

ELIXIR position paper on FAIR Data Management in the life sciences

Findable, Accessible, and Interoperable life-science data are reused

Biological sciences have a long tradition of open research data where vast data sets are made available for community reuse -- sometimes even prior to publication. ELIXIR, the European research infrastructure for life-science data, is committed to coordinating, integrating and sustaining deposition databases and supports European life scientists in making their data Findable, Accessible, Interoperable and Reusable (FAIR).¹ The coordinated action of national Nodes ensures harmonised data handling and management and provides the mechanism for FAIR data in collaborative European life-science projects.

FAIR data in the life sciences are stored in well-maintained, long-term sustainable repositories that are recognised and broadly supported by the community. ELIXIR maintains a list of recommended deposition databases for the life sciences.²

Wherever possible, persistent identifiers should be used according to community best-practice, datasets annotated using recognised and publicly-available standards and data should have a license or terms of use that clearly sets out conditions for reuse. This can easily be achieved by depositing data in an ELIXIR recommended deposition database. Much reuse takes place inside computational workflows; data archives should provide formats and annotation that support this.

Well-managed research data in the life sciences generates value far beyond the initial researcher's laboratory: in the research community, industry, education and society at large. The extensive reuse of data from life-science data resources is a testament to the societal value of open data and the FAIR principles.

ELIXIR underpins and drives good data management practice in the life sciences and is committed to making this data available within the framework of the European Open Science Cloud. We base this work on the following guiding principles for FAIR Data Management:

1. Open sharing of research data is a core principle for publicly-funded research and ELIXIR encourages all funders to adopt Open Data mandates.
2. Data Management is crucial part of good scientific practice and research excellence.
3. Whenever possible, biological research data should be submitted to the recommended community deposition databases.
4. All data submitted to Open Data archives must be annotated in accordance with community-defined standards.
5. ELIXIR Nodes are the national implementation of a harmonised FAIR Data Management programme for the life sciences.
6. FAIR data management requires professional skills and adequate resources.
7. Good research data management requires appropriate funding for data infrastructures.

These principles drive ELIXIR's engagement in the European Open Science Cloud and signals our commitment to the 1st EOSC Summit Declaration.

¹ Wilkinson MD, Dumontier M, Aalbersberg JJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, doi:[10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18).

² ELIXIR recommended Deposition Databases: <https://www.elixir-europe.org/platforms/data/elixir-deposition-databases>

ELIXIR's principles for FAIR Data Management

This document sets out the formal position of ELIXIR Europe on good data management practice in the life sciences and the practical implementation of FAIR Data Management principles by ELIXIR Nodes. It represents the collective recommendation of the 21 ELIXIR Nodes on how to implement FAIR in life sciences.

The principles set out in this document will guide ELIXIR's engagement in the European Open Science Cloud. The principles also signal our commitment to the First EOSC Summit Declaration.³ They provide specific recommendations for implementing FAIR and by building capacity and empowering our national Nodes, ELIXIR achieves the coordinated action requested in the EOSC Declaration.

Whilst the document focuses on data, we note that most of the principles outlined are equally applicable to other research assets such as software, training materials and other digital research objects.⁴

1. Open sharing of research data is a core principle for publicly-funded research and ELIXIR encourages all funders to adopt Open Data mandates

- Open access to data and results from publicly-funded research drives reuse and creates value for many different actors in the knowledge economy: other scientists, small companies, schools and other educational institutions.
- Science rests on the well-earned trust from voluntary participants and the public. Research data should be *as Open as possible, as Closed as necessary* to protect the privacy of individuals and legitimate commercial interests of companies.
- Data from human research subjects can be deposited in databases and reused through archives that support *Data Access Control*.⁵
- ELIXIR recognises that, while data and knowledge provided by ELIXIR Services will be accessible, this does not mean that the use of data is unencumbered: restrictions on the use of data may arise due to legal (e.g. data protection requirements, copyright protection, or license restrictions) or ethical considerations.
- ELIXIR Core Data Resources and Deposition Databases should reflect this commitment and have terms of use or a license that enables the reuse and remixing of data. The Creative Commons licenses CCo, CC-BY or CC-BY-SA all conform to the Open Definition (<http://opendefinition.org/licenses/>), as do equivalent open terms of use.⁶

2. Data Management is crucial part of good scientific practice and research excellence

- Good scientific practice includes scientific honesty and diligence (professionalism, forthrightness, transparency) in the stewardship of data, samples, and research results.
- Careful recording and reporting of data provenance, including appropriate data citations, is an integral part of data management.

³ Access the EOSC Declaration here: https://www.elixir-europe.org/system/files/eosc_declaration.pdf

⁴ For instance, the ELIXIR recommendations on Open Source Software are based around FAIR: Jiménez RC, Kuzak M, Alhamdoosh M *et al.* Four simple recommendations to encourage best practices in research software. *F1000Research* 2017, 6:876 (doi: [10.12688/f1000research.11407.1](https://doi.org/10.12688/f1000research.11407.1))

⁵ See e.g. European Genome Phenome Archive, EGA, <https://ega-archive.org/>

⁶ See e.g. RDA/CODATA Legal interoperability WG recommendations: doi: [10.5281/zenodo.162241](https://doi.org/10.5281/zenodo.162241)

- Data should be made available for peer review together with manuscripts; ELIXIR's recommended deposition databases work with publishers to encourage peer review and provide infrastructure supporting pre-publication review.
- Acknowledging data reuse and rewarding data sharing is essential; ELIXIR supports and actively promotes the *Joint Declaration of Data Citation Principles*.⁷

3. Whenever possible, biological research data should be submitted to the recommended community deposition databases

- Structured deposition databases have, during decades of operation, demonstrated the value of FAIR principles for experimental data produced in the biological sciences. These databases employ quality standards and formal metadata descriptions, both of which are critical requirements to ensure that the deposited data can be assessed by other scientists and actively reused.
- They are accessed and trusted by a large user-community on a daily basis and their use should be the default option for long term data stewardship.
- ELIXIR maintains a list of recommended deposition databases for experimental data.⁸
- Data management, including submission, is part of research project planning and evaluation, and is typically required as part of a funding application. Submission of data should be planned for early in the project's life cycle, to ensure FAIR data is captured accurately. Large distributed research projects should have dedicated resources for data coordination, validation and long-term access.
- Funders and journals should encourage researchers to publish their data in ELIXIR's recommended deposition databases, or their equivalent global partner databases, in the first instance.
- The extensive reuse of data from life-science deposition databases shows that combining technical data standards, community requirements and infrastructure for open data, generates high value for science and society.⁹

4. All data submitted to Open Data archives must be annotated in accordance with community-defined standards

- Wherever possible, ELIXIR recommends the use of deposition databases that adhere to community-defined standards and best practices, thereby promoting FAIR principles. If such deposition databases do not exist data can be made FAIR by annotation with community standards and well-known metadata services.
- The ELIXIR Interoperability Platform operates a set of such identifier, ontology and metadata standard services as a resource for the community to select appropriate standards for annotation of data in life sciences.¹⁰

⁷ Data Citation Synthesis Group: Joint Declaration of Data Citation Principles. Martone M. (ed.) San Diego CA: FORCE11; 2014 [<https://www.force11.org/group/joint-declaration-data-citation-principles-final>].

⁸ <https://www.elixir-europe.org/platforms/data/elixir-deposition-databases>

⁹ See e.g. <https://beagrie.com/static/resource/EBI-impact-report.pdf> and <https://dx.doi.org/10.12688/f1000research.7911.1>

¹⁰ E.g. identifiers.org, fairsharing.org, [Ontology Lookup Service](http://ontologylookup.org)

- Use of community best practice for persistent data identifiers¹¹ is critical for long term findability, drives interoperability and underpins data provenance, thus maximising the impact and reuse of life-science data.
- ELIXIR supports scalable data management by the development of data management practices, in close partnership with user communities and other research infrastructures.^{12,13}

5. ELIXIR Nodes are the national implementation of a harmonised FAIR Data Management programme for the life sciences

- Our mission to implement standards-based, FAIR data stewardship within every European life-science project and to help our user community to effectively meet funder requirements is centered around our 21 European Nodes: ELIXIR Nodes support the FAIR Data Management needs of national research projects.¹⁴
- ELIXIR's Nodes, which bring together over 180 institutes across Europe, are responsible for the implementation of FAIR principles within our member states: through these national bioinformatics infrastructures we drive the adoption of FAIR data management practice within national communities.

6. FAIR data management requires professional skills and adequate resources

- ELIXIR is committed to develop the expertise needed for data management in the life sciences through a comprehensive and scalable training programme in our Nodes.
- ELIXIR has established a European-wide Training Programme to train researchers and developers from the life science community in FAIR data management, and to increase the number of trainers (capacity building) with a dedicated Train the Trainer programme.¹⁵
- Through Bring-Your-Own-Data (BYOD), Software and Data Carpentry workshops and the strengthening of genome annotation expertise, ELIXIR Nodes actively build the capabilities needed in every research group for FAIR data stewardship.¹⁶
- ELIXIR Training materials are FAIR and made available through the training portal, TeSS (tess.elixir-europe.org).¹⁷
- ELIXIR collaborates with GOBLET (the Global Organisation for Bioinformatics Learning Education and Training, www.mygoblet.org) to work together in defining and implementing open materials, standards and best practices for training in the life sciences.¹⁸

7. Good research data management requires appropriate funding for data infrastructures

- ELIXIR's coordinated data management infrastructure is based on national investments (through research infrastructure roadmaps) in the services provided by ELIXIR Nodes,

¹¹ See e.g. McMurry JA, Juty N, Blomberg N, Burdett T, Conlin T, Conte N, et al. Identifiers for the 21st century: How to design, provision, and reuse persistent identifiers to maximize utility and impact of life science data. PLoS Biol. Public Library of Science; 2017 Jun 29;15(6):e2001414. <https://doi.org/10.1371/journal.pbio.2001414>

¹² For example: Hoopen PT, Finn RD, Bongo LA, Corre E, Fosso B, Meyer F, et al. The metagenomic data life-cycle: standards and best practices. GigaScience 2015 4:1. 2017 Jun 16. [doi:10.1093/gigascience/gix047](https://doi.org/10.1093/gigascience/gix047)

¹³ For example: EMBRIC Configurator services:

http://www.embric.eu/sites/default/files/deliverables/D4.2_Configurator%20service%20for%20EMBRIC.pdf

¹⁴ See e.g. <http://www.sib.swiss/services-resources/embedded-bioinformaticians>, <https://bils.se/resources/support.html>, <https://www.dtls.nl/fair-data/fair-data/fair-data-knowledge-expertise/>

¹⁵ 53 trainers in 12 Nodes trained in 2016-17, see: Train the Trainer Pilot Year Report -

<https://docs.google.com/document/d/1DPZb6XvRmkcOE1LKGR60-kX6lTuaZEr7lWKSay32PEE/edit>

¹⁶ Pawlik A, van Gelder CWG, Nenadic A et al. Developing a strategy for computational lab skills training through Software and Data Carpentry: Experiences from the ELIXIR Pilot action. F1000Research 2017, 6:1040 ([doi: 10.12688/f1000research.11718.1](https://doi.org/10.12688/f1000research.11718.1))

¹⁷ 773 training materials from 41 providers in life sciences are now in TeSS. A growing number encoded in standards such as bioschemas.org and cross-linked to tools (<https://bio.tools>) and standards (<https://fairsharing.org>)

¹⁸ https://www.elixir-europe.org/system/files/goblet-elixir_partnership-final.pdf

pooled national funding (through the ELIXIR Hub) for coordination of new activities (ELIXIR Implementation Studies) and European grants for service development and pan-European access.

- The long-term, international funding of research data repositories, such as ELIXIR's Core Data Resources, is a critical component of this infrastructure. ELIXIR works with the community and funders to encourage the development of appropriate funding models¹⁹ and global action to maintain data repositories.²⁰
- ELIXIR Core Data Resources²¹ represent our commitment to a community-defined, sustainable, data management infrastructure, based on objective indicators of use, governance, community engagement and impact. Inclusion is, and must be, based on transparent, expert-driven peer review.
- ELIXIR welcomes the recognition in the EOSC Declaration²² of the role for community data repositories and the need for long-term sustainable funding of Research Data Management infrastructures.

About ELIXIR

ELIXIR (www.elixir-europe.org) unites Europe's national bioinformatics Nodes into a single infrastructure that underpins the management and safeguarding the increasing volumes of data generated by publicly-funded research. The national ELIXIR Nodes and EMBL-EBI, coordinate and integrate bioinformatics resources across the member states and enable users in academia and industry to access services that are vital for their research.

ELIXIR is a distributed infrastructure; ELIXIR services are operated by our 21 European ELIXIR Nodes²³ which comprise over 180 leading universities and centres of excellence. These services include:

- Databases (both data deposition databases and knowledge bases)
- Analysis tools and software including data reproducibility mechanisms
- Interoperability services and expert support for FAIR data stewardship
- Compute services
- Data analysis and management expertise and support for national users
- Training

ELIXIR services support users across a broad spectrum of life sciences -- from rare diseases and human data through to marine metagenomics, plant sciences and biodiversity. ELIXIR has services for, and expertise in, data management across life-science disciplines, including genomics, proteomics, metabolomics and image data.

FAIR in ELIXIR

Safeguarding and developing the extensive reuse of life-science data -- i.e. making data Findable, Accessible, Interoperable and Reusable -- is the fundamental purpose of ELIXIR as an infrastructure, and permeates all major areas of ELIXIR's operations:

¹⁹ <https://www.elixir-europe.org/news/elixir-publishes-position-paper-interim-evaluation-horizon-2020>

²⁰ <https://www.elixir-europe.org/news/elixir-among-signatories-call-action-global-coalition-sustain-core-data-resources>

²¹ <https://www.elixir-europe.org/platforms/data/core-data-resources>

²² https://www.elixir-europe.org/system/files/eosc_declaration.pdf

²³ ELIXIR Nodes: <https://www.elixir-europe.org/about/elixir-nodes>

- **ELIXIR Data Platform** has developed a process for establishing a set of Core Data Resources²⁴ and a list of recommended Deposition Databases²⁵ that depend on evaluation against a set of indicators. The indicators have been mapped against FAIR principles.²⁶
- **ELIXIR Interoperability Platform** develops and maintains data standards, metadata standards, formats and methods to make data findable, accessible and interoperable -- and provide these through the services the community needs to make data FAIR.
- **ELIXIR Tools Platform** allows users to identify the best methods for making data FAIR -- through registries and benchmarking -- and addresses the issue of reproducibility of tools and workflows through containers.
- **ELIXIR Compute Platform** ties together national and European e-Infrastructure services to allow life-science users to authenticate, access, distribute and compute on FAIR data within a European Open Science Cloud.
- **ELIXIR Training Platform** trains researchers, developers and trainers in the tools and practices required to manage their data and make it FAIR, and builds a group of specialist trainers in the Nodes who can support users across Europe.
- **ELIXIR Data Management Plans Working Group** brings together ELIXIR experts across Nodes to align practices and identify opportunities and service gaps. This peer group is developing a Data Management Planning Portal Website.²⁷

²⁴ <https://www.elixir-europe.org/platforms/data/core-data-resources>

²⁵ <https://www.elixir-europe.org/platforms/data/elixir-deposition-databases>

²⁶ <https://f1000research.com/articles/5-2422/>

²⁷ <https://dmp.fairdata.solutions>