New Charter proposal for

Life Science Data Infrastructures IG

Adapted from the 2014 charter of the [ELIXIR Bridging Force IG](https://www.rd-alliance.org/groups/elixir-bridging-force-ig.html)

# Life Science Data Infrastructures IG

The life sciences are becoming increasingly data intensive, owing much to the huge improvements seen in large-scale gene sequencing and other molecular “omics” techniques. There is a need for large-scale sustainable and interoperable data management and storage methods that allow secure and easy access to and reuse of these highly complex data. Simultaneously, as omics-focussed life science research projects increasingly depend on more than one type of measurement, there is a widely felt need for the ability to integrate different data types. Phenotyping, bioimaging and biosample management can all be considered part of life science data, and omics-focussed life science data systems will need to interoperate with these. Different and diverse sectors of the life sciences, such as health, food production, bioindustry and the environment experience similar issues. Interoperability between different research domains is thus a necessary condition to allow for the emergence of cross-disciplinary data. It is no surprise that the life sciences gave birth to the FAIR principles.

The Life Science Data Infrastructures Interest Group is formed to serve as a bridge between life science data infrastructures in different regions of the world and relevant RDA Interest Groups, including both specific subtopics of the life sciences such as agricultural data, marine data, structural biology, toxicogenomics, as well as generic topics that can/should be applied in the life sciences such as big data analysis, federated identity management, and data publishing.

The IG aspires to have meaningful representation from diverse geographical regions including North and South America, the EU, UK and Eastern Europe; Africa, Asia and Oceania; and will actively seek participation from under-represented groups and the global South. The IG also aims for fair gender representation and inclusivity in all its activities.

## Collaborating Life Science Infrastructures

The interest group is currently an initiative of four regional infrastructure organisations:

* **ELIXIR** (<http://www.elixir-europe.org/>) has been established to build a sustainable pan-European research infrastructure for biological information providing support to life science research including medicine, agriculture, bioindustry and society. ELIXIR did originally take the initiative for this IG that started in 2014 as the ELIXIR Bridging Force IG and grew into the Life Science Data Infrastructures IG from initiatives in 2019.
* The **Australian BioCommons** (<https://www.biocommons.org.au/>) is building digital infrastructure to ensure Australian life science research remains globally competitive, by providing access to the tools, methods and training researchers require to respond to national challenges such as food security, environmental conservation and disease treatment
* In the USA, representation is through NIH Office of Data Science Strategy (<https://datascience.nih.gov>). Support for life science data infrastructure includes the **Science and Technology Research Infrastructure for Discovery, Experimentation, and Sustainability (STRIDES)** Initiative**. STRIDES** allows researchers to explore the use of cloud environments to streamline data use by partnering with commercial providers, Google, AWS and Microsoft Azure. Leveraging the STRIDES partnership with cloud service providers, NIH’s is connecting our various data systems in a critical step toward improving researchers’ access to all types of data. The [NIH Cloud Platform Interoperability (NCPI) effort](https://anvilproject.org/ncpi/) seeks to create a federated genomic data ecosystem and is a collaborative project between NIH and external partners (<https://datascience.nih.gov/nih-cloud-platform-interoperability-effort>).
* **H3ABioNet** (<https://www.h3abionet.org>) was established to develop bioinformatics capacity in Africa and specifically to enable genomics data analysis by H3Africa researchers across the continent. It has developed several key resources to serve African genomics data to the global scientific community.

At any time, this list will be open to other regional initiatives with a compatible remit. At no time will there be any limitations to membership of the interest group; this list only exists to provide an anchor to regional implementation of data standards in the life sciences.

## Existing collaborations

There are existing international collaborations on sub-topics of life science data for which we will make sure of appropriate representation in our RDA IG. Examples are:

* Global BioData Coalition (<https://globalbiodata.org/>), on sustainability of core data resources.
* Global Alliance for Genomics and Health (<https://www.ga4gh.org>), on standards for human genomic data sharing.
* Proteomics Standards Initiative (<https://www.hupo.org/Proteomics-Standards-Initiative>)

Our continued collaboration with the RDA FAIRsharing WG (<https://rd-alliance.org/group/fairsharing-registry-connecting-data-policies-standards-databases.html>) will also ensure we connect with data resources and standardization initiatives in life science and beyond.

# Co-chairs

The IG will be organised by a group of co-chairs that is representative of the geographic diversity.

Until now, Rob Hooft, Carole Goble and Bengt Persson, all three representing Europe / ELIXIR, have been co-chairs of the Bridging Force IG. They will step down and leave the lead of the changed group to new candidates with the proposed transformation.

Candidates to take up the role of Co-chairs in 2022 are:

* Jeff Christiansen; representing Australian BioCommons
* Nicola Mulder; representing H3ABioNet
* Rob Hooft; ad-interim representative of ELIXIR
* Susan Gregurick; representing NIH.

Co-chairs will be proposed by the represented infrastructures, and appointed by approval of the membership of the IG; co-chairship will be reconsidered at least every three years.

# Plans

Primary goal of the Interest Group will be to organise sessions at each plenary that focus on a (possibly timely) topic of interest to the participating infrastructures, and to bring together observations of interest from other sessions at the plenary. Secondary goals are to discuss problems encountered in life science data infrastructure in RDA context, and to make sure that RDA results are adopted in the life sciences where appropriate.

When a topic requires more in-depth examination, the setup of an RDA Working group will be considered. Topics could include:

* Interoperability of different kinds of large-scale (“omics”) data
* Sustainability of life science data identifiers
* Mechanisms for secure access to personal data, including authorisation and authentication issues
* Strategies for data storage allowing for computationally intensive analyses
* Distributed/federated data analysis

The tasks of the Life Science Data Infrastructures IG will be performed where possible in collaboration with other RDA Interest Groups and Working Groups.