating gender/sex analysis in R&I content⁶⁹. The known underrepresentation of women in higher-level positions in the natural sciences⁷⁰ is evident also in the ELIXIR infrastructure, where there is a marked gender imbalance at the PI, coordinator and task leader level. We will address this by promoting equal opportunities not just within the ELIXIR-EXCELERATE project and given time frame, but by creating a dedicated task within the management Work Package (WP12 Task 2.2) to develop best practise recommendation for ELIXIR on gender aspects and equal opportunities. ELIXIR will review the equal opportunity recommendations in access to training and services across ELIXIR Nodes, along with the developing Charter and Code of Access to research infrastructures. The diversity work planned within the CORBEL cluster proposal will provide another aspect to the review, and recommend sustainable strategies for building a diverse work force. In addition, collaborations with the emerging NIH BD2K Biomedical Data Science Training Coordination Centre will be sought, not only on bioinformatics training and shared infrastructure, but also on data science community topics such as gender balance and diversity. The good practise recommendations will be included into the ELIXIR *Handbook of Operations* and will therefore become an integral part of the procedures within the infrastructure. In addition, ELIXIR will aim for an improved gender balance within the consortium by paying special attention to and monitoring the number of women involved in the scientific work and in the decision-making structures, and the

⁶⁹ Gender Equality in Horizon 2020:

https://ec.europa.eu/programmes/horizon2020/sites/horizon2020/files/FactSheet_Gender_2.pdf

⁷⁰ She Figures 2012 – Gender in research and innovation, [Section 3](#)

appointment of qualified women for vacancies within the project. ELIXIR as an infrastructure acknowledges need to be aware of these biases in promoting gender balance and provide clear recommendations for equal opportunities.

3.4 Resources to be committed

To achieve the three goals of the project, the €1905 million funding requested will be spent over four years across the 13 WPs (Figure 11). €12.7 million will cover the costs for personnel for a total of 2099 person-months. The remaining funds are split between other direct costs (€2.3 million), subcontracting (€165,000), third parties (€80,000) and indirect costs (€3.7 million).

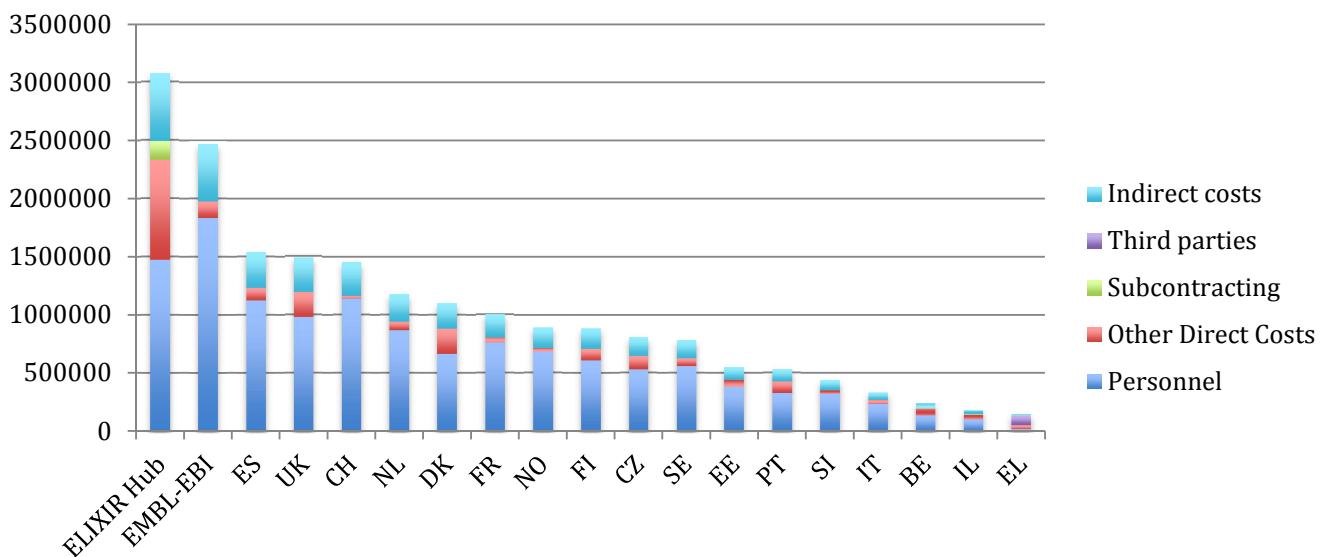


Figure 11: Distribution of funds amongst ELIXIR Nodes.

Personnel resources are distributed as follows across the 3 goals of the project (Figure 12):

- Goal 1: Deliver world-leading data services for academia and industry (WP1-9): 71.2% of the total person-months, including Use Cases (WP6 to 9) with 25.3%
- Goal 2: Increase bioinformatics capacity and competence across Europe (WP10-11): 20.9%
- Goal 3: Complete the management and organisational processes for an efficient distributed infrastructure (WP12-13): 7.7%

Consequently, the primary effort and funding will be for research and development of services and the use cases. This ensures ELIXIR-EXCELERATE remains focused on implementation and providing concrete examples and impact. Administrative related tasks are kept to the minimum level required for effective operations.

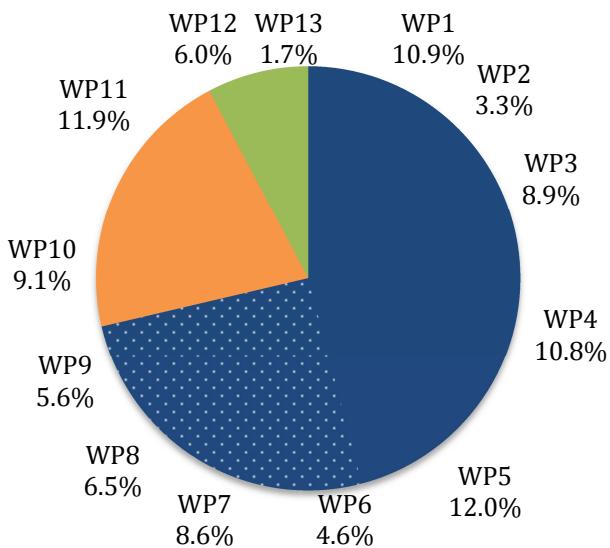


Figure 12: Distribution of resources (Person-months) between the Work Package, coloured according to the project goals: Goal 1 in blue (Use Cases dotted), Goal 2 in orange, Goal 3 in green.

The partners undertaking the main scientific Work Packages are leading institutions in the EU, and indeed worldwide. They bring to the task a wealth of resources, both collections of data and the expertise and knowledge to identify and access collections of value for user communities. This funding request does not include any major items (e.g. equipment) since the partner institutions are well funded through national level investments, possessing state-of-the-art equipment and facilities. The exemplar Use Cases are all independently funded to perform the research and are aligning their efforts in the expectation of creating long-term links and opportunities to explore emerging technologies. These contributions from external partners will therefore provide excellent added value over and above the EU funding, and are consistent with exploiting the value of existing investments and providing scientific projects with new synergies and interactions between them.

External speakers will be invited to the workshops organised by the project, for example in WP6, 7, 10 and 12 to maximise the scientific relevance of the workshop output. In addition, external speakers will ensure that the ELIXIR-EXCELERATE activities are aligned with other projects funded by the EC. The input of the external experts will facilitate the development of existing interoperability and the application of state-of-the-art methodology, which will be instrumental in the Use Cases WPs. Inviting external speakers, will enable reaching out to communities as per the aim of WP10. ELIXIR Board representatives for high-level WP12 working groups are not from partner institutions and will be invited. The travel and accommodation costs will be covered for the external experts in the Work Packages, while no further fees will be charged, and no honorariums will be awarded.

We are seeking to minimise the environmental impact and reduce carbon footprints and contain travel costs of such a large consortium through the co-location of meetings (in particular with the AHM) and extensive use of modern telecommunications and social media tools. All travel costs have been extensively reviewed and exceptions are well justified.

The overall financial plan is fully adequate to carry out the Work Packages proposed. Estimates of person-months needed for each have been agreed with the institutions carrying out the work (i.e. the initial estimation for ELIXIR Nodes has been systematically refined at the level of legal entities). Although the structure of the Nodes means that this project involves 41 legal entities, the existing collaboration between and within Nodes gives confidence that the resources are adequate.

The organisation and expertise of the Nodes within the ELIXIR landscape will provide a robust mechanism for sustainability.

Because of the project's nature, which involves establishing links between ELIXIR Nodes and users, most partners' budgets include either funds to organise meetings or a budget to attend them. Budget of partners of WP6, 7, 8, and 9 include funds to cover consumables for the Use Cases. As a consequence, partners with a low global budget have an "other direct costs" budget ('travel', 'equipment' and 'goods and services') exceeding 15% of the personnel costs. The costs are justified in Table 3.4b.

3.4.1 Summary of staff effort (Table 3.4a)

Sum of Effort (PM)	WP 1	WP2	WP3	WP4	WP5	WP6	WP7	WP8	WP9	WP 10	WP 11	WP 12	WP 13	Total
1 – HUB			8.0								126.0	36.0		170.0
1 – EMBL-EBI	12.0		77.0	34.0	12.0	28.0	12.0		34.0	6.0	14.0			229.0
2 - UOXF	6.0				51.4					4.0	44.0			105.4
3 - TGAC					10.6		21.2							31.8
4 - UNIMAN					30.0			6.0			30.8			66.8
5 - UTARTU	43.0								16.0	20.0	42.0			121.0
6 - NBIC				17.0	38.0		9.0	6.0	5.0		11.5			86.5
7 - CNIO	2.0	12.0	12.0					10.0		1.0	1.0			38.0
8 - CRG	11.0								33.0					44.0
9 – CIPF			3.6							3.6				7.2
10 – IRB	9.0				21.0			13.0						43.0
11 - UMA											2.0			2.0
12 – BSC		19.0			19.0			10.0						48.0
13 – CSIC										2.0	0.8			2.8
14 – UPF			14.1	23.5										37.6
15 - CNAG		9.6						38.4						48.0
16 - IMIM			5.5											5.5

17 – FCG					2.0	2.0			2.0	12.0			18.0
18 - INESC-ID	4.0				3.0				10.0				17.0
19 - IBET						48.0							48.0
20 – CCMAR					11.0								11.0
21 – CSC				48.0		11.0		18.0	4.0				81.0
22 – UiB	18.0								4.0				22.0
23 - NTNU							12.0						12.0
24 – UiT				2.0	36.0				4.0				42.0
25 – UiO					4.0			6.0		2.0			12.0
26 - SIB	9.5	23.0	66.0	2.0					6.0	14.0			120.5
27 - CNRS	9.0			12.0	10.0	10.0	12.0		5.0	4.0			62.0
28 - INRA						24.7							24.7
29 – CIRAD						12.0							12.0
30 - IP	12.0			6.0									18.0
31 – CNR					2.0	10.0	3.0			12.0			27.0
32 – LIU				12.0	32.0			6.0	24.0	4.0			78.0
33 – UL					6.0		15.0		30.0	42.0			93.0
34 – NIB						30.0							30.0
35 - UOCHB					3.2				8.0				11.2

36 – MU	18.2			40.0						26.6	7.2			92.0
37 – CESNET				24.0										24.0
38 – VIB						11.0			10.0					21.0
39 – DTU	76.0	6.0		6.0	10.0		12.0							110.0
40 – BSRCAF									12.0	2.0				14.0
41 – HUJ									8.0	4.0				12.0
Total Person Months	229.7	69.6	186.2	226.5	252.2	97.0	180.9	137.4	118.0	190.2	249.3	126.0	36.0	2099.0

3.4.2 ‘Other direct cost’ items (Table 3.4b)

P1 - EMBL	Cost	Justification
Travel	€ 110,923	Travel to All Hands meetings, WP meetings, specific task meetings, workshops, conferences, travel for external 10 person to participate in the workshops for WP7
	€ 16,000	ELIXIR Innovation & SME forum events, travel for external experts (WP13)
	€ 20,000	ELIXIR representation in conferences, travel for ELIXIR Hub personnel (WP13)
	€ 198,000	Travel for Heads of Nodes, All Hands and coordinators meetings (WP12)
	€ 23,000	Workshops, travel for external experts (WP12)
	€ 42,000	Workshops, travel for external experts (WP13)
	€ 10,000	Travel for external experts to ELIXIR Landmark meeting (WP12)
Equipment	€ 2,300	External storage (for backups and data exchange) WP6
Publications	€ 3,000	Publication for WP6
Other	€ 5,000	Printed materials & banner for each Node (WP13)
	€ 24,449	Kick-off workshops WP11, hosting workshops WP4, WP5, WP7
	€ 40,000	Organisation of ELIXIR Landmark meeting (WP12)
	€ 160,000	Organisation of the annual All Hands meeting (WP12)
	€ 120,000	Organisation of annual data, technical, and training coordinators meetings (WP12)
	€ 37,000	Hosting workshops in WP12
	€ 42,000	Hosting workshops in WP13
	€ 104,000	Organisation of ELIXIR Innovation & SME forum events (WP13)
	€ 40,000	ELIXIR representation in conferences (WP13)
	€ 3,000	Audit certificate
Subcontracting	€ 60,000	Training procurement (WP11)
	€ 40,000	ELIXIR video (WP13)
	€ 60,000	Legal consultancy (WP12)
Total	€ 1,160,672	

P2 - UOXF	Cost	Justification
Travel	€ 2,155	Travel and Subsistence
	€ 21,817	Travel and Subsistence
Consumables	€ 120,000	Workshop
	€ 1,000	Laptop
Total	€ 144,972	

P5 - UTARTU	Cost	Justification
Travel	€ 44,000	Participation at ELIXIR meetings
	€ 13,000	Computers for the personnel
Other	€ 2,000	Audit certificate
Total	€ 59,000	

P7 - CNIO	Cost	Justification
Travel	€ 37,000	Travel costs and accommodation for ELIXIR meetings and workshops

Total	€ 37,000
--------------	----------

P9-FVIB	Cost	Justification
Travel	€ 3,000	Collaborative travels are expected during the project
Equipment	€ 6,000	Replacement of equipment, as well as needs of new disk purchases are expected during the project
Publications	€ 3,000	Publications will be made under open license
Total	€ 12,000	

P11 - UMA	Cost	Justification
Travel	€ 2,000	To attend face to face training group meetings
Other	€ 10,000	To host WP11 UMA HPC workshop
Total	€ 12,000	

P13 - CSIC	Cost	Justification
Travel	€ 12,550	2 meetings per year for two people, including Coordination, plus two additional travels for one person to other Partners or to other meetings.
Total	€ 12,550	

P17 - FCG	Cost	Justification
Travel	€ 10,000	Work Package meetings (WP6,7,10,11)
	€ 6,400	Exchanges/site visits with other ELIXIR Nodes
	€ 13,600	Permanent WP meetings & annual meetings
Equipment	€ 5,000	Two workstations for exclusive use in this project
Total	€ 35,000	

P18 - INESC-ID	Cost	Justification
Travel	€ 22,500	Work Package meetings WP1,5,10
	€ 7,500	International meetings (RDA working groups)
	€ 5,000	Permanent WP meetings & annual meetings
Equipment	€ 5,000	2 workstations for exclusive use in this project
Publications	€ 2,000	Open access journal fees
Total	€ 42,000	

P19 - IBET	Cost	Justification
Travel	€ 4,000	Work Package meetings (WP7)
	€ 800	Exchanges/site visits with other ELIXIR Nodes
	€ 3,200	Annual network meetings
Equipment	€ 3,500	2 workstations that will be used exclusively for this project
Total	€ 11,500	

P20 - CCMAR	Cost	Justification
Travel	€ 4,000	Work Package meetings (wp6)
	€ 3,000	Annual network meetings
Equipment	€ 2,000	Workstation to be used exclusively in this project
Total	€ 9,000	

P21 - CSC	Cost	Justification
Travel	€ 53,400	The travel costs consist of project meetings, participation in relevant conferences and workshops and EC reviews.
Other	€ 20,000	Training costs (WP11)
	€ 5,000	Training costs (WP11)
	€ 20,000	e-Infrastructure events (WP3)
	€ 1,550	Audit
Total	€ 99,950	

P28 - INRA	Cost	Justification
Travel	€ 12,530	One two-days meeting per year with the partners for four persons and one congress for one person for outreach
Total	€ 12,530	

P35 - UOCHB	Cost	Justification
Travel	€ 6,000	8 trips WP5
	€ 10,000	12 trips WP10
	€ 10,000	12 trips multiple WPs -annual ELIXIR meetings
Consumables	€ 2,000	HW for communication platform for WP10
Other	€ 4,000	WP10 two workshops (4000€, in that 2000€ room rent and catering, 2000€ external speakers)
Total	€ 32,000	

P36 - MU	Cost	Justification
Travel	€ 23,600	27 trips multiple WPs: annual ELIXIR meetings (including accompanying WP and task specific events) - approx. 875 per trip (2-4 days) - 500 travel, 250 accommodation, 125 per diems
	€ 1,600	2 trips WP6: coordination workshops - approx. 800 per trip (2-3 days) - 500 travel, 200 accommodation, 100 per diems
	€ 12,800	16 trips WP2: scientific/technical meeting regarding implementation of tools interoperability and service registry - approx. 800 per trip (2-3 days) - 500 travel, 200 accommodation, 100 per diems
	€ 20,000	20 trips WP3: annually one WP coordination workshop, twice task (AAI, clouds) coordination, twice ad hoc technical meetings (3-4 days) - approx. 1000 per trip - 500 travel, 350 accommodation, 150 per diems
Consumables	€ 2,000	WP2: HW upgrades necessary for running ELIXIR tools and databases
Other	€ 4,000	WP6: two workshops (4000€, in that 2000€ room rent and catering, 2000€ external speakers)
	€ 6,000	WP7: training workshop for users of ELIXIR tools (6000€, in that 4000€ external speakers, 2000€ room and catering)
	€ 3,000	Mandatory audit (CFS)
Total	€ 73,000	

P38 - VIB	Cost	Justification
Travel	€ 37,500	Project meetings + dissemination @ international conferences (personnel paid on the project + own resources personnel; Yves Van de Peer,

		Frederik Coppens, Lieven Sterck, Nathalie Wuyts, Stijn Dhondt)
Equipment	€ 7,500	SQL server
Consumables	€ 11,000	Standard consumables dry lab (software, licences)
Total	€ 18,500	

P39 - DTU	Cost	Justification
Travel	€ 160,000	Costs of participants in hackathons; participation in scientific meetings.
Consumables	€ 54,000	We plan to arrange two hackathons per year, each hackathon is estimated to be 13,500 EUR (4 hackathons in total).
Other	€ 3,500	Audit certificate
Total	€ 217,500	

P40 - BRSCAF	Cost	Justification
Travel	€ 32,000	Participation of EL-Node representatives to ELIXIR meetings and of Node personnel into workshops organised by ELIXIR. 10 such participations per year at €800 per travel
Total	€ 32,000	

P41 - HUJ	Cost	Justification
Travel	€ 6,000	Meetings for Technical and academic
	€ 6,000	2 people for 2 meeting a year
Equipment	€ 5,000	Servers for large scale data
Consumables	€ 10,000	Storage, movable
Publications	€ 4,000	Publication fees
Other	€ 4,000	Communication, office
	€ 5,000	Office, editing support
Subcontracting	€ 5,000	Fellowships for students support
Total	€ 45,000	