ELIXIR Scientific Programme

2024-28   Full Programme

Contents

Foreword

Dear friends and colleagues

The Scientific Programme is a key document for ELIXIR. It sets out our vision for the coming five years and drives the strategy and priorities for our work within and between ELIXIR Nodes.

This document describes the details of our strategy and ambitions developed following extensive input from ELIXIR Heads of Nodes, Platforms, Communities and other expert working groups. A more concise version can be found in our Executive Summary brochure[1]. The ambitions and outcomes set out here have been used to develop the Financial Plan 2024-2028 and the corresponding annual budgets and work plans.

The ELIXIR 2024-28 Programme builds on many conversations in these working groups since early 2021. Starting with a formal round of consultations with the Nodes, Platforms and Communities (structured teleconference interviews with Heads of Nodes and other senior colleagues), we then held rounds of discussion with the Heads of Nodes group and the ELIXIR Scientific Advisory Board (SAB). At the end of 2021, we held a second series of open consultations, this time topic based, for instance to explore our plans for research data management and our links with industry. Moving into 2022 we held our first discussions with the ELIXIR Board in April, further discussion with the Heads of Nodes at the Prague retreat, and finally an event in Brussels in July with 30 external experts joining us to further help refine our plans. In addition, Node experts in the ELIXIR Platforms and Communities have been working hard to consolidate the ideas into detailed plans that set out how we will deliver on the ambitions.

The major theme of the Programme reflects the successes of the first ten years: ELIXIR will now operate, in production, a connected federated ecosystem for life science data rooted in common values and mutual trust. While much work remains to connect all national Nodes to this federated ecosystem, the transition to large-scale operation is a major success, reflecting the strong collaboration between Nodes in our first decade of operation. From a scientific perspective, human data and translational research will continue to be high priority, but ELIXIR will now also have a formal focus, with increased investment, in a joined-up effort that addresses the three key societal challenges of biodiversity, food security and pathogens.

Molecular data is at the heart of ELIXIR. The ELIXIR 2024-28 Programme maintains an emphasis on molecular and cellular sciences, underlining the importance of these areas as the source of scientific advancements in response to societal challenges. The focus also reflects ELIXIR's core mission to support scientists in curiosity-driven and investigator-led research in molecular and cellular biology, and to find innovative solutions to data from the exciting new methods and approaches in the field. We will address the data needs of this exciting research and help to translate it into solutions for health, the environment and food security.

Bringing life to this Scientific Programme has been a shared endeavour, and I would like to thank everyone who has contributed in the many discussions, workshops and drafting efforts.

Niklas Blomberg Director, Cambridge UK, November 2023

---

[1] https://elixir-europe.org/sites/default/files/documents/elixir-scientific-programme-2024-28-brochure.pdf

Together we accelerate the understanding of life

ELIXIR is rooted in the power of bioinformatics to transform data into biological insights, driven by advancements in the molecular and cellular sciences. ELIXIR's core mission is to support scientists in curiosity-driven and investigator-led research in molecular and cellular biology and help translate these advancements into sustainable solutions for food security, the bioeconomy, our health and the environment.

Well annotated, accessible, and ready-for-reuse data and software resources are the foundation of modern biology. The Human Genome Project ushered in a new era of data-led science aiming to map living organisms in molecular detail. It has been followed by many large-scale projects addressing crops, such as wheat and rice, and key organisms for molecular biology research such as mouse, rat and fruit fly. These projects demonstrate the power of the international consortia model to comprehensively map molecular functions and release large-scale datasets for use by the whole research community.

Collaborative genome sequencing led the way for large-scale data acquisition and drove technology development. The projects pioneered the coordination of international efforts based on national laboratories with open science as an organising principle, fostering innovation and enabling wide-spread participation. Equally important, these endeavours laid the technology foundation for routine large-scale measurements by developing the data management practices and software tools underpinning data-driven biology.

Today, most of Europe's research centres have the capacity for large-scale data generation enabling data-led experiments with large, comprehensive and complex data sets. Large-scale data collection is now routine: genome sequencing of new species, mapping of tissues at single-cell resolution, and tracking biological processes over time and in the context of whole ecosystems. As a result, we have an unparalleled collection of data, and ELIXIR's role is to provide life scientists with the means to find, access, integrate and analyse this data in the context of their own experiments. They must also be given the support, recognition and incentives to make their own data findable, accessible, interoperable, and reusable (FAIR) by others.

Putting these data to good use is challenging. Life science data are complex and making use of the data for secondary analysis, large scale modelling and advanced learning requires that the experimental context, limitations and possible bias are well characterised. To harness its full power, data must also be connected across scientific disciplines, driving the need for new partnerships and technological solutions. Careful cataloguing, curation and metadata annotation of data and software is critical to unlock all potential value. Modern life science builds on good research asset management. Scientists must be able to easily annotate the outcomes of each experiment and find and access data from the larger community to put results in context. FAIR scientific data and software are not only transforming life science research but feeding into society as a whole. The widespread calls for data from policymakers and international organisations show how politicians, government and the public have an expectation that science will provide the answers to the challenges societies face.

Societal challenges and scientific opportunities

We believe well-managed and open data, tools and standards are the lifeblood of future scientific progress, enabling collaborative and equitable research to meet the challenges facing global society.

Enabling and responding to advances in cellular and molecular research will drive the development and impact of life science data. Cellular and molecular research are in a state of accelerated development, with advances being driven by new experimental techniques. The rapid pace of development is on course to continue, and will impact both basic science and applications across the biomedical and life science domains. Exciting developments abound: the combination of imaging and molecular data now provide opportunities to study molecular processes in the cells, tissues and organisms where they take place.

Molecular sciences connect biodiversity, food security and pathogens. The application of molecular techniques is critical to understand the diversity and breadth of life on Earth. Driven by a rapidly changing climate and increasing contact between the natural world and humans, emerging infectious agents are very real threats for humans, farmed animals, crops and wildlife.

As a result, we have the challenge of supplying enough food to feed the whole of humanity. Solutions are likely to centre on engineered and improved crops, environment, agricultural practice, interaction between ecosystem and agriculture enabled by molecular sciences. Examples include the use of genomics to breed drought and disease tolerant crops in response to changes in climates and ecosystems, the role of the microbiome in healthy agricultural and native ecosystems, and the direct and indirect role of these ecosystems on human health.

The promise of genomics research to improve health and disease outcomes. A greater understanding of disease and the mechanisms that cause disease, brings an opportunity to advance personalised medicine and prevention approaches. With a greater understanding of the biological mechanisms of disease, comes a huge opportunity to make predictions. For diseases caused by a single gene this can be highly accurate. Understanding common, complex diseases can be more of a challenge as study cohorts need to be large and diverse. Equitable genomics research and healthcare can be improved. However, challenges remain around ensuring citizen engagement: genomic and other omics data could be used beyond health, raising ethical questions, risk of discrimination or threats to privacy.

Connecting data is the key to progress in the life sciences. To understand life we need to study biology at all scales: from the molecular details of protein binding to predict the risks of new viral variants, through studies at the cellular, organism and ecological level, to studies of global ecosystems to help understand and mitigate the impact of climate change on our environment, livelihood and health. This is where open, accessible research data and software makes a difference. Through the work with research communities, ELIXIR has early sight of new techniques, the impacts that they are likely to have, and the requirements they will put on data and research infrastructures. In the 2024-28 Programme, we aim to support the projects and technologies making the most difference to the challenges faced by all.

ELIXIR connects data, tools and countries to address societal challenges

ELIXIR's ambition is for Europe to have a robust, connected network of established national services, internationally recognised resources, and federated enabling technologies for data-centric life science research.

ELIXIR Platforms develop infrastructure as a distributed, pan-European network of services. To maximise the potential of scientific data, researchers need infrastructures accessible at scale. Data, workflows, models, analytics and software need to be interoperable and work across borders. ELIXIR has built a robust, decentralised model where national services can be accessed across national Nodes. As well as cross-European collaboration, each jurisdiction must continue to provide local resources, for example to comply with data sovereignty or funding constraints. The Federated European Genome-Phenome Archive (EGA) exemplifies this trend - with a federated network of resources to respect national data sovereignty, together with a robust legal and technical interoperability framework to allow cross-jurisdiction research. ELIXIR coordinates this distributed, pan-European network of services via the ELIXIR Platforms. The ELIXIR Platforms bring together experts from Nodes to develop ELIXIR's technical vision and drive the implementation of a common roadmap across five key infrastructure areas: Data, Tools, Interoperability, Compute and Training.

ELIXIR continues to build a European network of biological data resources and scientific software using common technical standards and best practices. ELIXIR Nodes support national research data management practices and help the mobilisation of data from national data processors to international repositories. We will build on our experience in open-source software (the 4OSS principles[2]), support software management plan development (though the Data Stewardship Wizard[3]), and use the structured description of algorithms (the DOME[4] recommendations for machine learning) to promote best practices and develop software stewardship as a companion activity to data stewardship. Using this approach, core software components become FAIR and connect to the tools ecosystem.

ELIXIR promotes sustainability, standards compliance and quality service provision, exemplified by the ELIXIR Deposition Databases[5] (EDDs), Core Data Resources[6] (CDRs) and Recommended Interoperability Resources[7] (RIRs) which have become a model for the Global Biodata Coalition and adopted by funders worldwide.

ELIXIR Communities connect the infrastructure services to research domains. ELIXIR Communities organise scientific or technology experts within a field of research. Communities are key contributors to ELIXIR, driving the development of standards, services and training in and across ELIXIR Platforms, and serving as the link to the wider research community. ELIXIR Communities currently cover a wide

---

[2] https://softdev4research.github.io/recommendations/
[3] https://ds-wizard.org/
[4] https://dome-ml.org/
[5] https://elixir-europe.org/platforms/data/elixir-deposition-databases
[6] https://elixir-europe.org/platforms/data/core-data-resources
[7] https://elixir-europe.org/platforms/interoperability/rirs

scientific spectrum, ranging from research fields such as structural bioinformatics, plant sciences, rare diseases and microbial biotechnology, through technological research communities such as proteomics, metabolomics, and the Galaxy workflow system.

ELIXIR is built on strong national Nodes. Tackling these challenges and maximising the associated opportunities requires human capital and strong Node operations. This Programme therefore has a strong focus on training and service sustainability. We must ensure that our Nodes, and all the individuals within them, are well equipped to run sustainable services and to participate in a distributed infrastructure. ELIXIR Nodes are made up of many different people: technical and scientific specialists such as data experts, curators, software engineers, scientific managers and web developers, through to those running the administration and professional operations of the infrastructure, including legal experts, project managers, impact evaluators, communications and outreach experts and governance experts. In this Programme we hope to provide opportunities, training and development for everyone, growing our culture of learning, knowledge exchange and support.

Priorities and ambitions of the 2024-28 Programme

Together we accelerate the understanding of life

ELIXIR's purpose is to accelerate the understanding of life by facilitating data-intensive research. Scientists should be able to analyse their data in the context of known research, build large, multidisciplinary datasets for advanced modelling and easily collaborate across borders.

ELIXIR's 2024-28 Scientific Programme is constructed from four tiers, each building on the one beneath. Each tier maps directly to a strategic goal of the 2024-28 Programme (Figure 1).

The top tier is science - our users. ELIXIR serves the life science community by promoting FAIR access to scientific data and by the provision of services for use by the community. As Europe's distributed research infrastructure for life science data, our purpose is to bring together national data services to advance research by enabling scientists to access and analyse life science data of unprecedented scale and complexity - deploying community agreed standards and easily accessing best practice in research data and software management. In 2024-28 Scientific Programme ELIXIR will continue to forge close partnerships with life-science research communities.

During the 2024-28 Programme, ELIXIR will operate, in production, a collaborative, connected ecosystem for life science data rooted in common values and mutual trust. ELIXIR's ambition is to connect life science data across countries and make this data FAIR. The data should be usable - and used - for new scientific research as well as for the broader society (for example, in education). We have structured the requirements for this collaborative, connected ecosystem for life science data in the three remaining tiers: technology, Nodes and people.

The foundation tier for ELIXIR's operations is our people. To successfully build a federated ecosystem that supports scientists across the whole of Europe each of our national Nodes needs highly qualified personnel to operate national endpoints in the federated infrastructure, for research data management and other advanced user support. This requires skills that are in short supply, we therefore run an ambitious training and capacity building programme which will be further expanded in the next programme.

ELIXIR is a distributed infrastructure based on Nodes - a coordinating centre for each ELIXIR member. ELIXIR Nodes operate the services that allow data to be discovered, accessed and analysed. The Nodes provide data management expertise and services nationally and support scientists with the brokering of data into international deposition databases. In addition, the Nodes provide national experts and agree on the scientific services and data resources that the country will contribute to the international infrastructure "commons". ELIXIR Nodes are funded via national funders, often through national research infrastructure roadmaps. A distributed infrastructure builds on national operations, we will therefore invest in our Nodes and the services they run.

Effective collaboration requires agreement on standards and technology for accessing and analysing data. ELIXIR develops shared technical platforms for research data management, discovery and

access, for tools and analysis and for user access to services across Europe building on the services and data resources within our Nodes. In the 2024-28 Programme we take the next step and put in production a collaborative, connected federated ecosystem for life science data rooted in common values and mutual trust.

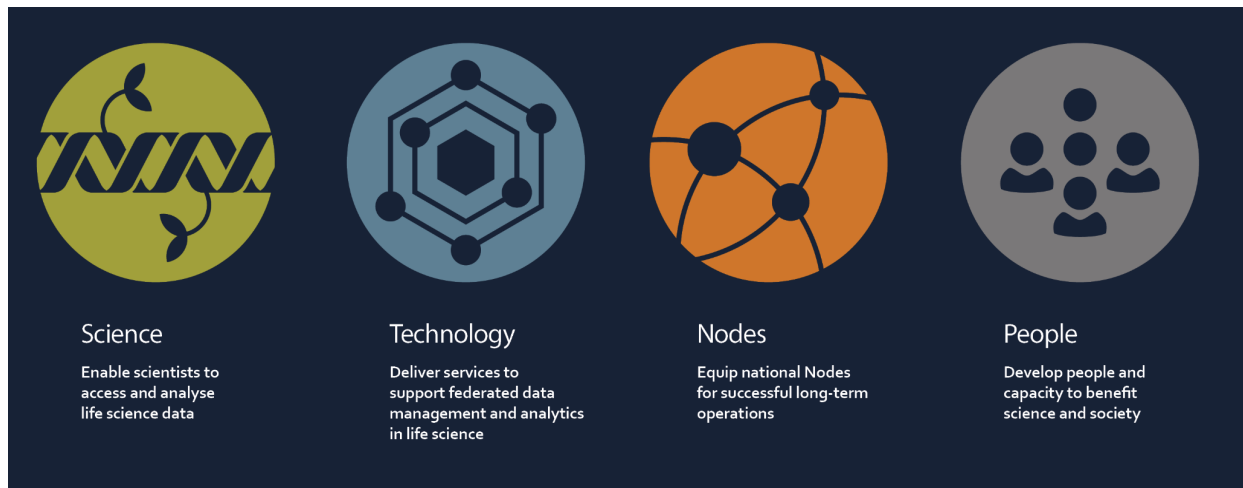*Figure 1 - The four tiers making up the ELIXIR infrastructure and their mapping to the four 2024-28 Programme priorities*

ELIXIR's 2024-28 strategic priorities

1 Enable scientists to access and analyse life science data

2 Deliver services to support federated data management and analytics in life science

3 Equip national Nodes for successful long-term operations

4 Develop people and capacity to benefit science and society

	ELIXIR Scientific Programme 2024-2028 The European life sciences data infrastructure			
Identity	**We enable scientists to access and analyse life science data**			
Values	• **We work to benefit everyone** - we enhance the quality, efficiency and novelty of life science research, leading to improvements in health and wellbeing • **We are trusted** - our services and resources are trusted and we are trusted to deliver on our commitments • **We work with a spirit of openness** - we are committed to Open Science, promoting sharing of knowledge and information throughout the discovery process • **We strive for excellence** - we provide a world-class service to support world-class science • **We work in an environment of respect** - we create an environment where everyone can be their best and enjoy their work			
Priorities	**Science** ENABLE SCIENTISTS TO ACCESS AND ANALYSE LIFE SCIENCE DATA	**Technology** DELIVER SERVICES TO SUPPORT FEDERATED DATA MANAGEMENT AND ANALYTICS	**Nodes** EQUIP NATIONAL NODES FOR SUCCESSFUL LONG-TERM OPERATIONS	**People** DEVELOP PEOPLE AND CAPACITY TO BENEFIT SCIENCE AND SOCIETY
Focus areas	Cellular and molecular research	Research data management and knowledge sharing	Node operations	Users of ELIXIR services
	Biodiversity, food security and pathogens	Reproducible analytics and infrastructure	National research and open science alignment	Managment and operational staff
	Human data and translational research	Federated service delivery	Industry and innovation	Technical and scientific staff
			Impact and sustainability	A diverse network of people

Figure 2 - A summary of the 2024-28 Programme priorities in the context of ELIXIR's purpose, identity and values

The 2024-28 Programme presents a series of new ambitions and desired outcomes. These are set in the context of the ELIXIR Platforms and ELIXIR Communities, whose successful structure will remain broadly unchanged. Focus Groups of experts from within and outside ELIXIR will be convened when new technologies, areas of practice, or partnerships are identified as being of strategic interest but not yet represented by Platform or Community activities.

Platforms

In the new Programme, ELIXIR Platforms will continue to manage the technology governance and architectural structure of ELIXIR services. Platforms will continue to be shaped by leading Node members who develop and operate services in order to address ELIXIR shared technical priorities.

In order to implement the ambitions of the 2024-28 Programme, the five Platforms (Data, Tools, Compute, Interoperability and Training) must continue to work closely together. Each Platform will be led by an executive committee of Node experts approved by the Heads of Nodes Committee, together with a coordinator appointed by the ELIXIR Hub. The Platform membership will continue to

include task leads from within the Programme and from affiliated driver projects, technical contributors from the Nodes, and technical and scientific staff from the Hub.

In the 2024-28 Programme, ELIXIR Platforms will oversee targeted technical development and information sharing workshops, review novel developments from projects across Europe and beyond, and recommend the adoption of new approaches and best practices to ELIXIR Node and multi-Node services. They will identify the ELIXIR infrastructure services and prioritise investments, and work together with the ELIXIR Communities and driver projects to identify gaps and requirements for new research and development.

Communities

ELIXIR Communities are communities of practice, bringing together experts from within and outside ELIXIR in a specific scientific or technology domain, using and contributing to ELIXIR services.

ELIXIR has seventeen Communities, working together in a variety of areas, with a main focus on either science or technology. In the 2024-28 Programme, each Community remains led by Community co-leads, supported by the ELIXIR Hubs Communities team. Community membership will continue to consist of the co-leads, Commissioned Service members, technical contributors from the Nodes, and technical and scientific staff from the Hub.

In the new Programme, it is important that ELIXIR Communities represent users and are co-developers of ELIXIR services. They will continue to lead ELIXIR Commissioned Services, coordinate research efforts, connect initiatives within ELIXIR and beyond, and bridge to European (and where necessary, international) projects and organisations. Their members represent ELIXIR and its services to the wider research community, they are ELIXIR's domain experts and work together with ELIXIR Platforms to drive projects to identify gaps and requirements for new research and development.

Science tier - Enable scientists to access and analyse life science data

Cellular and molecular research

Ambition 1.1 Connect the latest developments and established data resources to realise the potential of cellular and molecular biology

ELIXIR is rooted in bioinformatics, and cellular and molecular biology is core to our activities. Fundamental, curiosity-driven science is critical to finding solutions to the major challenges facing humanity today. The provision of a data-centric infrastructure for the rapidly developing molecular and cellular sciences is a core part of the 2024-28 Programme, covering both basic and applied research perspectives.

We will enhance our data services and software based around nucleic acids, proteins and other biomolecules to accommodate the latest developments in technology and analysis techniques. Cellular and molecular biology will create new demands for multi-omics and multi-modal analysis and we will develop methods and partnerships to address these. Modelling at all scales will become an increasing challenge and we will extend our expertise in reusable data and software to include FAIR models.

The challenges and opportunities

Enabling and responding to advances in cellular and molecular research will drive the development and impact of life science data. Research advances enabled by data are generating new techniques, which in turn increase the rate of data generation, putting the biosciences in a state of accelerated development. Developments in next generation sequencing, long read sequencing, mass spectrometry and imaging continue to allow exponential growth in data volumes and associated computational requirements. This will impact both basic science and applications across the biomedical and life science domains.

New advances and insights will be found through the integration of multi-modal data. Cellular and multi-omics approaches are now routine in fields from developmental biology to pharmaceutical target validation. Many techniques combine cell imaging data with molecular characterisation, for example in the growing technology family of spatial omics, combining multi-omics bioinformatics with multi-modal image interpretation. Multi-modal techniques and samples from diverse sources, including clinical and diagnostic applications in biomedicine, will require many of these analyses to be carried out in secure computing environments. As genomics data become more accessible, an increased number of genetic associations and polygenic risk scores will be uncovered. Cellular and molecular data, such as proteomics and other omics data, will be critical for researchers to translate these findings into understanding the molecular mechanisms of disease.

Novel modelling techniques rooted in large high-quality datasets are enriching our understanding of biology. Remarkable successes in artificial intelligence (AI) and machine learning (ML) such as the recent development of AlphaFold and the AlphaFold Protein Structure Database[8] have been enabled by the high quality, high volume data collected in the life sciences and the skills and computational infrastructures to analyse them. Rates of data generation and an increasing demand for rapid analysis and interpretation will require AI and ML to become part of the bioinformatician's standard toolkit. Along with ML approaches, our growing, linked datasets will support modelling approaches such as systems biology for health and bioindustrial applications and support for digital twins across scales from molecules to ecosystems.

Plans for 2024-28

ELIXIR can strengthen links between the core molecular and cellular data resources and connect technologies to accelerate impact. In the third Programme, ELIXIR will influence the development of data representation and analysis ecosystems for new experimental techniques and research practices that are interoperable, open, and FAIR by design. We will encourage the use and usability of existing ELIXIR Deposition Databases, contribute our expertise to the development of standards for new integrated datasets and support the development of reusable software and the sharing of bioscience data as FAIR digital objects across public and industrial research.

We will enhance ELIXIR services to create a connected network of data resources and software based on a shared understanding of biological concepts and biomolecular representation. This will be achieved through co-development and knowledge sharing, connecting the ELIXIR Communities and ELIXIR service teams, under the technical guidance of the ELIXIR Platforms. This will also seed the development of new ELIXIR services in support of integrated cellular and molecular biology.

ELIXIR will support standardisation and share best practices in multi-modal methods and knowledge representation in molecular structure, imaging and multi-omics technologies. Successes in structural techniques from three-dimensional bioinformatics, proteomics, and intrinsically disordered proteins will be extended to other molecule types such as nucleic acids and other biomolecules. Developments in multi-omics will link approaches in genomics, proteomics and metabolomics with single-cell and spatial techniques. We will partner with key European Strategy Forum on Research Infrastructures[9] (ESFRI) partners to achieve our goals. Euro-Bioimaging[10] will be our primary partner for microscopy and imaging data from the scale of molecules to organs and organoids, and Instruct-ERIC[11] will be our partner for data derived from biomolecular structure determination.

We will extend our tools, techniques and practices for FAIR and open data and software to the development, management and sharing of FAIR models developed through ML and systems biology approaches. We anticipate large volumes of high quality time-resolved data - developmental and metabolic - and the challenges posed by large volumes of spatially complex four-dimensional data,

---

[8] https://alphafold.ebi.ac.uk/
[9] https://www.esfri.eu/
[10] https://www.eurobioimaging.eu/
[11] https://instruct-eric.org/

and will continue to develop and enhance our data and analysis capabilities in face of this change. In this area we will again seek to partner with other ESFRI life science infrastructures, in particular, IBISBA[12] (the European Research Infrastructure for Industrial Biotechnology) for biotechnology applications and the application of modelling techniques.

Ambition 1.1 Connect the latest developments and established data resources to realise the potential of cellular and molecular biology

- A comprehensive set of open solutions is available to develop, manage and publish FAIR by design datasets and digital objects (for example, tools, workflows, data, models, operating procedures, case studies) for reproducible research in new areas of bioscience
- New and established ELIXIR services for cellular and molecular biology form a semantically interoperable and complementary network of FAIR data and tools which can be readily discovered, understood and integrated into novel analyses
- Demonstrated and comprehensive collections of interoperable tools, data, knowledge and training are available to researchers working on cross-modality problems involving structural, omics and imaging datasets in cellular and molecular biology
- Co-developed projects and knowledge sharing are established with EuroBioimaging and INSTRUCT research infrastructures to answer multimodal research questions
- FAIR collections of machine learning and systems biology models, supporting data and tools are routinely available, and an ecosystem of reusable models is available to answer basic and applied research questions
- ELIXIR services are in place to support the needs of the European Research Infrastructure for Industrial Biotechnology (IBISBA) research infrastructure and the European bioindustry

---

[12] https://www.ibisba.eu/

Biodiversity, food security and pathogens

Ambition 1.2 Mobilise and integrate life science data to support transnational research programmes in biodiversity, food security and pathogens

The areas of biodiversity, food security and pathogens represent critical societal challenges that Europe must address over the next decade, and which are strongly enabled by the application of molecular sciences and other data-intense disciplines. The importance of these scientific areas in contributing to societal challenges is also widely recognised by major transnational and national funding bodies and is reflected in their research strategies.

In the third Programme we propose to apply the operational model successfully used to coordinate human genomic data across biodiversity, food security and pathogens. There are a number of possible mechanisms to explore, including strategic investments in the Nodes to develop and enhance services operating across these scientific areas, and the coordination of externally funded projects with the Hub or a Node in the coordinating role.

Biodiversity, food security and pathogens link to interrelated societal challenges

We are at a critical point in being able to preserve the planet's rich biodiversity. If action isn't taken over the next decade, it seems almost certain the Earth will enter a sixth great extinction. The solutions to this challenge are highly complex, span many disciplines and will be as much about mitigating climate change as preserving wild habitats. The application of molecular sciences to assist the understanding and measurement of biodiversity is increasingly important. Indeed, it is likely that the application of molecular techniques is the only way by which we can ultimately understand the diversity and breadth of life on Earth and so devise effective mitigations to the challenges we face.

Arising out of the diversity of life and compounded by the increasing contact between the natural world and industrial methods of food production are the challenges presented by the recent COVID-19 pandemic and other emerging infectious agents, impacting humans, farmed animals, crops and wildlife. The pandemic has shown the versatility of ELIXIR services - applying technology and making infrastructure available rapidly and reliably to support a previously non-priority area. ELIXIR and its members have established themselves as a respected and trusted entity in infectious disease, health and pandemic preparedness.

The COVID-19 pandemic provided an opportunity to make strides in federated data sharing across country borders, and acted as a use case to test high performance computing applications in biology. Disease specific services such as the Infectious Diseases Toolkit[13] and the COVID-19 Data Portal[14] have emerged as key resources supported by ELIXIR and the European Commission (EC) funded projects it coordinates. The application of such toolkits and services will be ever more important to build on lessons learned when future pandemics emerge.
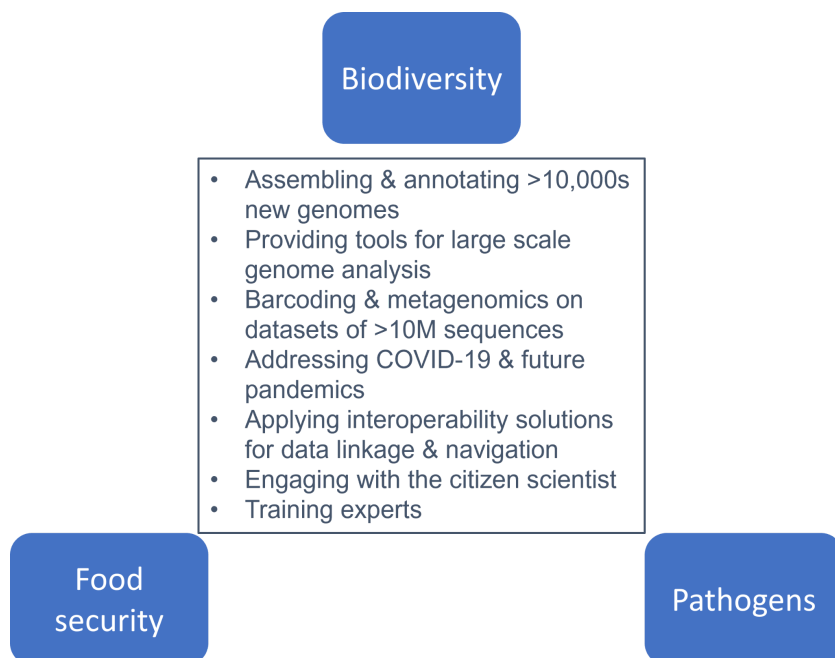
---

[13] https://www.infectious-diseases-toolkit.org/
[14] https://www.covid19dataportal.org/

Finally, and again closely related to the natural world, there is the challenge of supplying enough food for a growing population. Here, solutions are likely to centre on engineered and improved crops, continual development of agricultural practice and further understanding of the interactions between ecosystems and agriculture (agroecology). These solutions will be enabled, in part, by molecular sciences and facilitated through ELIXIR's tools and services. Examples are the use of genomics to understand the role of the microbiome in healthy agro and wild ecosystems, and breeding drought and disease tolerant crops in response to changes in climates and ecosystems. The direct and indirect (for example, through food) role of these ecosystems on human health, are aspects to which molecular sciences and bioinformatics can contribute.

Biodiversity, food security and pathogens are connected challenges with complementary solutions

Biodiversity, food security and pathogens are immense challenges for Europe and the rest of the world. Any potential solutions will be complex, multidisciplinary and long term. All three are strongly enabled by, and dependent upon, the molecular sciences and associated bioinformatics services. By taking a coordinated and aligned approach across the bioinformatic needs of biodiversity, food security and pathogens, ELIXIR can establish a collection of services, openly available, and widely applicable that complement and enhance each other. Figure 3 shows some ways how these areas overlap.

```
                    ┌──────────────┐
                    │ Biodiversity │
                    └──────────────┘

        • Assembling & annotating >10,000s
          new genomes
        • Providing tools for large scale
          genome analysis
        • Barcoding & metagenomics on
          datasets of >10M sequences
        • Addressing COVID-19 & future
          pandemics
        • Applying interoperability solutions
          for data linkage & navigation
        • Engaging with the citizen scientist
        • Training experts

  ┌──────────┐                    ┌───────────┐
  │   Food   │                    │ Pathogens │
  │ security │                    └───────────┘
  └──────────┘
```

Figure 3 - The inter-relationship of ELIXIR's scientific priorities

In terms of coordinating across these areas, ELIXIR is at an early phase of work. However, within each of these disciplines, much has already been achieved in the past few years both technically and

organisationally. ELIXIR is involved in multiple EC funded projects[15], and we have well established Communities in Plant Sciences, and Food and Nutrition. We also have a rapidly developing Community in Biodiversity and strong connections with many other infrastructures active in these areas (for example, DiSSCo[16], GBIF[17], LifeWatch ERIC[18], CoL[19], EMPHASIS[20], ANAEE ERIC[21]).

The beginning of the 2024-28 Programme will necessarily focus on defining a more detailed coordination and joint investment plan, much as was done for human data in the first ELIXIR Programme, before we make a suite of investments starting in mid-2024.

Ambition 1.2 Mobilise and integrate molecular data to support transnational research programmes in biodiversity, food security and pathogens

In the areas of biodiversity, food security and pathogens:
- A coordinated and comprehensive portfolio of activities is built via ELIXIR Communities (existing and newly created) and externally funded projects (national and transnational)
- Nodes broker biomolecular data into accessible reference collections for the research community
- Nodes are established as integral components of national data management programmes
- Nodes run a suite of services allowing molecular, cellular and imaging data to be routinely mobilised, connected and analysed
- Large and well-curated datasets are available for artificial intelligence analysis, high performance computing service use and advanced analytics
- Collaborations link to other disciplines and infrastructures to facilitate seamless data management, connectivity and analysis for the research community

---

[15] https://elixir-europe.org/about-us/how-funded/eu-projects
[16] https://www.dissco.eu/
[17] https://www.gbif.org/
[18] https://www.lifewatch.eu/
[19] https://www.catalogueoflife.org/
[20] https://emphasis.plant-phenotyping.eu/
[21] https://www.anaee.eu/

Human data and translational research

Ambition 1.3 Provide the infrastructure to support the discovery, access, sharing and analysis of human genomics data and linked phenotypic/other data on a massive scale

Human data is one of the key pillars of ELIXIR, a priority area since the first Scientific Programme in 2014. The mission to construct and operate a sustainable infrastructure for human genomics and translational data in Europe to support life science research and its translation to medicine continues from our previous Programme. The vision to facilitate discoverability, access, reception, storage and analysis of genomics data linked to other data types at an unprecedented scale is ongoing. Critical to this Programme is ELIXIR's contribution to demonstrating the impact of infrastructure in translating genomics research into medicine.

ELIXIR aims to be the sustainable coordinator for data management and analysis in the life sciences. For the infrastructure to operate in a federated landscape it is essential to coordinate across nationally funded services and community endorsed standards that can be managed in a workflow across data management requirements. Development of these standards and services, via ELIXIR Node technical experts, needs to be collaborative and transparent. ELIXIR, operating through its Member States, is well placed to develop such a human data management landscape.

The challenges

Progress towards effective data management of virtual cohorts of genomics data from millions of human participants has been made, but neither the scientific nor medical communities are able to use these data effectively and therefore it remains a significant area of priority. Also challenging is bridging across the needs of research and healthcare sectors that have largely developed infrastructure and solutions independently of one another, often using non-interoperable commercial software solutions. In today's era of genomic medicine, this needs to change.

Genomic technologies have advanced and it is projected that future DNA sequencing is most likely to be generated for healthcare in a clinical setting (for example, diagnosis, prevention, clinical trials and treatment) and not directly for research. Yet, research and innovation in the genomics field is critical to understanding disease and pharmacogenomics. To drive personalised medicine approaches in healthcare, research findings need to be integrated into the clinic. Likewise, findings from genomics in a healthcare setting can do much to facilitate research and innovation. To advance the genomics ecosystem, secure and scalable technologies need to be deployed and functional across these two domains.

The opportunities

In 2018 several European countries agreed, by signing a declaration, to step up efforts towards creating a European data infrastructure for genomic data and implementing common national rules

enabling federated data access. By 2022, 24 countries[22] are signatories to the declaration with most being ELIXIR Member States.

The declaration led to the launch of the 1+ Million Genomes (1+MG) initiative[23], which ELIXIR is committed to help implement due to shared objectives. In 2019, the 1+MG initiative recognised ELIXIR as a neutral broker to coordinate activities across the Member States and bring knowledge and learnings on developing infrastructure for management of sensitive human genomics data.

ELIXIR coordinated the Horizon 2020 project Beyond 1 Million Genomes[24] (B1MG) between 2020 and 2023, developing the framework, guidelines, recommendations and best practices required to enable large scale access to human genomes across countries, including developing a maturity model to aid implementation of genomics into healthcare systems.

ELIXIR was mandated by the 1+MG signatories to continue as a neutral broker and coordinate a Digital Europe project to establish a European Genomic Data Infrastructure[25] (GDI). The four-year project, which started in November 2022, brings together 20 countries, 18 of which have committed to be "fully operational/production ready" within the 1+MG infrastructure by the end of 2026 to enable routine access to genomic and phenotypic data across borders. ELIXIR will work to provide tools to support Node advancement and infrastructure-wide interoperability.

The ELIXIR infrastructure is focussed on research data and has not developed frameworks to be embedded into healthcare systems. For any healthcare system, the integration of genomic data for personalised medicine purposes requires extensive adjustments, including, but not limited to, the development of technical infrastructure, competencies and ethical, legal and social implications (ELSI) frameworks for acquiring, storing, sharing, interpreting and delivering genomic information.

Embedding infrastructure into healthcare is likely to be a national responsibility and out of scope for ELIXIR, although to ensure interoperability, collaboration with other research infrastructures (for example BBMRI-ERIC[26], ECRIN[27], EATRIS[28]), sectors (for example, industry), and standards organisations (for example, GA4GH) will continue to take place and will be further developed. This will happen in the context of EC projects such as GDI and through national programmes such as Health-RI [29] in the Netherlands. ELIXIR can adapt where necessary to support the needs of the evolving 1+MG initiative and will consider the interface to healthcare during this Programme.

Over the course of the 2024-28 Programme, frameworks and infrastructure in development or use for human genomics data can be reused for other European data spaces. The implementation of the European Health Data Space[30] (EHDS) will progress, the main objective of which is to develop

---

[22]Austria, Belgium, Bulgaria, Croatia, Cyprus, Czechia, Denmark, Estonia, Finland, Germany, Greece, Hungary, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Norway, Portugal, Slovenia, Spain, Sweden, UK.
[23] https://digital-strategy.ec.europa.eu/en/policies/1-million-genomes
[24] https://b1mg-project.eu/
[25] https://gdi.onemilliongenomes.eu/
[26] https://www.bbmri-eric.eu/
[27] https://ecrin.org/
[28] https://eatris.eu/
[29] https://www.health-ri.nl/
[30] https://www.european-health-data-space.com/

sustainable cross-border linkage of and access to a multitude of interoperable health datasets across Europe. Genomics data is a specific data type that will be included in the future EHDS, and whilst components of the EHDS can be managed independently, the data infrastructure needs to be interoperable to ensure maximum impact and utilisation. ELIXIR is leading a use case as part of the EHDS2 HealthData@EU[31] (secondary use) pilot project that will complete in 2024 and will introduce the requirements for federated genomic infrastructure in the larger context of EHDS developments.

Plans for 2024-28

ELIXIR will continue to build a comprehensive portfolio of infrastructure services, active partnerships with research communities and global strategic collaborations.

The ELIXIR Human Data Communities (HDCs), established in the previous Programme, ensure co-development and a long-term European strategy for management of sensitive human data. The three HDCs, Federated Human Data (FHD), Rare Diseases (RD) and human Copy Number Variation (hCNV), coordinate experts and interface with other ELIXIR Communities. They work with strategic projects such as Beacon and with partners such as GA4GH.

The aims of the Communities and the Beacon strategic project are:

FHD Community: enable sensitive human data to be internationally accessible for example via Federated European Genome-Phenome Archive (FEGA)
RD Community: create a federated infrastructure to enable researchers to discover, access and analyse different rare disease repositories across Europe
hCNV Community: implement processes to make the detection, annotation and interpretation of copy number variations easier
Beacon project: enhance data discoverability by developing the Beacon Protocol and Beacon Network

In the new Programme, progress will be made via externally funded projects to enable transnational discovery, access and analysis of human genomics and associated phenotypic data. Scientific use cases will be a focus, ensuring that services are available to specific research communities such as cancer, rare disease, common complex disease and infectious disease.

It is critical that HDCs maintain the flexibility and agility to build and operate a sustainable human genomics data infrastructure in Europe to support life science research and translation into medicine at scale, while also ensuring data that can be shared is shared in accordance with data protection laws, such as General Data Protection Regulation (GDPR) and ELSI policies to maintain participant trust.

Ambition 1.3   Provide the infrastructure to support the discovery, access, sharing and analysis of human genomics data and linked phenotypic/other data on a massive scale

---

[31] https://ehds2pilot.eu/

- Researchers (including research clinicians) use ELIXIR's infrastructure, operating within the European data space landscape, for human genomic, phenotypic (including broader set of high throughput omics), imaging and demographic data to support discovery, analysis, innovation and integration of research findings into the clinic and healthcare
- Millions of human genomes are discoverable and exploited in a biomedical setting through ELIXIR-supported infrastructure and community-endorsed standards, software, workflows and analysis environments across ELIXIR Nodes
- Federated analysis of data is demonstrated for use, for example, in rare disease, cancer, common complex disease, preventative disease, pharmacogenomic studies
- Partnerships linking human data to other disciplines and infrastructures are well established and mechanisms for data deposition are defined and agreed across Nodes
- Industry supports the deployment of the EU data spaces by providing solutions based on ELIXIR-supported standards

Technology Tier - Deliver services to support federated data management and analytics in life science

Research data management and knowledge sharing

Ambition 2.1   Provide data and knowledge management infrastructure to support open, data-driven research in the life sciences

ELIXIR's core purpose is to provide researchers with the infrastructure to enable the creation, analysis, management and re-use of biological data. Our members represent many of the world's leading centres of expertise in life science data, and we have an unparalleled opportunity to use the ELIXIR structures to develop and share that expertise with researchers throughout our membership and beyond.

In the 2024-2028 Programme, we will promote closer connections between the Core Data Resources and the consumers and contributors of their data; we will put the tools and expertise developed by the large scale data projects in the hands of individual research groups at all scales; and we will develop and promote our tools and expert networks to share best practice and technology in the creation and reuse of FAIR and open data.

The challenges and opportunities

The ELIXIR Core Data Resources represent a unique portfolio of critical biodata repositories which underpin global life science research over a variety of data types. The ELIXIR Deposition Databases extend these to provide a range of end points for most of the core research workflows in action today. The Recommended Interoperability Resources, include the standards, interchange formats and registries required for these services to work together. ELIXIR therefore provides many of the world's critical capabilities for integrated data-centred biological research. We must sustain this rich portfolio and make sure users can find, understand and reuse all data objects through registries, training and communication.

We need to put expertise in the creation and sharing of data and software in the hands of researchers. Life science researchers should be empowered to participate fully in every stage of the lifecycle of FAIR data production and reproducible analysis, and be connected to local, national and international technical capabilities and expertise. The ELIXIR data manager expert network developed in ELIXIR-CONVERGE and the Biocuration Focus Group, together represent the community of experts and users who are responsible for the creation and maintenance of ELIXIR's data resources, large and small.

Increasingly, smaller database projects are becoming more important; either as sources of novel data types based on new laboratory techniques or as curators of specialised, locally restricted data sets

such as clinical genomics. Wherever the data are, or come from, we need to support the researchers responsible for its management. The construction and use of aggregation databases to connect data types will enable new insights, and will need investment in the specialist communities who can create them.

ELIXIR's tools play a key role in building capacity and sharing best practices. ELIXIR has developed a range of tools based on the expertise of large data facilities which put the ability to create and share high quality FAIR data into the hands of individual researchers. The RDMkit[32], developed under ELIXIR-CONVERGE, and the FAIR Cookbook[33], developed during the FAIRplus project, form the nucleus of a growing framework of tools for knowledge sharing and research data management (RDM) expertise tailored for particular user requirements. Together with the Data Stewardship Wizard [34], a tool for data management planning, and FAIRsharing[35], the registry for databases and standards, ELIXIR has the basis of a complete research data management knowledge ecosystem for life science data stewards and practitioners. ELIXIR needs to sustain and develop these into a resilient, comprehensive capability.

Plans for 2024-28

ELIXIR will continue to build a European network of biological data resources using common technical standards and best practices, and the registries and knowledge resources which allow them to be discovered and connected. We will support and promote the CDR forum of experts responsible for large scale resources, and provide a voice for European data providers on a global stage. Node services which represent CDRs, EDDs and RIRs will be promoted through our research infrastructure and international partners.

We encourage the use of recognised deposition databases by their promotion, extending the portfolio of EDDs, and by investing in support for novel data types - in particular, supporting developments in experimental technology and the new data types and challenges they present. We will invest in technologies that allow the CDRs and EDDs to become more inclusive and better connected. ELIXIR Nodes will increasingly support brokering of data from national data processors to international repositories as we develop and promote expedited pathways for researchers to deposit their data through Node-supported data brokering models.

We will use the toolkit of curation, data representation, interoperability and application programming interface (API) technologies developed for our large scale resources to encourage the development of high quality smaller specialised databases for specific communities. This work will require the development of the skills of biocurators or the addition of biocuration skills to the capabilities of many more researchers. We will also develop techniques and tools to integrate hard-to-reach sources of public data, for example the long tail of data in supplementary material and simple datasets deposited in generic archives such as Zenodo[36], into some of the CDRs. These activities will increase the diversity of contributors to our datasets. Mechanisms to recognise original contributions and the

---

[32] https://rdmkit.elixir-europe.org/
[33] https://faircookbook.elixir-europe.org/content/home.html
[34] https://ds-wizard.org/
[35] https://fairsharing.org/
[36] https://zenodo.org/

value added by research participants are required, from sample collection to data curation, methods development and analysis. We will continue to develop our approach to recognition and data citation so that aggregation becomes a positive and impactful route to re-use for the providers of component data types.

We will provide comprehensive guidance for data stewardship practitioners in the life sciences by continuing to develop the RDMkit, the FAIR Cookbook, and the wider FAIR services ecosystem. We will promote FAIRifying data resources and encourage a mindset that data should be FAIR by design. To support the expansion and quality of training in biocuration, data stewardship and research software engineering alongside other bioinformatics skills we will enhance the 'train-the -trainer' programme to support training providers with tools for data and software centred-training and common approaches to e-learning; develop a dynamic European registry of trainers, using metrics for training delivery as part of a researcher's contribution recognition; and develop curricula on data stewardship as a transferable competency framework.

We will support the recognition of training activities, including the design of resources, making resources FAIR, maintaining teaching material, and adding entries in training registries. Through this we will develop a framework for certification of supported training material and events and contribute to the recognition of training as a practice and the certification of training as part of a researcher's continued professional development. We will develop standard guidelines on the skills, responsibilities and recognition of all of the professionals involved in biocuration, data stewardship, and research software engineering in the life sciences, in dialogue with our partner international organisations - the International Society for Biocuration[37], the Research Data Alliance[38], and the Research Software Alliance[39]. ELIXIR members will be encouraged and supported in their contributions to these organisations, and technologies and best practices will be shared in both directions.

Ambition 2.1 Provide data and knowledge management infrastructure to support open, data-driven research in the life sciences

- Our key data services, represented by the Core Data Resources forum, have strong connections through ELIXIR Nodes to user and developer communities and international expert groups
- The human and machine usability of our large scale resources is enhanced and data deposition is enabled through standards and brokering
- Our research data management infrastructure allows researchers and research communities at all scales to easily and cost-effectively make all Europe's life science data FAIR, enhancing the usability of large scale resources and supporting community-driven data curation
- The ELIXIR knowledge management infrastructure drives attribution and recognition of data depositions, biocuration, software development and other open science contributions using well-recognised global frameworks

---

[37] https://www.biocuration.org/
[38] https://www.rd-alliance.org/
[39] https://www.researchsoft.org/

- The ELIXIR research data management knowledge ecosystem is sustained, enabling the connection of new services and the maintenance and co-creation of new content
- A strong and extensive training programme is an integral part of the research process, co-developed by life scientists and service developers, supporting all aspects of life science knowledge management, and connected to appropriate international expert groups such as the Research Data Alliance, the Research Software Alliance and the International Society for Biocuration

Reproducible analytics and infrastructure

Ambition 2.2 Develop our infrastructure to enable end-to-end management of life science software, workflows and containers

Research in the life sciences is characterised by complex, multi-step workflows using a range of different representations, data management technologies, software architectures and compute environments. ELIXIR's users need support in understanding, finding, using and contributing to this endeavour in recognised and reproducible ways.

In the 2024-28 Programme we will develop a full software lifecycle foundation, from the development of software management plans to the registration and archiving of software and algorithms. We will extend and standardise the technologies and registries for the description and deployment of complex multi-platform workflows; and we will give users access to the resources they need from the growing portfolio of cloud and high performance computing services available to research.

The challenges and opportunities

We need a solid foundation of practices and tools to support the software lifecycle. ELIXIR has developed a Tools Ecosystem, since rebranded the "ELIXIR Research Software Ecosystem",[40], which acts as a metadata exchange platform, to coordinate registries and services for software, using standards such as Common Workflow Language (CWL)[41], RO-Crate[42], EDAM[43], Bioschemas[44], and biotoolsSchema[45]. Building on the content aggregated and curated over the last years, the ELIXIR Research Software Ecosystem makes data accessible beyond API calls and serves as the central tool-metadata hub. This needs to be sustained and adopted into practice, and extended to cover the growing complexity of research software.

We must support the complex workflows our users develop and depend on. Researchers and large scale projects need solutions to support the rapid development, deployment and scale-up of data-centred software, workflows and containers, taking account of data governance and legal aspects, in a consistent and adaptable way. ELIXIR has supported this through many activities, including EOSC Life[46], WorkflowHub[47], the Galaxy analytics ecosystem[48], the ELIXIR::GA4GH Federated Analysis Systems Project (FASP) Cloud and AAI (Authentication and Authorization Infrastructure) driver project[49], a range of applications in human data, and implementation studies on

---

[40] https://elixir-europe.org/internal-projects/commissioned-services/tools-platform-ecosystem/
[41] https://www.commonwl.org/
[42] https://www.researchobject.org/ro-crate/
[43] https://edamontology.org/page/
[44] https://bioschemas.org/
[45] https://github.com/bio-tools/biotoolsSchema
[46] https://www.eosc-life.eu/
[47] https://workflowhub.eu/
[48] https://usegalaxy.org/
[49] https://www.ga4gh.org/driver_project/elixir-cloud-and-aai-for-human-data/

container deployment[50] [51]. Containers can be built from BioContainers[52], registered in WorkflowHub and linked to the Galaxy Community, and integrated with Life Science Login[53] for identity and access management. This work is complex, and needs to be developed further and simplified to be put in the hands of research teams.

ELIXIR has led efforts with EOSC[54] (European Open Science Cloud) partners to coordinate technical, operational and funding provisioning of hybrid cloud solutions for data and compute services in Europe, focussing on federating private and national community clouds with public clouds. One of the key enablers for federated analytics is ensuring our Node experts are connected to the development of the EOSC. Our 2022 ELIXIR strategy for EOSC[55] has focussed on participation in the task forces of the EOSC Association[56] along with other activities through the EOSC Partnership[57].

For example Life Science Login is a foundational, federated service in which ELIXIR's vision and technology have given us a leadership role. We can continue to play an active role in influencing the EOSC Strategic Research Roadmap and Innovation Agenda[58] (SRIA-MAR). Key gaps to be addressed in this ecosystem include an interoperable cost transfer/payment system, access control (including ongoing GA4GH passport integration) and deployment for sensitive data use cases, and the extension of the approach to other providers of large scale e-infrastructure for research, particularly public cloud and high performance computing facilities.

Plans for 2024-28

We will build on our experience in open-source software, support for software management, and the structured description of algorithms to promote software lifecycle best practices and develop software stewardship as a companion activity to data stewardship. Using this approach, core software components become FAIR and connect to the ELIXIR Research Software Ecosystem. We will promote the software management plan, and use interoperable data approaches such as RO-Crate FAIR Digital Objects[59] for supporting reproducible software packaging. We will continue to promote and develop standards for registration of re-usable *in silico* methods, for example recommendations for the description of machine learning (based on the DOME recommendations[60]). In the ELIXIR Research Software Ecosystem, ELIXIR's core software component registries (bio.tools[61] and Biocontainers) and a framework for software quality assessment and benchmarking (OpenEBench[62]) cross-reference and specify software, versioning, inputs and outputs and we will continue to develop these into complex workflows. We will build training delivery programmes to ensure ELIXIR Node staff have the skills

[50] https://elixir-europe.org/internal-projects/commissioned-services/making-container-services-integratable/
[51] https://elixir-europe.org/internal-projects/commissioned-services/2022-ecp3/
[52] https://biocontainers.pro/
[53] https://lifescience-ri.eu/ls-login/
[54] https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science/european-open-science-cloud-eosc_en
[55] https://zenodo.org/record/7120997#.ZMEcXXbMK16
[56] https://eosc.eu/
[57] https://www.eosc.eu/partnership
[58] https://www.eosc.eu/sria-mar
[59] https://doi.org/10.3897/rio.8.e93937
[60] https://dome-ml.org/
[61] https://bio.tools/
[62] https://openebench.bsc.es/

required so that interoperable tools can be more easily combined and deployed as complex workflows over distributed data.

To support the development of workflows, we will continue to develop some of our core technologies. The WorkflowHub registry developed through the Tools and Interoperability Platforms and EOSC Life project will become part of the ELIXIR Research Software Ecosystem, using CWL, Bioschemas and EDAM as the description framework and packaged using RO-Crate, an implementation of FAIR Digital Objects (FDO) for workflows. We will extend its integrations with Galaxy, Sapporo-WES[63] in Japan and direct collaborations with the Australian BioCommons[64] and the US-led Workflows Community Initiative[65]. Sustaining the service will require the involvement of additional ELIXIR Nodes.

The GA4GH Workflow Execution Service[66] (WES) sends Task Execution Service[67] (TES) demands to distributed analysis points for real world implementations, and features in ongoing developments in the CINECA project[68], the 1+MG initiative, and the BY-COVID project[69] to enable secure, transnational access to human cohort data. Ongoing work will integrate GA4GH passport standards into workflow services using ELIXIR-driven AAI expertise. New projects such as GDI will accelerate the implementation of these proofs-of-concept in real world use cases in the nations and regions and enable the growth of federated analytics. To put these developments in the hands of large numbers of users, the ELIXIR Galaxy Community provides highly scalable open APIs to over one hundred thousand users and we will work to extend support for the GA4GH cloud APIs by the software developers, IT operations and system administrator stakeholders who build and maintain these popular services.

We will promote the adoption of standards and new technologies allowing Nodes to collectively provide solutions for advanced analyses with the goal of internationally federated cloud and high performance computing (HPC) services. At the European level we will grow and develop partnerships in the provision of large scale computing and storage and trusted research environments with advanced HPC providers alongside EOSC through EuroHPC[70], leveraging ELIXIR Node participation, and key government-industry partnerships such as GAIA-X[71] where we will also develop our representation.

In collaboration with the Galaxy Community and with the support of HPC centres, ELIXIR will establish a legal framework and accounting mechanisms and help develop a more unified strategic approach taking into account new EC procurement models. The new Programme will share best practices such as the development of improved user interfaces to support users by the de.NBI cloud[72] infrastructure from the ELIXIR Node in Germany and the Ga4GH-enabled Hypatia Cloud[73] of ELIXIR

---

[63] https://github.com/sapporo-wes/sapporo/
[64] https://www.biocommons.org.au/
[65] https://workflows.community/
[66] https://ga4gh.github.io/workflow-execution-service-schemas/docs/
[67] https://ga4gh.github.io/task-execution-schemas/docs/
[68] https://www.cineca-project.eu/
[69] https://by-covid.org/
[70] https://eurohpc-ju.europa.eu/index_en/
[71] https://www.data-infrastructure.eu/GAIAX/Navigation/EN/Home/home.html
[72] https://www.denbi.de/cloud/
[73] https://hypatia.athenarc.gr/index.php

Greece. Use cases identified across the Nodes will suggest further directions for development and common features for presentation and adoption of ELIXIR's Compute infrastructure.

Ambition 2.2 Develop our infrastructure to enable end-to-end management of life science software, workflows and containers

- A mature ecosystem is in place to support the software lifecycle from the software management plan to deployment and long-term retention of code
- Research software engineers in the life sciences have guidelines, training and registries to ensure software and workflows are open and reproducible, managed using the techniques developed for FAIR and open research data
- Our work on co-developing standards with organisations such as GA4GH enables the composition of authentication, data, compute and workflow services using only open, standards-based APIs
- Dialogue with the international Galaxy developers enables the adoption of open interfaces, increased integration of ELIXIR services, and adoption throughout the ELIXIR Communities addressing their diverse needs including sensitive workflows
- Established protocols, services and technical standards together with availability of key data resources allow life science researchers to make effective use of Europe's new and expanding high performance computing facilities (EuroHPC) as well as cloud services (EOSC, GAIA-X).
- Accounting for resource use across infrastructure services allows service providers and research users to connect resources to requirements based on agreed priorities

Federated service delivery

Ambition 2.3 Develop a federated pan-European technical infrastructure for life scientists to access data, storage and compute services

ELIXIR has developed as an infrastructure based on the complementary strengths of its Node services. Our ambition is for European researchers to have straightforward access to a robust, connected network of established national services, internationally recognised resources and enabling technologies for data-centric life science research.

In the 2024-2028 Programme we will develop ELIXIR services as composable elements based on semantic interoperability and the adoption of standards-based APIs; we will recognise, enhance and promote the core services and service collections to provide the common foundation for diverse research programmes; and we will sustain and strengthen the key enabling services on which the others depend through a federal delivery model for governance and operation.

The challenges and opportunities

ELIXIR has grown as a network of national services, complementing each other and providing an unparalleled portfolio of data, expertise and technology capabilities with capacity building enabled by its international scale. Through ELIXIR-EXCELERATE[74], ELIXIR-CONVERGE[75] and previous ELIXIR scientific Programmes we have built a set of complementary services for bioinformaticians based on the strengths of our ELIXIR Nodes. The challenge now is to bring them together into a robust network using shared approaches and enabling technology. A focus on implementing standards allows the easy confederation of services. For example, the ELIXIR AAI, the base for Life Science Login for researcher identity and access management, is firmly established in day-to-day operations and support provision across the life sciences.

ELIXIR's implementations of the GA4GH standards including passports, data repository and cloud services have demonstrated their utility and influenced their design. These become services, for example through the Beacon network, promoting the finding and re-use of shareable genomic data. The Federated EGA exemplifies this trend - with a federated network of resources to respect national data sovereignty, together with a robust legal and technical interoperability framework to allow cross-jurisdiction research. Likewise, in plant sciences, the dissemination of standards (MIAPPE[76] and BrAPI[77]) in linked databases and tools ease data integration and exchange.

ELIXIR has highlighted the role of sustainability, standards compliance and service focus in biological knowledge sharing, exemplified by the ELIXIR Deposition Databases, Core Data Resources, and Recommended Interoperability Resources. The Core Data Resource approach has become a model for the Global Biodata Coalition[78] and adopted by funders worldwide. Our commitment to

---

[74] https://elixir-europe.org/about-us/how-funded/eu-projects/excelerate
[75] https://elixir-europe.org/about-us/how-funded/eu-projects/converge
[76] https://www.miappe.org/
[77] https://brapi.org/
[78] https://globalbiodata.org/

interoperability has supported the development of Bioschemas to allow biological knowledge to be found and linked between diverse resources. We need to continue to identify, promote, and find ways to sustain the data-centred resources on which life scientists depend.

Increasingly, services developed by individual Nodes are being shared, adapted and reused between Nodes, or operated by multiple Nodes working together. For example multiple ELIXIR Nodes run services based on FAIRDOM-SEEK[79]; ELIXIR partners have built and together run the Training e-Support System (TeSS)[80], allowing researchers to browse, discover and organise life sciences training events and material, aggregated from ELIXIR Nodes and third-party providers. TeSS represents a unique capability adopted by infrastructures worldwide. The shared operation and continued co-development of services offer opportunities for a more resilient service network, but they require the development of new models for governance and support which ELIXIR can facilitate.

Plans for 2024-28

In the 2024-2028 Programme, ELIXIR will develop its infrastructure as a distributed, pan-European network of stable, consolidated national and multinational services avoiding unnecessary redundancy and duplication, increasingly forming a coherent federation. Each jurisdiction will also continue to provide local resources to comply with data sovereignty or funding constraints. To enable this, we will build a robust, decentralised architecture which accommodates both models - dependable international services with interoperable interfaces, and common standards and maturity models to support national service development and provision. This will require a strong sustainability framework which recognises the necessity for blended funding from multiple sources but also leverages the advantages of ELIXIR and the ELIXIR Nodes as long-term, strategic partners with the ability to contribute maintenance, service provision and enabling activities for priority services.

Member State commitment to the national funding of resources will be recognised and the impact of contributions to shared, Europe-wide capabilities measured and communicated. For example, trusted research environments for sensitive data are a developing area of national capability in which some Nodes will be able to act as hosts for protected data types, from highly sensitive clinical-genomic data to intellectual property in the biotechnology industries. We will enhance capabilities among our leading centres of expertise in this area, and develop packaged solutions, our trust framework, training networks, guidelines and reusable operational frameworks to increase the capability of individual Nodes and services.

We will continue to identify, promote and find ways to sustain the data-centred resources on which life scientists depend. We will support research workflows requiring specialist collections of different recommended services, from ELIXIR and beyond, to cover all the aspects of specific scientific domains or common analytic methodologies. This includes developing and promoting specialist datasets, technology and training to accelerate the adoption of best practice; promoting the FAIR principles to support the goal of reusability; and investing in mechanisms to connect and showcase ELIXIR services with their users.

---

[79] https://seek4science.org/
[80] https://tess.elixir-europe.org/

Our architectural approach will favour semantic interoperability (describable data which can be dynamically interrogated and connected algorithmically) over manual annotation and mapping of connections between data sources. This will be complemented by the development of APIs and common query language adoption (syntactic interoperability), so dependent services and client applications can be rapidly developed. Our technology ecosystems approach uses workflows, shared ontologies, standardised interfaces (APIs) and registries to loosely coupled services and data to fulfil diverse and dynamic requirements. We will extend this approach, pioneered by the ELIXIR Research Software Ecosystem and the Bioschemas initiative, across all of the ELIXIR technology areas.

We will identify key infrastructure services in all areas, from the provision of infrastructure to registries of standards and ontologies, support their development and operation, and formalise a framework to recognise and prioritise new ones. Building, sustaining and improving our infrastructure services will increasingly make use of developed capabilities of partner research infrastructures and computing initiatives such as EOSC, EuroHPC, and GAIA-X. Federation requires interoperability from the policy and ELSI levels to the API, data, semantics levels and we will continue to develop approaches to ensure all are covered.

The success of Life Science Login will be extended to ensure that all services have the ability to use common representations of researcher identity, affiliation, accreditation and resource allocation in national, pan-European and international collaborative projects. This will form a standard service for user authentication, authorisation and, increasingly, accounting. This infrastructure will include a framework of services for delivering training and enabling scientific trainers to develop their own material and contribute to the promotion of ELIXIR services. This will include a centralised, collaborative structure for designing, co-developing and maintaining training material; further development of the TeSS and promotion of the standardised representation of training events and materials (making training FAIR); enhancement of the training metrics database; and developing a global standard for training impact assessment. Training will focus on ELIXIR Community needs and, by coordinating the available expertise, it can be adapted to the requirements of emerging areas.

Ambition 2.3 Develop a federated pan-European technical infrastructure for life scientists to access data, storage and compute services

- A comprehensive set of agreed technical standards, protocols and cost-recovery models allows researchers to perform distributed and/or federated analysis over national data services, clouds and secure processing environments
- Researchers can seamlessly access resources across Europe via a single, secure login providing authentication, authorisation, auditing and accounting
- ELIXIR Deposition Databases and Core Data Resources are at the heart of a distributed network of European data repositories with technical interoperability and standardised impact assessment
- Specialist Communities can find, develop and share collections of services, operating procedures and guidelines to meet their complex needs through familiar technologies
- ELIXIR's key infrastructure services are identified, developed, recommended and sustained
- ELIXIR Nodes contribute to a growing portfolio of multi-Node services, supported, developed and promoted by a distributed network of experts

Node Tier - Equip ELIXIR Nodes for successful long-term operations

Node operations

Ambition 3.1 Nodes have the operational capabilities to contribute to a European infrastructure for data

ELIXIR is a distributed research infrastructure bringing together 25 Nodes, which in turn connect over 245 leading institutes and universities in bioinformatics service provision. ELIXIR Nodes need to be well-coordinated, professionally run organisations that meet the needs of national stakeholders by providing high-quality support, expertise and services (described in Tier 1 and 2) that are aligned with national research priorities. And to be successful over the long term, these national operations require skilled personnel (described in Tier 4) as well as dedicated funding for service provision and coordination.

As ELIXIR Nodes operate within different national contexts; they vary in size, maturity, focus and national funding level. However, all Nodes benefit from operational efficiency, sound governance structures, appropriate funding, relevance to other national priorities (including scientific domains and open science policies), effective industry engagement, processes to capture and demonstrate impact, good coordination and communication between Node partners, and effective outreach to users and collaborators. Some of these areas are described in more detail in ambitions 3.2 (national research priorities and open science), 3.3 (industry and innovation) and 3.4 (impact and sustainability).

The organisational capabilities of Nodes need to grow to meet the demands of increasing usage and the requirements arising from close collaboration with many national partners, as well as European and international collaborators. New Europe-wide developments, such as the creation of common European data spaces, will rely on national implementation, with ELIXIR Nodes increasingly being seen as crucial components of the infrastructure needed to address those objectives. This will require Nodes to have effective communication with a range of transnational collaborators and stakeholders. It is expected that services supporting European data spaces will be co-owned, co-developed and co-operated by several ELIXIR Nodes, increasing the technical and financial resilience of the infrastructure services ELIXIR provide to life science researchers across Europe.

Since the inception of ELIXIR, much support has been given to the development and successful operations of ELIXIR Nodes, through capacity building, knowledge exchange and the sharing of best practice in Node operations. Recently, this has taken place through training within the ELIXIR-CONVERGE project, through formal coordination groups such as meetings of Node Coordinators and various ELIXIR Focus Groups representing impact, industry and innovation and communications. Nodes benefit from efficient communications and outreach channels to national partners and external stakeholders including users, other research infrastructures, funders and

policymakers. The 2024-2028 Programme will support ELIXIR Nodes in establishing and maintaining these critical networks.

Ambition 3.1 Nodes have the operational capabilities to contribute to a European infrastructure for data

- Nodes have in place appropriate governance structures and successful procedures for effective coordination and operations
- ELIXIR enables knowledge sharing between Nodes and establishes an implementation toolkit for adoption of good practice
- Nodes have effective communication channels with national partners and relevant stakeholders

National research and open science policy alignment

Ambition 3.2 Nodes are aligned with national research priorities and open science policies

The COVID-19 pandemic has demonstrated that ELIXIR Nodes are the go-to partners for managing large volumes of data from national research programmes, for example, through rapid development of pipelines connecting data from large cohort studies and other national facilities. This is true in other areas, like the management of sensitive human data, where ELIXIR Nodes are seen by other Member States as crucial to the deployment of the 1+MG initiative.

ELIXIR Nodes operate in the context of national open science and research data policies, and they not only manage digital assets produced by research infrastructures but also with many other national, European and global projects. The national research environments are, and should be, diverse and reflect the strengths and priorities of individual countries. At a local level, Nodes coordinate their activities in response to the national research priorities (for example, ELIXIR Norway's focus on genomes of marine species and ELIXIR Portugal's expertise in the genome of the cork oak tree).

An increasing number of national funders and charitable foundations are developing formal policies relating to open science along with guidelines to support research applicants with the management of digital assets from their research. Traditionally this has focussed on data management, but increasingly, recognition is being given to other assets including software and workflows. In many cases these policies and guidelines reference resources that are run by ELIXIR Nodes.

This welcome development is likely to have two benefits: firstly, ensuring that more digital assets are made FAIR and available for reuse, and secondly, providing the grounds to secure national funding to sustain and develop the referenced resources. However, many national funders do not yet have open science policies or guidance, and when they do, they don't always reference ELIXIR resources. Seeking to address this over time, through supporting Nodes to work with national funders and developing activities that highlight the benefits of ELIXIR resources, will be a priority.

Ambition 3.2 Nodes are aligned with national research priorities and open science policies

- National funders recognise ELIXIR Nodes as key components of the data infrastructure when funding major research initiatives
- National funders increasingly reference ELIXIR resources in open science policies and guidelines

Integrating Nodes with industry and innovation

Ambition 3.3   Nodes are fundamental to national industry ecosystems and engage in EU-wide industry efforts

The range of industry organisations which could benefit from ELIXIR's resources is vast. It includes organisations with a requirement for, or expertise in, bioinformatics, data sharing and infrastructure provision. Such organisations encompasses start-ups and small to medium sized enterprises (SMEs), large enterprises and industry associations such as the Pistoia Alliance[81], EFPIA[82], EuropaBio[83] and CEBR[84]. Industry relationships with ELIXIR are multifaceted and include users of ELIXIR services, collaborators and partners in projects, potential funders of activities in Nodes and providers of services used by academic partners within ELIXIR.

According to the 2022 Deloitte Global Life Sciences Outlook, a significant number of new digital-related jobs will be created by 2025 in the life sciences, with 40 percent of core skills changing, and life sciences competing within and outside the industry for the same digital and data talent[85]. In an increasingly competitive business landscape, efficiency game-changers like AI are fast becoming industry essentials. For this reason, the work of ELIXIR to enable users in academia and industry to access services vital for their research is crucial for boosting open innovation in the European life sciences ecosystem.

Several challenges need to be considered when defining industry engagement objectives for the 2024-28 Programme: the often divergent interests between academia and industry, the perception of limited applicability of academic resources to industry needs, and the lack of traceability of industry users of ELIXIR's digital services.

Building Node industry strategies

The 2019-2023 Programme set a target for each Node coordination group to contain at least one industry contact person. Currently, 80 percent of Nodes either have an industry officer working on a dedicated industry strategy or the responsibility sits under another role, such as Head of Node or Node Coordinator.

Considering the wide range of industry activities, an Innovation and Industry Focus Group was established during the ELIXIR-CONVERGE project for ELIXIR members to share industry engagement experiences. This Focus Group will continue running throughout the 2024-28 Programme, encouraging Nodes to discuss challenges and opportunities, to build a collaborative approach to common problems, and to increase the industry activities in their portfolios.

---

[81] https://www.pistoiaalliance.org/
[82] https://efpia.eu/
[83] https://www.europabio.org/
[84] http://www.cebr.net/
[85] https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Life-Sciences-Health-Care/gx-lshc-dei-global-life-sciences-outlook-report.pdf

Connecting expertise through matchmaking

In the 2024-28 Scientific Programme, ELIXIR will continue to seek high-level input via the Industry Advisory Committee[86] (IAC). Additionally, IAC members will be encouraged to be ambassadors for communicating ELIXIR across their relevant networks. More activities will focus on connecting expertise between academia and industry including giving more visibility to the ELIXIR vacancy webpage[87], which also displays jobs in industry, and linking ELIXIR experts with specific companies to enable collaboration and sharing information on topics of interest.

Expanding networking activities and capacity building for industry

ELIXIR's support for industry related events will expand with the roll out of a new scheme to support Node industry engagement days, which aim to help Nodes build strong links within their national industry ecosystems, and with virtual events where interested ELIXIR partners will have more technical discussions with targeted companies on a theme of interest to a Platform, Community or Focus Group. The Innovation and SME Forums[88] and Bioinformatics Industry Forums[89] will continue with the former focusing on domain-specific themes (such as microbiome, human data, microbial biotechnology), hosted by ELIXIR Nodes in life science hotspots around Europe, and the latter expanding to enable discussions on transversal themes (such as AI and interoperability) between ELIXIR partners and SMEs, multinational companies and industry clusters and associations.

Building collaborations through projects

ELIXIR will continue working closely with industry partners in multi-stakeholder projects, such as those funded through the Innovative Health Initiative[90], and encourage ELIXIR members to build collaborative partnerships with the industry via the Knowledge Exchange Scheme[91]. In this Programme, ELIXIR will support the continuation of relevant industry activities resulting from former projects such as FAIR implementation, human data and COVID-19 research data.

Capturing outcomes

To showcase ELIXIR's efforts around industry engagement, ELIXIR will continue to capture and publish examples highlighting industry use of ELIXIR resources and also establish Programme-wide performance indicators for industry engagement. This will demonstrate the growing impact of ELIXIR resources in the creation of an open innovation ecosystem in the life sciences. During 2024-28, ELIXIR will perform periodic evaluations of European life science ecosystems, examining growth and connections to ELIXIR, and launch new research into the benefits to industry of ELIXIR beyond the current focus on the open data.

---

[86] https://elixir-europe.org/industry/advisory-committee
[87] https://elixir-europe.org/about-us/vacancies
[88] https://elixir-europe.org/industry/forums
[89] https://elixir-europe.org/industry/suppliers-forum
[90] http://www.ihi.europa.eu/
[91] https://elixir-europe.org/industry/knowledge-exchange-scheme

Ambition 3.3 Nodes are fundamental to national industry ecosystems and engage in EU-wide industry efforts

- Industry, particularly small and medium-sized enterprises, benefit from the services, technologies and expertise of ELIXIR Nodes
- ELIXIR Nodes and industry collaborate and innovate to form mutually beneficial ecosystems
- ELIXIR develops a portfolio of case studies communicating the value to industry of public investments in open data infrastructures
- ELIXIR engages in key EU-level industry initiatives, providing solutions based on ELIXIR supported standards

Demonstrating impact and supporting sustainability

Ambition 3.4   Nodes are empowered and supported in their efforts towards long-term sustainability

Demonstrating the scientific, socio-economic and societal impacts of ELIXIR is central to ensuring the long-term sustainability of its services beyond 2028. Work under this portfolio is guided by the 2020 recommendations[92] made to ELIXIR by a high-level expert group, which assessed the progress of the European Strategic Forum on Research Infrastructures (ESFRI) and other world class research infrastructures towards implementation and long-term sustainability.

Developing a community of practice and empowering Nodes

In line with its distributed nature, ELIXIR will continue to take a decentralised and pragmatic approach to evaluating and communicating its impact[93]. It will continue to develop a community of practice, through empowering staff in Nodes to cater for the array of research infrastructure stakeholders[94] (local, regional and national authorities; research infrastructure funders, initiators, and host countries; civil society; and the scientific community) they must engage with. It is essential that Nodes can confidently demonstrate the performance and impact of their work to national and international funders and policymakers.

Drawing from a range of past, current and future EU[95] and ELIXIR[96]-funded projects, ELIXIR is working to equip Nodes in skills and knowledge related to impact evaluation. Capacity building takes the form of formal training delivered by internal and external experts in the field, knowledge-exchange events, and the application of the new skills and knowledge in the context of Nodes' particular needs, for example through Node-relevant case studies. In parallel, a wealth of useful resources (including worked examples, slides, specialised bibliographies, tools and methods, teaching materials) is maintained on the ELIXIR intranet, and will be made public-facing during the new 2024-28 Programme.

A publicly available impact dashboard

Capacity building efforts at Node-level go hand in hand with Hub-driven efforts to demonstrate and showcase the performance and impact of ELIXIR as a whole. In particular, ELIXIR will continue to

---

[92] https://ec.europa.eu/info/publications/supporting-transformative-impact-research-infrastructures-european-research_en

[93] https://doi.org/10.1111/apce.12328

[94] https://doi.org/10.1787/3ffee43b-en

[95] https://ri-paths-tool.eu/en (2018-2020);
https://elixir-europe.org/about-us/how-funded/eu-projects/converge/wp4 (2020-2023);
http://pathos-project.eu (2022-2025)

[96] https://elixir-europe.org/internal-projects/staff-exchange-programme/measure-performance-impact (2019-2020); https://elixir-europe.org/internal-projects/commissioned-services/impact-evaluation (2021-2023)

develop an impact dashboard[97] and maintain underlying indicators, to make it easier for funders and other policy stakeholders to grasp the scale of usage of its flagship services including bioinformatics training activities, as well as the scientific and socio-economic impacts of ELIXIR-related activities more broadly. The information on the dashboard is directly relevant to a range of requests for information on ELIXIR's public value, which are regularly issued by national-level funders as well as influential stakeholders such as ESFRI as part of their monitoring of research infrastructures.

The 2019-2023 Programme saw a marked increase in impact-related activities across ELIXIR, supported by significant capacity building and knowledge exchange. The 2024-28 Programme will see a consolidation of this strategically important portfolio, with more systematic approaches being implemented across activities and projects, a levelling-up of Node capacities in this field, as well as an expansion of efforts to externally communicate ELIXIR's scientific, socio-economic and societal impacts.

ELIXIR resources will be allocated to support the realisation of this strategic ambition, for instance through internally funded projects (such as self-standing Commissioned Service projects and dedicated impact work packages, impact-focussed staff exchange projects) and through positioning and bidding for EU funding related to impact. Further, ELIXIR will make use of external experts in impact evaluation to provide specialist training and also undertake *ad-hoc* impact studies for selected areas of open science where significant ELIXIR investment has been made.

The national funding landscape in ELIXIR Member States varies considerably between countries and changes over time. All ELIXIR Nodes rely on national funding for at least part of their activities, even if this is for the operation of specific services. In many countries effective dedicated national investments have been made to support the coordination activities of the Node, enabling the country to participate fully in ELIXIR's activities and gain maximum return on investment.

However, the complexities posed by the heterogeneous funding landscapes across ELIXIR Member States is recognised as a major issue, as in many countries no dedicated funding for coordination support has been provided to the Node. In addition, ensuring alignment of ELIXIR's Europe-wide activities of the scientific Programme with the operations of the Nodes - when some receive no funding, and others that do receive funding from national roadmap awards have different investment cycles and policy objectives - is the key challenge facing ELIXIR overall.

Providing opportunities for Nodes to exchange good practice in securing national funding and exploring opportunities to exploit alternative funding sources will remain a priority, and will complement the work undertaken to help demonstrate the impact of ELIXIR's activities at all levels.

---

[97] https://elixir-europe.org/about-us/impact

Ambition 3.4 Nodes are empowered and supported in their efforts towards long-term sustainability

- Nodes staff are equipped with the skills and knowledge to confidently demonstrate the performance and impact of their work to national and international funders and policymakers
- ELIXIR has systems in place (for example, to track the number of users, training beneficiaries and publications) to demonstrate public value to a range of stakeholders, including the European Strategic Forum on Research Infrastructures
- ELIXIR works with, and learns from, evaluation and impact specialists through EU- and internally-funded projects
- Information and good practice in securing funding from national and other funding sources is shared across Nodes

People tier - Develop people and capacity to benefit science and society

All of ELIXIR's activities depend on people. It is people with specific expertise and skills who design, implement and maintain the ELIXIR infrastructure. The expertise and skills required are manifold and at times highly specific and specialised. Finding qualified and highly skilled employees is a challenge for many ELIXIR Nodes, along with the need to retain staff and knowledge. To maintain and develop ELIXIR's sustainability and strength, the 2024-28 Programme includes specific investments to build capacity and enable further development of ELIXIR's people network.

ELIXIR Nodes require skilled people to manage the services and operations to deliver impact and ensure sustainability. With the increasing maturity of ELIXIR's federated services comes increasing complexity in the management and sharing of sensitive genomic data, Nodes therefore need solid organisational, operational and managerial capacity to be effective. Nodes must be well-governed with a robust quality review process and strong service delivery in an increasingly interdisciplinary space. These tasks need to be delivered by Node personnel and collaborators who understand the complexities of the landscape where ELIXIR operates and are able to effectively articulate and maximise impact.

Building on ELIXIR's substantial track-record in training, our People development strategic tier in the 2024-28 Programme extends current efforts to develop and sustain a training infrastructure and support its people with knowledge and capacity building activities. The focus is on providing end users and infrastructure staff with the skills needed to operate and make use of a distributed European-wide data infrastructure.

Users of ELIXIR services

Ambition 4.1 All users of ELIXIR services have access to training, resources and expertise

Any service or resource only reaches its full potential if end users are sufficiently knowledgeable and well-trained. Following FAIR principles, good training is not only about the quality of the course itself (the contact and exchange with experienced trainers, service developers or experts) but also the accessibility of the training and the quality of training material.

To address these factors, ELIXIR has developed a strong and successful training community spanning all Nodes, and it develops and maintains ELIXIR's Training Portal (TeSS). Both of which allow services to be scaled to support the data and training needs of Europe's estimated 500,000 life scientists.

ELIXIR acts to strengthen national training programmes, expand bioinformatics training capacity and competence across Europe, and empower researchers to use ELIXIR's services and tools. Reaching many research institutes across ELIXIR Nodes and beyond , the training creates users who can effectively utilise services and increase self-sufficiency. Additionally, the ELIXIR peer-to-peer network described in Ambition 4.2 will enable the exchange of expertise, bringing individuals across Nodes closer together.

During the 2024-28 Programme, continued efforts will be made to enhance the impact and reach of training provided to users of ELIXIR services and resources. A particular focus will be on training in the FAIR creation and stewardship of all types of digital assets, including data, software, methodology, predictive models and images. To maximise the reach of material and expertise, Nodes will be supported in their training provision. This will include efforts to recognise centres of excellent training, and promoting the sharing of good practices between Nodes. Prioritising national capacity building will be supported by the train-the-trainer programme, maintaining the state of the art of training content, and being responsive to local training requirements.

Ambition 4.1   All users of ELIXIR services have access to training, resources and expertise

- ELIXIR gives users access to a Europe-wide training programme for the creation and stewardship of all types of digital asset
- ELIXIR develops and maintains ELIXIR's Training Portal (TeSS)
- ELIXIR strengthens and scales up national user training programmes by supporting trainer development, fostering knowledge sharing and exchange of materials, building capacity and driving adoption of good practice between Nodes (based on the FAIR training principles)

Management and operational staff

Ambition 4.2 Nodes are well equipped to efficiently manage and run a distributed infrastructure

Successful research infrastructures rely on managerial and operational excellence as well as technical and scientific knowledge. Ambition 4.1 covers management and operational aspects, whereas technical and scientific aspects are described in Ambition 4.2.

People working in research infrastructures have varied roles and career paths, and a shift from technical or scientific roles to managerial or administrative positions is common. ELIXIR aims to build capacity, specifically for middle management level, and support Node personnel to run efficiently organised Nodes as part of a distributed research infrastructure.

The challenges facing management and operational staff in ELIXIR Nodes will be amplified in the next decade by the combined pressures of rapidly evolving technologies, novel scientific advances and increasing requirements for distributed operations. A case in point is the European Genomic Data Infrastructure project (GDI), initiating in 2022, coordinated by ELIXIR and involving 20 member states. The size and complexity of the project requires Node operational expertise in areas such as business and organisational management, legal, ethical and social aspects of work. Additionally, similar efforts in other scientific projects are expected to emerge.

Given ELIXIR's distributed set up, the closest colleagues in terms of expertise may reside in a different country. This creates additional challenges to maintaining and developing a sense of belonging to the ELIXIR community, while ensuring connection with colleagues to share best practices. The 2024-28 Programme will focus on creating peer-to-peer networks (i.e. communities of practice) to allow members, particularly with an operational role, to come together as a community and share good practices.

In the 2024-28 Programme we will use external training providers for general training, either in national settings or carried out in larger, European-wide initiatives. In parallel, specific training required for optimal functioning of the ELIXIR research infrastructure will be developed and offered internally. Both internal and external training are crucial to sustain and grow the skills and capacity of people in ELIXIR at all levels. To achieve these outcomes, a number of existing instruments such the Executive Masters in Management of Research Infrastructures (EMMRI) course[98] will be used, and new tools such as those developed in the RITrainPlus project[99] and the ELIXIR Training Programme in Management (ELITMa), emerging from the ELIXIR-CONVERGE project, will be introduced. In addition, tools such as the Staff Exchange Programme[100] and mentoring (see Ambition 4.4) will be used to further develop management and operational staff.

---

[98] https://emmri.unimib.it/
[99] https://ritrainplus.eu/
[100] https://elixir-europe.org/internal-projects/staff-exchange-programme

This part of the 2024-28 Programme will increase the capacity of people to fulfil their roles, assume more responsibilities and advance to leadership positions, as well as to support others.

Ambition 4.2 **Nodes are well equipped to efficiently manage and run a distributed infrastructure**

- Personnel are supported to run efficiently organised Nodes, with good understanding of distributed research infrastructures in general, and ELIXIR operations in particular
- A peer-to-peer network (a community of practice) enables ELIXIR people, particularly those with an operational role, to receive support and guidance and actively share knowledge with others
- Management and operational staff are well prepared to meet the organisational, business and ethical, legal and social implications (ELSI) of the Genomic Data Infrastructure project and emerging life science and research infrastructure challenges

Technical and scientific staff

Ambition 4.3 ELIXIR technical and scientific staff have the skills needed for operational excellence and impact

ELIXIR members, distributed across the infrastructure, are experts in all aspects required to sustain and advance a European research infrastructure. Much of this knowledge is highly specialised and rapidly changing, creating challenges for both knowledge dissemination to existing staff and the onboarding of newcomers. In addition, there is a growing requirement for experience of federated operations, with associated business and ELSI skills, for example, in the coordination of the European Genomic Data Infrastructure project.

Technical knowledge is currently disseminated within ELIXIR through existing member-based operational structures (such as Community, Platform and Node-based activities). For example, Galaxy training efforts (such as the GTN Smörgåsbord[101]), RDMkit activities[102] and learning paths offering targeted sets of training material published on TeSS. Further examples are activities of the ELIXIR Node coordinator group, data stewards and data managers (for example, specialised training in software development, IT operations, system administrator and data management plans). The goal for 2024-28 is to maintain and increase these activities, enabling training by the experts in a specific field to the individuals with the greatest training need.

To support organisational and operational excellence, ELIXIR makes use of existing offers (for example, the RITrainPlus project[103]), national-level training programmes and internal ELIXIR expertise. Technical experts need to be supported in learning new technologies to meet the growing need for expertise on Life Science Login, TeSS and GDI, and more generally in open science, cloud resources, federated operations and research data management. Depending on requirements, specific modules will be developed for internal ELIXIR training and knowledge sharing (similar to the approach used in ELITMa in section 4.2).

ELIXIR's 2024-28 Programme highlights the importance of life-long learning, allowing individuals and the research infrastructure as a whole to stay abreast of technical developments, strive for excellence, and see training as an inherent part of working life.

---

[101]https://gallantries.github.io/posts/2021/12/14/smorgasbord2-tapas/?utm_source=org-website&utm_medium=web&utm_campaign=smorgasbord
[102] https://rdmkit.elixir-europe.org
[103] https://ritrainplus.eu/

Ambition 4.3 ELIXIR technical and scientific staff have the skills needed for operational excellence and impact

- Nodes are supported to become training providers and national centres of excellence in the FAIR stewardship of all types of digital assets
- A training programme enables all Nodes to maximally benefit from technical roadmaps and associated services
- Technical and scientific staff are well prepared for the requirements of emerging life science and research Infrastructure challenges (for example, the Genomic Data Infrastructure)

A diverse people network

Ambition 4.4   ELIXIR is a diverse network of people who are appropriately credited for their work

Over the past ten years, ELIXIR has built a community of people who work together to achieve common goals. This network of over 850 people is diverse in many different aspects, not just the cultural mix that comes with 23 member countries, but also in terms of background, education, expertise and skills. The benefits of the diversity are clear: together ELIXIR can develop common objectives and achieve shared outcomes across the ELIXIR community.

ELIXIR has built this people network to develop and implement a shared vision through many ways. These include events (for example, the All Hands Meeting, the Biohackathon, Platform and Community meetings), workshops, joint projects, Commissioned Services and operational structures (Platforms, Communities and Focus Groups). The people in ELIXIR usually belong to one or several operational structures, which strengthens the collaboration and connects the people within the infrastructure.

Any ambitions for the development of people must recognise that careers are life-long. In the 2024-28 Programme we will provide a set of mechanisms to promote long-term development, peer-to-peer learning and establish communities of practices. These will include promotion of the staff exchanges efforts to encourage participation by more Nodes. Following consultation, we will put in place a knowledge sharing and capacity building programme to meet the needs for ongoing learning and topical exchange. The Programme will aim to harness and share existing knowledge within and across Nodes in a way that is of benefit to all participants.

Furthermore, traditional publication-based approaches do not adequately reflect the contributions of infrastructure members to managing and sustaining the ELIXIR infrastructure. The use of a unique identifiers (i.e., such as ORCID[104]), and novel approaches towards the recognition of publishing data, software and others are examples of new practices which can be encouraged. In the 2024-28 Programme, we aim to support and grow people by ensuring fair recognition and credit, by providing mechanisms to promote long-term development following the life-long learning principle, and by giving opportunities to share and exchange expertise with current and past ELIXIR members.

ELIXIR has become an established, large, distributed and virtual organisation, which is based on key values of open science, inclusivity and trust. ELIXIR's people see these values as important and are proud of their efforts to realise benefits to research, industry and ultimately the public good. In the 2024-28 Programme, ELIXIR will focus on nurturing this sense of belonging. We will focus on establishing mechanisms to ensure that this international, interdisciplinary, distributed network welcomes newcomers, helps them to understand how ELIXIR works and allows them to connect across Europe and the globe.

---

[104] https://orcid.org/

Recognising the benefits of diversity, ELIXIR will increase efforts in equity, diversity and inclusion (EDI) and seek to understand how to best promote the participation of under-represented groups. It is critical for ELIXIR to support those groups in leadership positions, and we will prioritise efforts to establish a long-term plan for increasing the diversity of individuals both in the leadership roles and in the infrastructure as a whole. The EDI working group and emerging initiatives, such as ELEAD[105], will be expanded and will define a process to monitor the current situation and track improvements over time. Additionally, ELIXIR will use its position to advocate for diversity in experimental design and awareness of gender in data analysis settings.

Ambition 4.4 ELIXIR is a diverse network of people who are appropriately credited for their work

- ELIXIR provides mechanisms to recognise community and organisational contributions to research infrastructure development, operations and impact
- Mechanisms for recognition of contributions and the long-term development of people are in place
- ELIXIR operates a set of people development tools (for example, knowledge sharing and capacity building schemes) which are well understood and utilised by all Nodes
- ELIXIR is a well-connected network of people with a strong sense of belonging
- ELIXIR has a long-term plan for increasing the diversity of individuals in the infrastructure, and equity, diversity and inclusion (EDI) support and training ensures a diverse and inclusive selection process for all ELIXIR leadership roles

---

[105] https://elixir-europe.org/internal-projects/commissioned-services/elead

Partnerships

Introduction to partnerships

The world of bioinformatics is highly interconnected, with dependencies established between many different services and organisations within countries, across Europe and globally. Whilst ELIXIR connects over 245 universities and research institutes, there are many other organisations and initiatives with which ELIXIR must collaborate to further the aim of helping life scientists access and analyse life science data. These include global standard-setting organisations, national bioinformatics initiatives in countries outside ELIXIR, research infrastructures with data-related needs, and industry and technology-focussed initiatives in domains within the scope of ELIXIR. Working in partnership with these organisations helps ELIXIR better understand user needs, ensure there is no unnecessary duplication of effort and improve service interoperability across the globe.

The most critical initiatives will require formal partnerships or collaboration agreements to be established or maintained during the period 2024-2028. Other initiatives will have strategic relevance to ELIXIR, but implemented through lighter, ad-hoc collaborations. In other cases, collaborations will be fostered and developed through EU-funded projects or enabled through national research infrastructure roadmap funding.

Emerging national bioinformatics infrastructures within Europe

Whilst ELIXIR has a large number of members, there are many other countries that are part of ESFRI and the European Research Area that are not members. In some of these countries, efforts are taking place within the scientific community to coordinate bioinformatics service provision nationally and collaborate with ELIXIR. In other countries ministries are actively working to join ELIXIR.

These countries include but are not limited to Austria and Romania (who joined ELIXIR as Observers in 2023), Croatia, Latvia, Poland, Romania and Slovakia. ELIXIR will seek to support these emerging bioinformatics communities as they develop and provide for an appropriate form of collaboration with existing Members.

European technology initiatives

From its foundation, ELIXIR has been a leader in its vision of a network of open data resources and the services required to connect and make use of them. Key ELIXIR experts were among the authors of the original FAIR principles[106], and the life sciences with their large and complex data resources collected through collaborative worldwide activity have long been at the forefront of the technical requirements and solutions for data storage, structuring and reusability.

ELIXIR has set out to provide the necessary infrastructure to support this activity at a European scale, organising across three dimensions - national, technological and disciplinary.

European Open Science Cloud (EOSC)

The ambition of the European Open Science Cloud (EOSC)[107] is to develop a web of FAIR data and services for science in Europe. EOSC will be a multi-disciplinary environment where researchers can publish, find and re-use data, tools and services, enabling them to better conduct their work.

The EOSC Association was formed in 2020 and is the legal entity established to govern EOSC and has since grown to over 200 Members and Observers. The Association membership is jointly responsible for delivering the objectives agreed in the Memorandum of Understanding signed by the European Union and EOSC Association to form the official EOSC Partnership[108]. The EOSC ecosystem is being co-created in a series of funded projects and initiatives from Member States and Associated Countries. The EOSC Association plays an important role in helping to coordinate and steer these investments via its task forces and other governance structures.

While it is expected that the majority of EOSC-wide services are provided by the e-infrastructures (e.g. EGI[109], EUDAT[110], OpenAIRE[111], and GÉANT[112]), generic services and resources offered by the science cluster communities will also be offered. The architectural framework for EOSC is directed by a set of EOSC Association task forces[113] for different technology focus areas - such as Technical interoperability of data and services and Long-term data preservation.

ELIXIR has long identified EOSC as a significant strategic partner, with a position paper[114] in 2019 and the 2019-2023 Scientific Programme highlighting the development of the relationship as a strategic objective.

ELIXIR and EOSC have worked together and participated in EOSC projects on many aspects of the underpinning technologies and end users application enabled by the shared vision. This has enabled

---

[106] https://doi.org/10.1038/sdata.2016.18
[107] https://eosc.eu/about-eosc/
[108] https://eosc.eu/partnership/
[109] https://www.egi.eu/egi-infrastructure/
[110] https://eudat.eu/
[111] https://www.openaire.eu/
[112] https://clouds.geant.org/services/infrastructure-services/
[113] https://eosc.eu/advisory-groups/
[114] http://bit.ly/2QTja9L

the sharing and co-development of technologies and the demonstration of utility and impact for the life sciences.

In 2022 an ELIXIR EOSC Focus Group comprising over 100 individuals agreed on a new ELIXIR 2022 Strategy[115] which identified the following objectives to take us into the new Programme:

1. **Define and develop EOSC for the life sciences.** ELIXIR, in partnership with biomedical science research infrastructures and other stakeholders, will have defined and developed EOSC for the life sciences, allowing users to access fully operational services via ELIXIR portals and to be supported by ELIXIR Node experts.

2. **Recommended Resources become core parts of the EOSC.** ELIXIR Recommended Resources, such as Core Data Resources, Deposition Databases and Recommended Interoperability Resources, become core parts of the EOSC for life sciences

3. **Provide sustainable, cloud-based workflows.** ELIXIR will provide sustainable, cloud-based workflows that represent the gold standard for workflow-based biological data analysis, demonstrating the central role of ELIXIR in the European life science data landscape.

4. **Build the capabilities of the user communities and infrastructure operators.** ELIXIR will build the capabilities of the user communities and infrastructure operators required to make the best use of the EOSC across all our Nodes.

EuroHPC

EuroHPC is a Joint Undertaking (JU) for advanced high performance computing which allows participating countries to pool their resources together and boost their scientific excellence. The EuroHPC JU has organised the first round of pre-exascale capacity in Europe and ELIXIR Nodes are tightly connected to all three consortia that are deploying pre-exascale computing capacity for research use. These new computing capacities will be deployed in Finland, Italy and Spain and ELIXIR Nodes in these countries are key partners in the coordination of these efforts.

As a research infrastructure, ELIXIR plays a key role in bringing the researcher requirements to the EuroHPC consortium. EuroHPC also aims to widen the use of that supercomputing infrastructure to public and private users, thus supporting research and public-private partnerships in Europe. These partnerships will help drive ELIXIR's industry collaborations forward.

Gaia-X

Gaia-X is an innovative technology concept which is gaining considerable support across Europe in developing architecture standards and policies for European data spaces. While primarily aimed towards business use cases, representatives from academia, research and science ensure that the Gaia-X principles are universally applicable and allow full control by the data owners - both organisations and citizens - over what data is processed and wheres.

---

[115] https://zenodo.org/record/7120997/

Gaia-X is based on decentralised architecture that facilitates connecting data with appropriate services so data can be exchanged and processed in trusted environments. Gaia-X itself is organised as a distributed organisation, with national Gaia-X hubs capturing national initiatives, and allowing interested parties to engage with Gaia-X locally. The Gaia-X AISBL[116] is the organisation that provides the governance structure to align the various specific requirements to iterate and refine the architecture, policy and rules, and standards emerging from Gaia-X related activities. Core technical components of Gaia-X are the Gaia-X Federation Services[117] which facilitate building trusted interconnections between services. ELIXIR's requirements and input will be shared with the Gaia-X AISBL via ELIXIR members with direct involvement through their affiliation with Gaia-X member organisations.

Partnerships with ESFRIs

Since its inception in 2006, the European Strategic Forum on Research Infrastructures (ESFRI) process has led to the establishment of over 50 ESFRI projects and landmarks[118]. Many of these ESFRI operate in the life sciences, and therefore have touchpoints or connections with ELIXIR. Some operate in the e-infrastructure domain, and others operate in the environment and biodiversity domains, which will be an increasingly active area for ELIXIR.

Several ESFRIs provide access to technologies that generate data for which ELIXIR has solutions, and many also provide complimentary services that, along with ELIXIR's activities, help to improve the life science data ecosystem for users.

In recent years, ELIXIR has forged close collaborations with key ESFRIs including, for example, BBMRI-ERIC, EuroBioImaging and EMPHASIS, which has led to joint meetings, workshops and formal collaboration strategies. EU-funding, in particular through Horizon Europe 'cluster' projects such as EOSC Life and European Open Science Cloud-related projects like BY-COVID has enabled ELIXIR to develop links with all of the mature ESFRIs in the life sciences.

The consolidation of research infrastructures and their services is considered favourable to stakeholders, providing clarity to users and better value to funders by removing duplication of effort. As the landscape of ESFRI projects evolves continuously with each new iteration of the ESFRI roadmap, ELIXIR will need to reach out to and establish synergistic collaborations with new ESFRIs and consider how to support scientific communities when projects have moved off the roadmap.

In the Programme 2024-2028, ELIXIR will seek to formalise collaboration strategies with a small number of critical ESFRIs where there is deemed to be strategic benefit. Examples of the areas that are likely to be of strategic relevance to ELIXIR are described in the scientific and technical tiers. These include ESFRIs covering health, image technologies, biodiversity and industrial biotechnology.

---

[116] https://gaia-x.eu/
[117] https://www.gxfs.eu/
[118] https://www.esfri.eu/esfri-roadmap-2021/

The Global Alliance for Genomics and Health

Securely accessing personal genomic data at a population scale, and across national borders, is an enormous challenge that will require significant investments in national and international infrastructure. The scope and complexity of this infrastructure is beyond the capacity — and jurisdiction — of any single organisation; international collaboration is needed to provide recognised, secure, standardised, documented and interoperable services.

Building on the Framework for Responsible Sharing of Genomic and Health-Related Data[119], the Global Alliance for Genomics and Health (GA4GH) is an international, nonprofit alliance established in 2013 to accelerate the potential of research and medicine to advance human health. Bringing together over 600 leading organisations working in healthcare, research, patient advocacy, life science and information technology, the GA4GH community works together to create frameworks and standards to enable the responsible, voluntary and secure sharing of genomic and health-related data.

The ELIXIR::GA4GH strategic partnership, was founded on a collaboration agreement from 2017, and formalised as a partnership in 2019. It is imperative to continue this relationship in the 2024-28 Programme to build upon the foundation that has been laid for an even stronger collaboration. This will enable coordination of the generation and implementation of the GA4GH suite of interoperable standards and policy frameworks to overcome the technical and regulatory hurdles to genomic data sharing with a focus on Europe.

The number of interactions between ELIXIR and GA4GH has increased as a result of the organisations' expanding memberships. ELIXIR experts continue to contribute to GA4GH projects and workstreams to ensure that genomic standards are developed. The focus of GA4GH is to agree genomic data and analytics standards to ensure interoperability at an international level. ELIXIR focusses on European settings by developing, adopting and implementing standards for European use.

The ongoing partnership ensures that European efforts are aligned with international efforts, preventing the creation of silos and ensuring genomic data sharing is enabled to benefit human health, equitably. GA4GH standards will be recommended as part of the 1+MG Trust Framework and via proof of concept workflows established for specific use cases such as rare disease. Implementation and demonstration of these standards, across a range of use cases, will be a focus of the 24-28 Programme using the infrastructure deployment (delivered by the European Genomic Data Infrastructure project, GDI). Another focus is to ensure the future strategic roadmaps of ELIXIR and GA4GH align and work towards identifying and developing standards where there are gaps in workflows, for example, in federated analysis.

---

[119] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4685158/

Strengthening collaborations beyond Europe

ELIXIR's international engagement is guided by an international strategy[120], which describes in a high-level manner the current ELIXIR activities of international relevance, and planned implementation actions to reinforce ELIXIR's global significance and impact. We will continue to develop this portfolio in the next Programme, building on the successes of earlier work.

Strategic and technical international engagement. Strategic international engagement will remain a priority, for instance the development of collaboration strategies, participation at international events (for example, the International Conference on Research Infrastructures, plenaries of the Research Data Alliance), and support to sibling organisations such as the Global Biodata Coalition, which aim at developing sustainable funding solutions for core databases. Technical engagement will be expanded, for example, projects linked to the GA4GH, participation in technical working groups and initiatives (for example, Global Organisation for Bioinformatics Learning, Education and Training [121] (GOBLET), Common Workflow Language, Galaxy), as well as inviting international partners on ELIXIR-coordinated grants. ELIXIR's commitment to reaching out beyond Europe is evidenced by the extent of the (often in-kind) international engagement taking place across ELIXIR with technical, scientific and policy partners.

Building on strong links with the Australian BioCommons. Bioinformatics is a global science and ELIXIR strongly benefits from interacting and collaborating with research infrastructures beyond Europe that also have open science as a founding principle. ELIXIR and the Australian BioCommons renewed their collaboration strategy in 2023, owing to the considerable mutual benefits that it has brought to the two research infrastructures in a range of technical domains of implementation, such as tools, compute and training. Looking forward, the collaboration strategy will expand to better capitalise on our synergies (ELIXIR Communities and their Australian counterparts).

Deepening the relationship with the US National Institutes of Health. For a number of years, ELIXIR has cultivated a relationship with the US Office of Data Science Strategy[122] (ODSS) at the National Institutes of Health (NIH). This relationship will continue to be developed during the new Programme, mainly through informal means such as joint events at international conferences, representation at various meetings and technical workshops, and the facilitation and sponsorship of study visits or secondments. During the new Programme, ELIXIR will cultivate other such collaborations with other national-level bioinformatics communities beyond Europe, where there is mutual benefit and alignment in goals.

Growing ELIXIR's impact in policy. ELIXIR is a highly visible research infrastructure in international policy circles, owing to active engagement by the ELIXIR Hub and Nodes with a range of bodies. These include the G7[123] (technical working groups, Group of Senior Officials on research infrastructures), the Organisation for Economic Cooperation and Development[124] (various work strands of the Global Science Forum), and several members of the United Nations family such as the

---

[120] https://f1000research.com/documents/8-1583/
[121] https://www.mygoblet.org/
[122] https://datascience.nih.gov/about/odss/
[123] https://en.wikipedia.org/wiki/G7/
[124] https://www.oecd.org/

World Health Organisation[125] (for example, infectious diseases), the Convention on Biological Diversity[126] (for example, Digital Sequence Information), and UNESCO[127] (open science). ELIXIR will continue to be recognised as an infrastructure of global relevance and a partner of choice for intergovernmental organisations and fora.

---

[125] https://www.who.int/
[126] https://www.cbd.int/
[127] https://www.unesco.org/en

Appendix: Background to the Programme

A1: About ELIXIR

We enable scientists to access and analyse life science data

A collaborative infrastructure to support collaborative research. ELIXIR is a distributed infrastructure bringing together experts from across Europe to enable life science researchers throughout the world to access and analyse life science data. The 23 ELIXIR Member States (along with EMBL-EBI) each form a national Node composed of one or more centres of excellence in bioinformatics. Each Node coordinates services, standards and resources, and collaborates with experts in other Nodes to create a sustainable Europe-wide infrastructure for biological data.

Life science is a global community uniting nations with the shared goal of understanding life to benefit humanity and the planet. As life science evolves into a data science, the mechanisms previously relied on to ensure knowledge sharing and efficient progress, such as global conferences, peer reviewed publications and international societies, are no longer sufficient. Data infrastructures need to evolve with the science and ELIXIR is at the forefront of this revolution.

Life science has shared challenges and shared solutions. Common challenges and opportunities are familiar across modern life science: managing the increase in diverse and complex data generated by high-throughput technologies, combining different types of data and making them visible and accessible to the scientists who need them, finding right tools for analysis (and the compute and storage to use them), and accessing training and sharing expertise. Critically, these common technical challenges are currently limiting factors in the speed of addressing many of the wider challenges facing society, such as pathogens, biodiversity and food security.

Given the scale and international scope of the issues facing society, there is an urgent need for coordinated and trusted networks of experts who work across borders and have a technical level of understanding. ELIXIR is a mature infrastructure with a long track record of addressing the challenges of data-centric life science and has the resources, both in terms of people and proven technological solutions, to be well positioned to rise to challenges of the future.

ELIXIR excels at adapting and repurposing. ELIXIR manages foundational resources, including databases, software tools, training materials, standards and cloud storage, that are readily adaptable to different life science disciplines. This ability to quickly pivot and meet emerging scientific needs is illustrated by ELIXIR's contribution to the COVID-19 pandemic. Prior to 2020, ELIXIR did not have an infectious disease programme, yet was able to rapidly realign and refocus existing resources and expertise on serving the global COVID-19 research community.

The ELIXIR structure is an adaptable one, whereby technical solutions can be matched to life science challenges, and learnings from one area can be readily applied to another. By working together, ELIXIR offers sustainable, forward-thinking and impactful solutions which tackle the foundational challenges inherent in leading-edge life sciences.

National Nodes are the beating heart of ELIXIR. At the heart of ELIXIR's expert network lie the 25 Nodes who provide services and resources, a strong set of expertise and skills, and national level coordination. This structure of national Nodes connected to a central secretariat or Hub and collaborating via transnational working groups, forms a sustainable grass roots network which understands national needs, builds links with local users and data providers, and looks outwards to enhance synergies and collaborations with other nations.

Increasingly, the data federation model is being adopted as a powerful solution to enable data access whilst respecting national data sharing regulations. ELIXIR is proud to be in the vanguard of data federation with its support and vision contributing to the creation of the Federated EGA to enable sensitive data access across Europe.

ELIXIR's people are its backbone. Groups of experts form the foundational backbone of ELIXIR. There are groups organised by scientific discipline, by technical focus and by special interest. There are working groups for each project work package and groups dedicated to training and knowledge exchange. Each of these groups overlap and interconnect in ways that would be impossible to unravel, but which provide an immense depth of experience to guide ELIXIR's operations.

ELIXIR also has numerous international and industry collaborations - yet more connections between groups of people bringing a range of expertise and fresh perspectives from outside the initiative. The expertise and skills within ELIXIR are not limited to technical competencies but span a range of project management, operations and scientific administrative capacity, in both the Hub and the Nodes.

A successful and sustainable funding model. The activities described in ELIXIR's Scientific Programme are funded by ELIXIR Member States as defined in the ELIXIR Financial Plan, by joint European research infrastructure projects (funded by the European Commission and other international funders), and by the national contributions to Nodes for service operation and coordination (funded by national funders). The Programme not only drives the implementation of ELIXIR's scientific strategy but also guides future EU grant applications by ELIXIR by setting priority scientific themes for future engagement.

A2: ELIXIR's governance structure

The ELIXIR governance structure is defined in the ELIXIR Consortium Agreement[128] and is represented below:

- **ELIXIR Board**: the highest decision-making body in ELIXIR.

- **Scientific Advisory Board**: advises the ELIXIR Board and the Director on ELIXIR's scientific strategy and reviews ELIXIR Node applications and monitors their performance.

- **Industry Advisory Committee**: gives advice and guidance on industry needs to the ELIXIR Board, Director and the Heads of Nodes Committee.

- **Director**: responsible for implementing the ELIXIR Board's decisions, for governing the ELIXIR Hub and for implementing ELIXIR's scientific programme and financial plan.

- **Heads of Nodes Committee:** consists of the heads of the national infrastructures, which have obtained the status as a Node. The Committee advises the ELIXIR Board and the Director to develop ELIXIR's scientific and technical strategy, including its Programme.
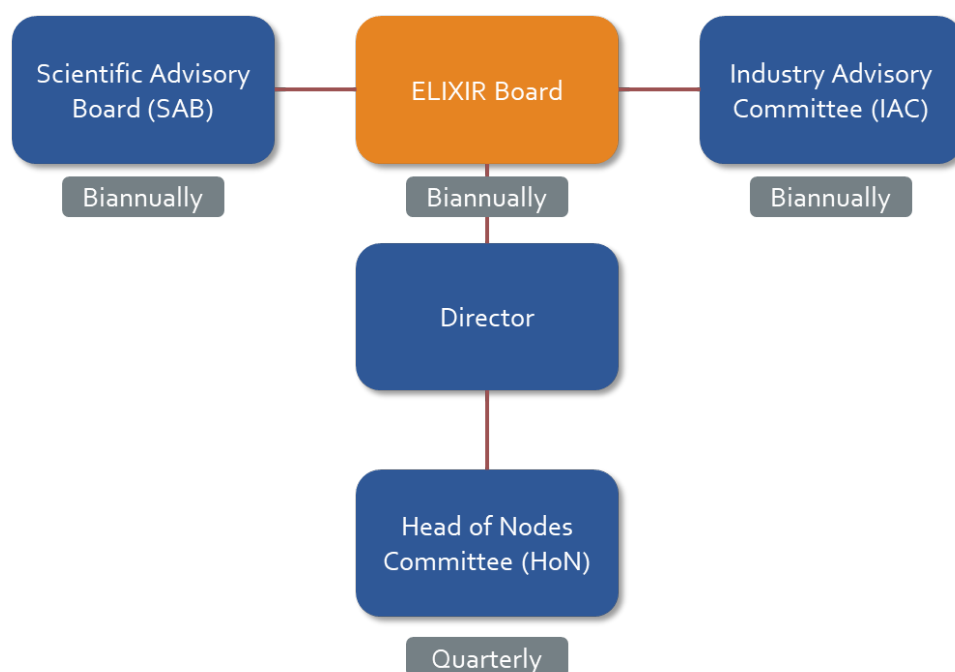
Figure A: ELIXIR's governance structure, including meeting frequency

---

[128] https://drive.google.com/file/d/0B7btK9HAXhx1dEVIMmRsaEpDSzA/view?usp=sharing

A3: EU Funding 2024-28

Since ELIXIR's inception, EU funding has provided a substantial source of income, supporting activities in both the ELIXIR Hub and Nodes, complimenting the budget provided by Member States via the ELIXIR Financial Plan[129]. In the EU financial period 2021-2027, ELIXIR will seek to participate in relevant research and innovation programmes, in particular Horizon Europe as well as the Digital Europe and EU4Health Programmes. Participation in these programmes will enable ELIXIR partners to develop new services, deploy infrastructure and connect with large-scale users. Ensuring that ELIXIR is aware of and can respond to new funding calls in these programmes will remain important, given they will provide both funding and opportunities to collaborate closely with other initiatives.

Within Horizon Europe, the research infrastructure programme will remain a key opportunity, allowing ELIXIR to collaborate with other ESFRIs and e-infrastructures on flagship data-related projects that address specific global challenges and projects allowing ELIXIR to develop generic services including data, software and workflows, interoperability resources, training and compute. Increasingly, large collaborative research projects in the 'Health' and the 'Food, Bioeconomy, Natural Resources, Agriculture and Environment' parts of Horizon Europe will present opportunities for ELIXIR to partner and support their data management and infrastructure needs.

New Horizon Europe Partnerships[130] have been established in a number of areas within the scope of ELIXIR. Partnerships bring together funding from the European Commission, as well as other stakeholders including Member States and sometimes industry, allowing consortia to work collaboratively. ELIXIR will monitor and when relevant seek to engage in and with the following partnerships:

- Innovative Health Initiative
- European Partnership for Personalised Medicine
- European Partnership on Rare Diseases
- European Partnership on Pandemic Preparedness
- European Partnership for OneHealth/Antimicrobial resistance
- ERA for Health Research
- Partnership for Chemical risk Assessment (PARC)
- European Open Science Cloud
- Partnership for EuroHPC
- AI, Data, Robotics
- European Partnership for rescuing biodiversity to safeguard life on Earth
- Accelerating farming systems transition: agroecology living labs and research infrastructures
- European Partnership for Animal Health
- Agriculture of Data
- Safe and Sustainable Food Systems

---

[129] https://f1000research.com/documents/8-1642

[130] https://research-and-innovation.ec.europa.eu/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-europe/european-partnerships-horizon-europe_en

In addition to Partnerships, the following Missions[131] will also present funding and collaboration opportunities for ELIXIR:

- Cancer
- Healthy Soil
- Healthy Oceans

The emergence of data spaces in domains of relevance to ELIXIR, such as the European Health Data Space, will also present opportunities for ELIXIR through funding via the Digital Europe and EU4Health Programmes, building on foundations established during ELIXIR's current Programme period. ELIXIR will seek to engage in other data spaces as they are created, ensuring that the research data expertise and services provided by partners can be connected to and interoperable with the data types and standards being developed.

EU funding opportunities will provide ELIXIR with the opportunity to implement activities in many new areas of work including in One Health, marine sciences, food and biodiversity, where until now, ELIXIR has not been as active as in the human data related domains. ELIXIR will also explore opportunities provided through other international funding sources or charitable foundations when in scope of ELIXIR's Programme and where there is sufficient added value.

---

[131]https://research-and-innovation.ec.europa.eu/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-europe/eu-missions-horizon-europe_en

A4: Impact of the 2019-2023 Programme

ELIXIR's 2019-2023 Scientific Programme has had an impact in a number of areas. These are depicted below in relation to ELIXIR's key stakeholders including life scientists and bioinformaticians, other research infrastructures, policymakers and citizens. For further information, *Martin et al.*[132] summarises ELIXIR's journey in building impact monitoring capacity in the Hub and Nodes.

For life scientists and bioinformaticians

Sustainable technical solutions	Human data management
Creating services, registries and guidelines, for example: ● Life Science Login ● RDMkit and Data	Building pan-European infrastructures for sensitive data: ● Federated European Genome-phenome Archive (EGA) ● Beacon network ● Human Data Communities
COVID-19 response	**Node capacity building**
Leveraging established infrastructures and networks to rapidly respond to COVID-19: ● EU project funding ● COVID-19 Data Portal and data brokering ● Rapidly development of software, workflows and compute resources	● Supporting 24 Nodes as national centres of excellence in life science data management, enabling local outreach to users and industry ● Connecting and strengthening Node network ● Supporting technical and training coordinators
Scientific and technical community building	**EU grant coordination and participation**
● Building and strengthening ELIXIR's Platforms for technical activities ● Growing ELIXIR Communities in life science domains ● Initiating Focus Groups for new and upcoming areas	● Coordinated 8 EU projects ● Participated in 21 EU projects
Strategic collaborations	**Industry outreach**
● European Open Science Cloud (EOSC)	● 10 Innovation and SME forums in 5 different countries

---

[132] https://onlinelibrary.wiley.com/doi/10.1111/apce.12328

• GA4GH (Global Alliance for Genomics and Health) Strategic Partnership • Research Data Alliance (RDA) • Global Organisation for Bioinformatics Learning, Education and Training (GOBLET) • Australian BioCommons	• 4 knowledge exchange schemes • 3 Bioinformatics industry forums • 20 industry newsletters to >1,100 subscribers
Training and knowledge exchange • >880 training courses, including train-the-trainer • 30 staff exchanges • >70 travel grants administered • 34 ELIXIR participants in the Executive Masters in Management of Research Infrastructures (EMMRI), representing 11 Nodes	**Commissioned Services** • Funding for 82 short-term technical projects, to advance the work of the Platforms and Communities, and to help connect resources across the Nodes • >€16,500 in funding assigned
Scientific publications • >600 ELIXIR-supported publications and >21,000 citations on the development and operation of bioinformatics resources (databases, tools, cloud, standards and training)	**High impact events** • 4 BioHackathon Europe events • 5 All Hands meetings

For research infrastructures

Research infrastructure visibility • Working with funders to increase visibility and appreciation of sustainable RIs • For example, ELIXIR heavily referenced in 2023 OECD *Very Large Research Infrastructures* report[133]	**Research infrastructure collaboration** • Promoting cross-infrastructure dialogue and bringing RIs together via EU grants • Integrating key life science services and enabling cross domain interoperability

For policymakers

Open science awareness • Ensuring policymakers are aware of open science and promote FAIR (Findable, Accessible, Interoperable & Reusable) • >100 recorded engagement actions with policy stakeholders	**Cross-border sharing of sensitive data** • Developing mechanisms to enable the access and analysis of sensitive data within the jurisdictional data protection regulations of each state • Federated EGA initiated with 5 inaugural nodes

---

[133] https://www.oecd.org/publications/very-large-research-infrastructures-2b93187f-en.htm

● >120 mentions of related to ELIXIR and its work in policy documents	

For citizens

Value for money	Publicly available science
● Improving resource efficiency and value for money for publicly funded science through innovation	● Enabling outputs of publicly funded science to be reused by others
COVID-19 response	**Awareness**
● Supporting a rapid response to COVID-19 through existing cross-border collaborative networks and infrastructures	● Raising the awareness of bioinformatics and open science, including their socio-economic benefits

ELIXIR HUB

info@elixir-europe.org

South Building

Wellcome Genome

Campus

Hinxton

CB10 1SD, UK

ELIXIR GLOSSARY