










OPINION ARTICLE

The ELIXIR Biodiversity Community: Understanding short- and long-term changes in biodiversity [version 1; peer review: awaiting peer review]

Robert M. Waterhouse ¹, Anne-Françoise Adam-Blondon², Bachir Balech ³,
Endre Barta⁴, Katharina F. Heil ⁵, Graham M. Hughes ⁶, Lars S. Jermiin ⁷,
Matúš Kalaš ⁸, Jerry Lanfear ⁵, Evangelos Pafilis⁹, Aristotelis C. Papageorgiou¹⁰,
Fotis Psomopoulos ¹¹, Niels Raes ¹², Josephine Burgin¹³, Toni Gabaldón ^{14,15}

¹Naturalis Biodiversity Center, Leiden, South Holland, The Netherlands

²EMBL-EBI, Wellcome Genome Campus, Hinxton, England, CB10 1SD, UK

³Centro Nacional de Supercomputacion, Barcelona, Catalonia, Spain

⁴Institut de Recerca Biomedica, Barcelona, Catalonia, Spain

⁵Department of Ecology and Evolution, SIB Swiss Institute of Bioinformatics, Universite de Lausanne, Lausanne, Vaud, 1015, Switzerland

⁶INRAE, BioinfOmics, Plant Bioinformatics Facility, Universite Paris-Saclay, Gif-sur-Yvette, Île-de-France, 78026, France

⁷Istituto di Biomembrane, Bioenergetica e Biotecnologie Molecolari, Bari, 70126, Italy

⁸Institute of Genetics and Biotechnology, Magyar Agrar- es Elettudományi Egyetem, Gödöllő, Pest County, Hungary

⁹ELIXIR, Wellcome Genome Campus, Hinxton, England, CB10 1SD, UK

¹⁰School of Biology and Environmental Science, University College Dublin, Dublin, Leinster, Ireland

¹¹Systems Biology Ireland, School of Medicine, University College Dublin, Dublin, Leinster, Ireland

¹²Department of Informatics, Universitetet i Bergen, Bergen, Hordaland, Norway

¹³Institute of Marine Biology, Biotechnology and Aquaculture, Hellenic Centre for Marine Research, Heraklion, 71003, Greece

¹⁴Department of Molecular Biology and Genetics, Democritus University of Thrace, Alexandroupolis, Greece

¹⁵Institute of Applied Biosciences, Centre for Research and Technology Hellas, Thessaloniki, Greece

V1 First published: 15 May 2023, 12(ELIXIR):499
<https://doi.org/10.12688/f1000research.133724.1>

Latest published: 15 May 2023, 12(ELIXIR):499
<https://doi.org/10.12688/f1000research.133724.1>

Abstract

Biodiversity loss is now recognised as one of the major challenges for humankind to address over the next few decades. Unless major actions are taken, the sixth mass extinction will lead to catastrophic effects on the Earth's biosphere and human health and well-being. ELIXIR can help address the technical challenges of biodiversity science, through leveraging its suite of services and expertise to enable data management and analysis activities that enhance our understanding of life on Earth and facilitate biodiversity preservation and restoration. This white paper, prepared by the ELIXIR Biodiversity Community, summarises the current status and responses, and presents a set of plans, both technical and community-oriented, that should both enhance how ELIXIR Services are applied in the biodiversity field and how ELIXIR builds connections across the many

Open Peer Review

Approval Status Awaiting Peer Review

Any reports and responses or comments on the article can be found at the end of the article.

other infrastructures active in this area. We discuss the areas of highest priority, how they can be implemented in cooperation with the ELIXIR Platforms, and their connections to existing ELIXIR Communities and international consortia. The article provides a preliminary blueprint for a Biodiversity Community in ELIXIR and is an appeal to identify and involve new stakeholders.

Keywords

White Paper, ELIXIR Strategy, Community Roadmap, Biodiversity Networks, Biodiversity Informatics, Environmental Systems, Data Science



This article is included in the [ELIXIR gateway](#).



This article is included in the [Ecology and Global Change gateway](#).

Corresponding author: Robert M. Waterhouse (robert.waterhouse@gmail.com)

Author roles: **Waterhouse RM:** Conceptualization, Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing; **Adam-Blondon AF:** Writing – Review & Editing; **Balech B:** Writing – Review & Editing; **Barta E:** Writing – Review & Editing; **Heil KF:** Project Administration, Writing – Review & Editing; **Hughes GM:** Writing – Review & Editing; **Jermiin LS:** Writing – Review & Editing; **Kalaš M:** Writing – Review & Editing; **Lanfear J:** Conceptualization, Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing; **Pafilis E:** Writing – Review & Editing; **Papageorgiou AC:** Writing – Review & Editing; **Psomopoulos F:** Writing – Review & Editing; **Raes N:** Writing – Review & Editing; **Burgin J:** Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Gabaldón T:** Conceptualization, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: The authors acknowledge the following grant support: Swiss National Science Foundation (SNSF) grant number PP00P3_202669 to RMW; Swiss State Secretariat for Education, Research and Innovation (SERI) contract number 22.00173. *The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

Copyright: © 2023 Waterhouse RM *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Waterhouse RM, Adam-Blondon AF, Balech B *et al.* **The ELIXIR Biodiversity Community: Understanding short- and long-term changes in biodiversity [version 1; peer review: awaiting peer review]** F1000Research 2023, 12(ELIXIR):499 <https://doi.org/10.12688/f1000research.133724.1>

First published: 15 May 2023, 12(ELIXIR):499 <https://doi.org/10.12688/f1000research.133724.1>

Introduction

Biodiversity threats and challenges

Biological diversity—or biodiversity—refers to the variety and variability of life on Earth, encompassing genetic and species diversity at the levels of populations, communities, and ecosystems. Biodiversity reflects the ever-changing natural balance that has evolved over billions of years, sustaining communities of interdependent and interacting organisms. Those balances form the basis of a healthy Earth, including the ecosystem functions that support human well-being (*i.e.*, ecosystem services). With growing demands on nature due to human activities, the Anthropocene is upsetting this balance and is consequently witnessing an unprecedented loss of biodiversity globally (WWF, 2022; Johnson *et al.*, 2017). These declines pose a grave threat to humanity, the severity of which is increasingly recognised by international organisations, regional bodies, national governments, and society.

Strategies to protect and restore biodiversity are wide-ranging in scope and scale, with policies and actions that require broad support to be feasible and effective *e.g.*, goals 12-15 of the 17 Sustainable Development Goals (SDGs) adopted by the United Nations (UN, 2015). Biodiversity research aimed at building the knowledge and resources that inform management practices and policy is equally wide-ranging, often bringing together researchers from different disciplines, such as taxonomists, ecologists, evolutionary biologists, and informaticians. This is particularly true for the growing field of interdisciplinary research taking advantage of molecular sequence data, which recognises the relevance of and advantages offered by genetic and genomic data in biodiversity assessment, monitoring, conservation, and restoration (Hoban *et al.*, 2021; Lewin *et al.*, 2022). Connecting such molecular sequence data with biodiversity research infrastructures (see *Extended Data (Waterhouse, 2023)*) and resources is a critical step towards facilitating exchange of knowledge, sharing, and interoperability of large and complex datasets (Waterhouse *et al.*, 2022).

The need for informatics solutions to address these challenges inspires many scientists from across the ELIXIR Nodes to increasingly engage in different aspects of biodiversity research. This stems from a natural alignment with ELIXIR's overarching mission to support the management of public research data, integrate and coordinate life science resources, and foster the development of innovative services and technical solutions in Europe (Harrow *et al.*, 2021). Here we present the ELIXIR Biodiversity Community, comprised of researchers from different disciplines, united by a shared recognition of the main societal and informatics challenges, as well as key scientific and organisational opportunities; how these connect with ELIXIR Platforms and other ELIXIR Communities, as well as with the wider “ecosystem” of biodiversity projects and infrastructures; and set out our roadmap for building on ELIXIR expertise to grow the ELIXIR Biodiversity Community and engage with the development of resources and infrastructures to support biodiversity research.

Societal challenges and global responses

Biodiversity represents the variety of organisms on the planet at all taxonomic levels, a result of a long and complex evolutionary process. Biodiversity is essential for life itself, for the adaptation of populations, species, communities, and ecosystems towards rapid change in biotic and abiotic parameters, including climate change. From a human standpoint, biodiversity forms the foundation of ecosystem services that are indispensable for human well-being and a healthy planet, and has long been a source of adaptive solutions or innovations in several critical areas such as food production. Despite its importance, biodiversity has been declining at a mass-extinction-level rate (IPBES, 2019) over the last decades. The unsustainable human development model has increased pressures on biodiversity, through climate change (IPCC, 2022; Wezel *et al.*, 2020), invasive species, habitat loss and degradation, and the depletion of natural resources (IPBES, 2019). The decline of biodiversity at this rate often creates unpredictable threats and changes to ecological oscillations, such as the increased risk of new human diseases (Frumkin & Haines, 2019), the collapse of ecosystem services, the degradation of natural resources, and the increased possibility of a global food crisis (FAO, 2019).

At the same time, scientists and naturalists do not even know what is being lost, as around 80% of biodiversity at the species and population levels remains undescribed and/or underrepresented in inventories and databases (Mora *et al.*, 2011; Costello *et al.*, 2013; Moura & Jetz, 2021; Bispo *et al.*, 2022; Boekhout *et al.*, 2022; Chimeno *et al.*, 2022). Furthermore, most research and monitoring efforts tend to focus on a limited number of biodiversity levels or elements. While there is significant literature around biodiversity loss (*e.g.*, a Scopus query [13.09.2022] for “biodiversity loss” returns 33,324 documents), there is a very limited effort in reviewing biodiversity using high-throughput data (Scopus query [13.09.2022] for “Biodiversity loss” AND (“omics” OR “genomics” OR “metagenomics”) returns only 1,795 documents). This clearly indicates a bias in reporting, which has repercussions on the decision-making process pertaining to biodiversity conservation efforts (Gadelha *et al.*, 2021). This brings forward an additional challenge of shifting perspectives from narrow, low-throughput efforts towards more holistic, high-throughput initiatives, including better citizen scientist contributions towards these efforts. Humanity may miss important solutions to key problems for its survival, such as the loss of important genetic variants among wild plants, animals, and microbes for agriculture (Nic Lughadha *et al.*, 2020) and for dealing with health issues (Marselle *et al.*, 2021).

Following the 1992 United Nations Convention on Biological Diversity (CBD), governments and international organisations have responded to the decline of biodiversity with policies, and restoration and protection strategies. However, the initial goals of these have not been reached and biodiversity decline continues accelerating (IPBES, 2019; Turvey & Crees, 2019; WWF, 2022). For the new targets set by the post-2020 global biodiversity framework (GBF, 2023) to succeed, research is considered to be key, especially the interaction between science, society, and policy makers (Blicharska *et al.*, 2019; Hermoso *et al.*, 2022; Nature, 2022), with net improvements by 2050 to achieve the CBD's vision of "living in harmony with nature by 2050". Today, scientists recognise the important roles that genetic and genomic data can play in biodiversity discovery, assessment, monitoring, conservation, and restoration, to ensure the long-term resilience of ecosystems (Hoban *et al.*, 2020; Gadelha *et al.*, 2021; Segelbacher *et al.*, 2022; Formenti *et al.*, 2022; Theissinger *et al.*, 2023). The contribution of genomics and bioinformatics towards these targets, and many of the associated technical and scientific challenges are described in Waterhouse *et al.* (2022), together with the possible contribution of the ELIXIR European Strategy Forum for Research Infrastructures to meet them.

Scientific opportunities in biodiversity research

Biodiversity researchers are increasingly realising the potential offered by modern technologies, particularly in genomics, to create new opportunities for developing tools and resources that will transform the field. These opportunities lie primarily in the types of scientific applications that are becoming more feasible and scalable through continued advances in genomics technologies alongside enhanced data management systems. A long-term vision sees a future where sequence-based biodiversity monitoring at scale becomes the default and provides the means for ecosystem biodiversity characterisation in space and time. In support of realising these opportunities, ongoing global and regional efforts are building capacity to generate catalogues of reference DNA barcodes (International Barcode of Life, iBOL, BIOSCAN) (Hobern, 2021) and genomes by the Earth BioGenome Project (EBP) (Lewin *et al.*, 2018, 2022) as well as the European Reference Genome Atlas (ERGA, 2023), or both by the Biodiversity Genomics Europe (BGE, 2023) project. Along with this increased production, concurrent development of the necessary tools and resources will greatly enhance our abilities to:

- Maintain and query increasingly comprehensive reference DNA barcode and genome catalogues, improving taxonomic coverage and differentiation (including of cryptic species), and coordinating the efforts of various initiatives under global and regional umbrellas;
- Connect and integrate these molecular resources with other biodiversity data (traits, observations, literature, *etc.*), using an increasingly standardised and harmonised taxonomic framework as the common backbone;
- Use these integrated resources for applied data-driven science to understand the diversity of extant life on Earth, how that diversity functions and interacts, and how it responds to changing environmental pressures;
- Implement monitoring of lesser-known or complex ecosystems, including for enhancing understanding of species interactions and dynamics, as well as for species discovery and exploration of "dark taxa" *e.g.*, Rahman *et al.* (2022);
- Include assessments of within-species, population-level genetic diversity to support characterisations of their evolutionary histories and predictions of their future prospects in the face of ongoing climatic changes;
- Operationalise the assessment of Essential Biodiversity Variables (EBVs) across taxa and spatiotemporal scales, focusing on species distribution and abundance (Kissling *et al.*, 2018; Jetz *et al.*, 2019);
- Engage with naturalists and citizen scientist groups through the use of new technologies that help build a democratised monitoring framework and improve characterisation of ecosystem biodiversity in space and time (Robinson & Peres, 2021);
- Evaluate biodiversity declines, as well as population-level adaptation and migration processes, in the context of anthropogenic activities (*e.g.*, climate change and urbanisation consequences), and understand key aspects necessary to restore ecosystem functions to help prioritise biodiversity conservation, restoration, and "rewilding" efforts (*e.g.*, particularly relevant to at-risk biodiversity hotspots).

Organisational opportunities and ELIXIR's roles

The field of biodiversity assessment and research, from an organisational context, is broad, complex, and distributed. There are a multitude of organisations that operate across international borders, within countries, and at a local level

(see *Extended Data* (Waterhouse, 2023)). This landscape is further demarcated along scientific and technical lines, with organisations that focus on taxonomies, ecology, molecular sciences, and method development (necessitated by the increasingly large and complex amount of data being generated). ELIXIR, perhaps uniquely, stands as a hub for the molecular sciences and bioinformatics at an international and national level across many scientific disciplines (Waterhouse *et al.*, 2022). Biodiversity research and infrastructures increasingly rely on molecular data, so ELIXIR is well placed to lead organisational alignments and collaborations: from a core set of partners across Europe mainly within the field of molecular sciences, to an expanding variety of partner organisations that focus on other biodiversity-related research and resources (see below for examples from the ecosystem of biodiversity projects, resources, and infrastructures). Importantly, this extends beyond the data themselves as FAIRification of digital research objects, championed by ELIXIR's Services and Platforms, is increasingly recognised as essential in biodiversity research. Opportunities to help coordinate and align organisational activities in the biodiversity domain arise naturally from ELIXIR's established European-wide "network of networks" approach, connecting to existing initiatives at both the national and international levels. With ELIXIR's strengths in molecular sciences, a "hub and spokes" model would help augment opportunities to connect molecular-focused bioinformatics tools, protocols, and resources with the many other biodiversity-related infrastructure and stakeholder organisations.

Informatics challenges facing biodiversity infrastructures and resources

The variety of existing biodiversity data infrastructures and resources is a testament to the long-standing recognition by multiple stakeholders of their importance, currently reflected in the growing European and global commitments to prevent further biodiversity decline and ensure the long-term health of ecosystem services. This heterogeneity, however, gives rise to many challenges, both technical in terms of data analysis (due to inadequacies of existing methodologies), data integration and data interaction, and at the level of the scientific community, which faces a heterogeneous landscape of infrastructures and resources that can be difficult to navigate. The methodological and logistical challenges range from scaling up (needed to be able to process the increasing amounts of complex molecular data) to the management of these data and working on connecting them to other biodiversity research infrastructures (Waterhouse *et al.*, 2022). The biodiversity community needs to proactively seek common solutions (without unnecessary duplication of effort) that enable molecular technologies to advance biodiversity research. To this end, informatics solutions will need to be developed to address the practicalities of common challenges, such as:

- The need to constantly incorporate knowledge-based updates and resolve conflicts to maintain standardised taxonomies that serve as a dynamic framework that facilitates interoperability across research infrastructures;
- Building data and metadata brokering services that support coordinated community engagement to ensure good data management through technical infrastructures for aiding and automating data submission;
- Developing the means, through text mining and curation, to identify and liberate in digital form invaluable historical or baseline data trapped in the literature (including those published in non-English sources), or in museum and other natural history collections;
- Improving the accessibility of research results through publications (*e.g.*, by making published traits, tables, treatments, specimens, figures *etc.*), citable and reusable (*e.g.*, through nanopublications), and including identifiers of cited elements (genes, specimens, taxonomic names, treatments);
- Improving and harmonising currently highly heterogeneous metadata collection standards to promote the adoption of community best practices that will maximise findability, accessibility, interoperability, and reusability of digital research objects (*i.e.*, drive biodiversity research towards FAIR compliance);
- Scaling up of services for data and metadata management to keep pace with and accommodate the increases in data production (*e.g.*, genomics) and collection (*e.g.*, Essential Biodiversity Variables);
- Developing frameworks that deliver an increasingly integrated and interconnected landscape of biodiversity research infrastructures, utilising developments in application programming interfaces and Semantic Web services;
- Ensuring widespread access to high-performance computing (HPC) and HPC-deployable software and data-management systems, including containers and workflows, to enable decentralised efforts while promoting standardisation.

The ELIXIR Biodiversity Community: An “ecosystem” of projects

Tackling the biodiversity crisis at a general level is not going to be resolved through a single action, but instead requires a complex set of interacting actions that are co-dependent but usually funded separately. ELIXIR can assume a key leading role in a subset of those actions, focused on data management and the molecular sciences, but even at the level of ELIXIR, there are a multitude of funded projects at a transnational, national, and local level. These form a complex network of interacting projects that have distinct but related aims, usually focused on establishing communities and connections and/or building new technical solutions to help with data access, storage, or analysis. ELIXIR can serve a critical function here, as a fundamental aspect of its mission is to make connections and coordinate across complex activities. [Table 1](#) lists a subset of ongoing projects across Europe and within ELIXIR member states that illustrate the breadth of activities underway.

Table 1. Summaries of a selection of transnational and national biodiversity-related projects in which ELIXIR Nodes are involved.

Project	Node/Funder	Summary details/description
ARISE	Netherlands	ARISE (Authoritative and Rapid Identification System for Essential biodiversity information) is a digital infrastructure with a mission to provide semi-automated identification of all multicellular species in the Netherlands (van Ommen Kloeke et al., 2022).
BiCIKL	E.C. (coordinated by Pensoft)	BiCIKL (Biodiversity Community Integrated Knowledge Library) will catalyse a culture change in the way biodiversity data is identified, linked, integrated and re-used across the research cycle. We will cultivate a more transparent, trustworthy and efficient research ecosystem.
Biodiversity Genomics Europe (BGE)	E.C. (coordinated by Naturalis Biodiversity)	By bringing together Europe’s key practitioners in two fundamental DNA-based technologies - DNA barcoding and genome sequencing - the BGE consortium aims to streamline the rollout of these methods across Europe.
Biodiversity Digital Twin (BioDT)	E.C. (coordinated by CSC - IT CENTER FOR SCIENCE LTD.)	The Biodiversity Digital Twin prototype will provide advanced models for simulation and prediction capabilities, through practical use cases addressing critical issues related to global biodiversity dynamics.
Curated collections of DNA barcode marker	Italy	A reference collection of COXI mitochondrial DNA genes based on the integration of sequence and taxonomy data of BOLD and ENA (Balech et al., 2022).
e-BioDiv	Switzerland	Open Biodiversity FAIR-ification Services for Biospecimens stored in Swiss Natural History Museums
Earlham Institute Barcoding the Broads	UK	A Wellcome-funded programme of public engagement events and activities to explore biodiversity on the Norfolk Broads, led by the Earlham Institute as part of the work on the Darwin Tree of Life project.
ELIXIR Norway	Norway	Dedicated national ELIXIR Node funding (2022-2026) includes a focus on biodiversity and connections to other biodiversity infrastructures and projects in Norway (e.g., the Earth Biogenome Project Norway: EBP-Nor).
Establishment of an ELIXIR Contextual Data Clearinghouse	ELIXIR (Implementation study)	The objective is to develop and deploy an “ELIXIR Contextual Data Clearinghouse” for extending, correcting and improving publicly available annotations on records in sample and sequencing data resources.

Table 1. *Continued*

Project	Node/Funder	Summary details/description
Molecular Biodiversity Greece Community (MBGC)	Greece	Greece is a biodiversity hotspot and to this end, a network of networks covering different disciplines of molecular biodiversity research has been developed. MBGC aims to channel the flow of information amongst researchers, institutions, policy makers, stakeholders and local communities, remaining aligned to all relevant initiatives and infrastructures, at the national, European, and global level.
NFDI4Biodiversity	Germany	Network of diverse biodiversity data (not only molecular). Data are provided by research organisations and projects (e.g., GBOL), public authorities, professional societies and citizen initiatives. Data Management oriented. The production of the data itself is done through use cases.
Phylogenetic methodology	Ireland	A range of analytical tools is being developed to augment the bioinformatics tool kit for comparative genome analysis.
Pole National de Données de Biodiversité	France	National centre of data on biodiversity: the data are provided by the same diversity of channels as in Germany and the role of PND is to support FAIR data management.

Connections with ELIXIR Platforms and Communities

ELIXIR as a Research Infrastructure is structured around (technological) Platforms as well as (user) Communities. Both of these interact on an ongoing basis, mutually supporting each other's efforts. The ELIXIR Biodiversity Community is already collaborating with some of these and aims to broaden interactions to fully leverage the available potential and resources. Some examples of current and future interactions with ELIXIR Platforms (Tools, Compute, Data, Training, and Interoperability) are:

- The Tools Platform provides services for finding software tools and web portals (Bio.tools (Ison *et al.*, 2019), including the <https://biodiversity.bio.tools> subdomain to be populated by the ELIXIR Biodiversity Community), software containers (BioContainers (da Veiga Leprevost *et al.*, 2017)), and workflows (WorkflowHub (Goble *et al.*, 2021)); for assessing tools (OpenEBench (Capella-Gutierrez *et al.*, 2017)); and the best practices in providing research software (Jiménez *et al.*, 2017)). EDAM ontology enables annotation and search of tools and other research objects by application domain, task, or data (Black *et al.*, 2022); and an extended coverage of biodiversity research concepts could be achieved *via* engagement with the Biodiversity Community.
- Specifically for the Compute Platform: User accessible compute, potentially controlled user access *via* Authentication and Authorisation Infrastructure (AAI).
- Community data-management support, and integration with ELIXIR Core and Deposition Data resources. The European Nucleotide Archive (ENA) is a critical data deposition resource for biodiversity genomics data. A concrete example of metadata management workflow is that developed between biodiversity scientists, the Data Platform, and the Biodiversity Community Integrated Knowledge Library (BiCIKL) project (Penev *et al.*, 2021): a metadata management workflow employs the PlutoF tool for biodiversity data and metadata management (Abarenkov *et al.*, 2010), and the ELIXIR Data Platform services.
- Networks of tool/infrastructure users and developers to augment the Training Platform offerings (e.g., with specific courses covering aspects such as: genome annotation, meta-data brokering, *etc.*) and more complete learning paths, covering entire workflows (e.g., from sequencing to annotation, possibly covered *via* Galaxy).
- A growing necessity in the biodiversity field towards connected data, as championed by the Interoperability Platform, concretely touching on resources like: RO-Crate and link to specimens, RDMkit, FAIRsharing, Bioschemas and the FAIRcookbook. The ELIXIR Biodiversity Community aims to bring together researchers producing the data, in all their varied forms, with informaticians developing interoperability solutions, to help overcome the challenges of data heterogeneity in the field.

Table 2. Examples of links between the ELIXIR Biodiversity Community and other ELIXIR Communities.

Community	Shared activities
Food & Nutrition	Conceptualisation and implementation of interoperability data models able to integrate, standardise and harmonise data from different disciplines: metagenomics, metabolomics and transcriptomics.
Galaxy	Thousands of tools, including hundreds for biodiversity and microbial/microbiome analysis, are ready to be used on publicly-accessible HPC resources, together with workflows for data processing, which can be versioned, annotated, and shared for reuse. The European Galaxy server (https://usegalaxy.eu) offers access to 2700+ tools and workflows. Galaxy-Ecology is its subdomain piloted by the French ELIXIR Node. A training material repository (https://training.galaxyproject.org) is open for everyone to use and contribute to, providing slides, hands-on tutorials, and other material on using Galaxy to analyse data, with 260+ tutorials in 20+ topics including ecology, microbiome, and climate. Integration of PlutoF and other biodiversity tools into Galaxy could be carried out together with the Biodiversity Community in the near future.
Marine Metagenomics/ Microbiome	Meta-genomic workflows and data archiving. Marine sample metadata annotation guidelines.
Plant Science	Taxonomy framework; coherent/consistent metadata standards for samples (see also interoperability PF (platform), MIAPPE (Minimum Information About Plant Phenotyping Experiments)). Alignment between the MIAPPE standard and exchange formats and the relevant TDWG (Biodiversity Information Standards) standards and exchange formats. Integration and linking different plant data types.

With regards to links between the ELIXIR Biodiversity Community and other ELIXIR Communities, these are already foreseen, and a number of synergies have been clearly identified. Some examples can be found in [Table 2](#).

A global network of biodiversity projects and infrastructures

ELIXIR entered the European Strategy Forum for Research Infrastructure's (ESFRI) first roadmap in 2006 and reached its Landmark status in 2016 ([ELIXIR, 2021](#)). As a distributed research infrastructure, ELIXIR coordinates, integrates, and sustains bioinformatics resources across European countries and helps address the Grand Challenges across life sciences, from marine research, *via* plants and agriculture, to health research, medical sciences, and biodiversity informatics. ELIXIR provides services in seven scientific domains including "Evolution and phylogeny" and "Genes and genomes" (<https://elixir-europe.org/services>) that link the activities of the ELIXIR community to the wider landscape of life-science research infrastructures (RIs) and international projects. As RIs mature and FAIRness has become the standard to achieve interoperability between RIs, it is opportune to outline the global network of interrelated projects and infrastructures, in which ELIXIR operates to maximise synergy and to avoid redundancy.

The relationships between different aspects of biodiversity data are well captured by the biodiversity knowledge graph of Roderic Page ([Figure 1](#)). The key activities of ELIXIR are captured by the molecular domain; the biodiversity knowledge graph clearly indicates how molecular data are related to the wider spectrum of biodiversity data that are targeted by other RIs and projects. The ELIXIR Biodiversity Community benefits from connections to RIs and projects in the biodiversity domain, an overview of which can build on the landscape analyses of the ESFRI roadmaps of [ESFRI 2018](#) ([ESFRI, 2018](#)) and [2021](#) ([ESFRI, 2021](#)), the partners of the Alliance for Biodiversity Knowledge, and the research infrastructure contact zones analysis between 10 biodiversity infrastructures, including ELIXIR ([Smith et al., 2022](#)). Additional to the data types considered by Page ([Figure 1](#)), the contact zones analysis considers 'observations' and 'collections', or groups of specimens, as elements of the biodiversity data domain. This recognition of the variety of types of biodiversity data and the importance of integration has been key to the establishment of many RIs and research projects, for example: the [Alliance for Biodiversity Knowledge](#); [Biodiversity Genomics Europe](#); [Biodiversity Heritage Library](#); [Biodiversity Community Integrated Knowledge Library](#); [iBOL BIOSCAN](#); [Biodiversity Literature Repository](#); [Catalogue of Life](#); [Data Observation Network for Earth](#); [Distributed System of Scientific Collections](#); [Earth BioGenome Project](#); [European Marine Biological Resource Centre](#); [Environmental Research Infrastructures](#); [Encyclopedia of Life](#); [European Open Science Cloud](#); [European Reference Genome Atlas](#); [Europa Biodiversity Observation Network](#); [Global Biodiversity Information Facility](#); [Global Earth Observation System of Systems](#); [Global Soil Biodiversity Initiative](#); [International Barcode of Life](#); [iNaturalist](#); [LifeWatch ERIC](#); [Long-Term Ecosystem Research in Europe](#); [Microbial Resource Research Infrastructure](#); [National Ecological Observatory Network](#); [Open Traits Network](#); [Plazi](#); [Pôle national de données de biodiversité](#); [Swiss Institute for bioinformatics Literature Services](#); [Soil Biodiversity Observation Network](#); [Treatment-Bank](#); [World Register of Marine Species](#).

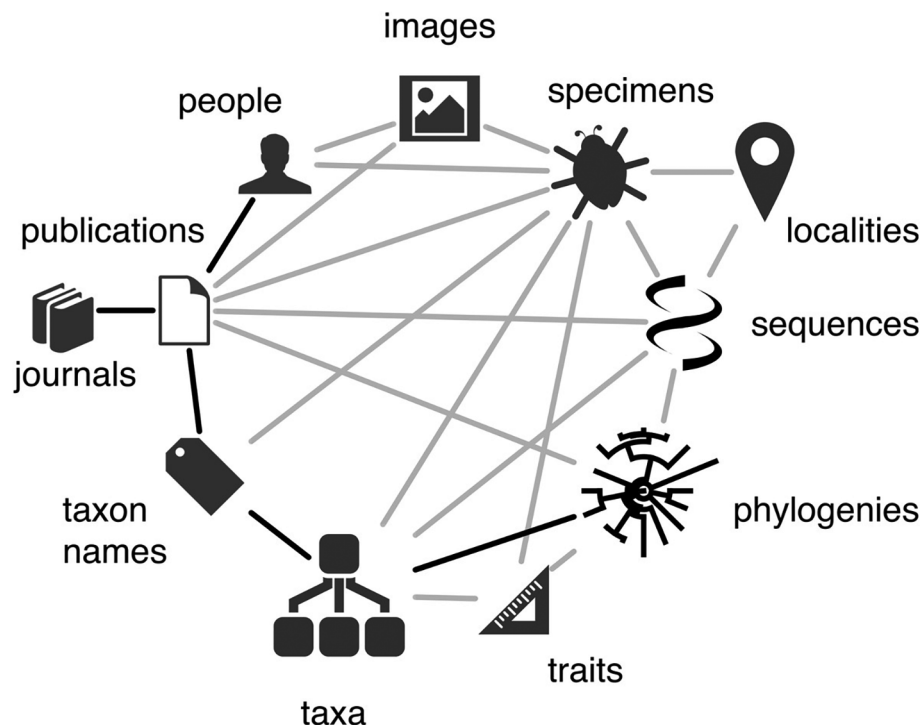


Figure 1. The biodiversity knowledge graph defined by Roderick D.M. Page (2013, 2016). Genomics data comprise one facet of the biodiversity knowledge graph, where questions and approaches in biodiversity research traverse the paths in this graph, and where all parts of the graph are constantly 'evolving' and growing. Wikimedia Commons [CC-BY-4.0](https://creativecommons.org/licenses/by/4.0/).

In addition to the above examples and in the context of ELIXIR the following two examples highlight ongoing activities in the field of biodiversity and in the context of the European research sphere.

Example: Biodiversity Community Integrated Knowledge Library (BiCIKL)

Several ELIXIR Nodes are involved in European projects with focus on biodiversity. The BiCIKL project is building the Biodiversity Knowledge Hub (BKH) - a single knowledge portal to interlinked machine-readable FAIR data - using unique stable identifiers on specimens, sequences, taxonomy and publications (Penev *et al.*, 2021). A set of core global biodiversity databases (GBIF, ENA, PlutoF, Plazi, DISSCO, OpenBioDiv, ToL, *etc.*) are contributing with the aim to develop services to augment the interlinking of biodiversity contents, starting with biotic interactions. The project is also financing competitive implementation studies to develop transnational resources.

Example: European Open Science Cloud (EOSC)

The European Open Science Cloud initiative (2023) intends to offer a federated and open multi-disciplinary environment where tools, data and services can be published, sought, and re-used. Via enabling seamless access and FAIR management EOSC aims to develop a Web of FAIR Data and services for science, innovation and education in Europe through which value-added services can be offered. The EOSC-Life initiative connects 13 life science 'ESFRI' research infrastructures to create an open, digital and collaborative space for biological and medical research. Among the EOSC-Life "FAIR" published data and catalogued services (by participating RIs), ones related to biodiversity are included. The workflow for marine Genomic Observatories data analysis is such an example (EBI, 2021).

Conclusions: A roadmap for the ELIXIR Biodiversity Community

Considering the context discussed above, the ELIXIR Biodiversity Community aims to contribute towards the global aim of tackling the biodiversity crisis by making possible a future where:

- Large-scale sustainable data production services are meeting the routine needs of hundreds of laboratories and thousands of citizen scientists for sequence-based biodiversity research and monitoring;

- A set of well-connected, stable and long-term infrastructures among which ELIXIR is supporting a growing portfolio of stakeholders by improving their access to, and integration of well-curated, high-quality, richly annotated and connected molecular data.
- State-of-the-art computational tools are available for large-scale projects related to biodiversity, including data standardisation initiatives.

To advance towards these ambitious goals, longer-term and within one year of the Community establishment, the ELIXIR Biodiversity Community proposes a roadmap. Table 3 shows five long-term objectives for the ELIXIR Biodiversity Community to address. The current focus is on the informatics, databases, and tools more than on the biological questions, so as the Community grows, it will be important to widen the diversity of its membership to ensure that the technical developments will serve the needs of biodiversity researchers.

Table 3. The ELIXIR Biodiversity Community long-term objectives.

Objectives	Approaches
Identify and support large-scale stakeholders in the biodiversity domain	<ul style="list-style-type: none"> • Support alignment of large-scale projects and transcending initiatives to result in high-quality, interoperable data and metadata • Build routes for the community to access and add to the knowledge (curation) of growing resources <i>e.g.</i>, trait measurements, observations beyond geolocations • Include primary production sectors affecting biodiversity
Connect and align biodiversity infrastructures	<ul style="list-style-type: none"> • Identify infrastructures contributing to the worldwide effort to sequence and catalogue biodiversity data • Leverage ELIXIR networks and Communities to facilitate linking between infrastructures through collaborative projects • Increase interoperability in biodiversity infrastructures through alignment of taxonomies and data/metadata standards • Include relevant citizen science infrastructures (<i>e.g.</i>, Atlas of Living Australia, iNaturalist, eBird)
Assist with policy decision making	<ul style="list-style-type: none"> • High-level alignment of strategy and policy in the biodiversity data domain • Support reconciliation of the interests of primary producers in biodiversity-rich environments
Aspire to achieve an ecosystem of production services for sequence-based biodiversity monitoring	<ul style="list-style-type: none"> • Identify gaps in the platforms/frameworks that exist to support the biodiversity data life cycle • Address all ELIXIR tools and services and where they can be plugged into the biodiversity data ecosystem • Coordinate and integrate services that support workflows through all stages of the process: from sampling, taxonomic identification and vouchering, sequence generation, annotation, cataloguing and further application of the data • Result in a network of services that meet the route needed by hundreds of labs and thousands of citizen scientists
Connect to and leverage the full potential of ELIXIR	<ul style="list-style-type: none"> • Establish the network of Nodes • Invest in training • Focus on community integration and re-use (rather than disjointed efforts) • Connect with other ongoing ELIXIR efforts

Data availability

Underlying data

No data are associated with this article.

Extended data

Figshare: Extended Data 1: Biodiversity RIs & Projects. <https://doi.org/10.6084/m9.figshare.22723432> (Waterhouse, 2023).

This project contains the following extended data:

- Extended_Data_1_Biodiversity_RIs_Projects.xlsx (A non-exhaustive list of biodiversity research infrastructures, collected as part of the development of the ELIXIR Biodiversity Community white paper 2022-2023.)

Data are available under the terms of the [Creative Commons Zero “No rights reserved” data waiver](#) (CC0 1.0 Public domain dedication).

Acknowledgements

The authors thank all members of the ELIXIR Biodiversity Community, the ELIXIR Hub, and the ELIXIR Heads of Nodes, for useful feedback during the development of this Community White Paper.

References

- Abarenkov K, *et al.*: **PlutoF—a Web Based Workbench for Ecological and Taxonomic Research, with an Online Implementation for Fungal ITS Sequences.** *Evol. Bioinforma.* 2010; **6**: EBO.S6271.
[Publisher Full Text](#)
- Balech B, Sandionigi A, Marzano M, *et al.*: **MetaCOXI: an integrated collection of metazoan mitochondrial cytochrome oxidase subunit-I DNA sequences.** *Database.* 2022:baab084. 2022.
[Publisher Full Text](#)
- BGE: **BGE - Biodiversity Genomics Europe.** 2023. (Accessed April 7, 2023).
[Reference Source](#)
- Bispo A, Willenz P, Hajdu E: **Diving into the unknown: fourteen new species of haplosclerid sponges (Demospongiae: Haplosclerida) revealed along the Peruvian coast (Southeastern Pacific).** *Zootaxa.* 2022; **5087**: 201–252.
[Publisher Full Text](#)
- Black M, *et al.*: **EDAM: the bioscientific data analysis ontology (update 2021).** 2022.
[Publisher Full Text](#)
- Blicharska M, *et al.*: **Biodiversity's contributions to sustainable development.** *Nat. Sustain.* 2019; **2**: 1083–1093.
[Publisher Full Text](#)
- Boekhout T, *et al.*: **Trends in yeast diversity discovery.** *Fungal Divers.* 2022; **114**: 491–537.
[Publisher Full Text](#)
- Capella-Gutierrez S, *et al.*: **Lessons Learned: Recommendations for Establishing Critical Periodic Scientific Benchmarking.** *Bioinformatics.* 2017.
[Publisher Full Text](#)
- Chimeno C, *et al.*: **Peering into the Darkness: DNA Barcoding Reveals Surprisingly High Diversity of Unknown Species of Diptera (Insecta) in Germany.** *Insects.* 2022; **13**: 82.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Costello MJ, May RM, Stork NE: **Can We Name Earth's Species Before They Go Extinct?** *Science.* 2013; **339**: 413–416.
[PubMed Abstract](#) | [Publisher Full Text](#)
- EBI: **A workflow for marine genomic data analysis.** 2021. (Accessed April 7, 2023).
[Reference Source](#)
- ELIXIR: **ELIXIR | ESFRI Roadmap 2021.** 2021. (Accessed April 7, 2023).
[Reference Source](#)
- EOSC Portal: **EOSC Portal.** 2023. (Accessed April 7, 2023).
[Reference Source](#)
- ERGA: **The European Reference Genome Atlas (ERGA) initiative.** erga. 2023. (Accessed April 7, 2023).
[Reference Source](#)
- ESFRI: **ESFRI Strategy Report and Roadmap 2018.** 2018. (Accessed April 7, 2023).
[Reference Source](#)
- ESFRI: **ESFRI Strategy Report on Research Infrastructures.** 2021. (Accessed April 7, 2023).
[Reference Source](#)
- FAO: **The state of the world's biodiversity for food and agriculture.** Rome: FAO Commission on Genetic Resources for Food and Agriculture; 2019.
- Formenti G, *et al.*: **The era of reference genomes in conservation genomics.** *Trends Ecol. Evol.* 2022; **37**: 197–202.
[Publisher Full Text](#)
- Frumkin H, Haines A: **Global Environmental Change and Noncommunicable Disease Risks.** *Annu. Rev. Public Health.* 2019; **40**: 261–282.
[Publisher Full Text](#)
- Gadelha LMR, *et al.*: **A survey of biodiversity informatics: Concepts, practices, and challenges.** *WIREs Data Min. Knowl. Discov.* 2021; **11**.
[Publisher Full Text](#)
- GBF: **Kunming-Montreal Global Biodiversity Framework.** 2023. (Accessed April 7, 2023).
[Reference Source](#)
- Goble C, *et al.*: **Implementing FAIR Digital Objects in the EOSC-Life Workflow Collaboratory.** 2021.
[Publisher Full Text](#)
- Harrow J, *et al.*: **ELIXIR-EXCELERATE: establishing Europe's data infrastructure for the life science research of the future.** *EMBO J.* 2021; **40**: e107409.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hermoso V, *et al.*: **The EU Biodiversity Strategy for 2030: Opportunities and challenges on the path towards biodiversity recovery.** *Environ. Sci. Policy.* 2022; **127**: 263–271.
[Publisher Full Text](#)
- Hoban S, *et al.*: **Genetic diversity targets and indicators in the CBD post-2020 Global Biodiversity Framework must be improved.** *Biol. Conserv.* 2020; **248**: 108654.
[Publisher Full Text](#)
- Hoban S, *et al.*: **Global Commitments to Conserving and Monitoring Genetic Diversity Are Now Necessary and Feasible.** *Bioscience.* 2021; **71**: 964–976.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hoborn D: **BIOSCAN: DNA barcoding to accelerate taxonomy and biogeography for conservation and sustainability** Adamowicz, S, editor. *Genome.* 2021; **64**: 161–164.
[PubMed Abstract](#) | [Publisher Full Text](#)

- IPBES: **Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services.** *Zenodo.* 2019.
[Publisher Full Text](#)
- IPCC: *Climate Change 2022: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge, UK and New York, NY, USA: Cambridge University Press; 2022.
[Reference Source](#)
- Ison J, et al.: **The bio.tools registry of software tools and data resources for the life sciences.** *Genome Biol.* 2019; **20**: 164.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Jetz W, et al.: **Essential biodiversity variables for mapping and monitoring species populations.** *Nat. Ecol. Evol.* 2019; **3**: 539–551.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Jiménez RC, et al.: **Four simple recommendations to encourage best practices in research software.** *F1000Res.* 2017; **6**: 876.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Johnson CN, et al.: **Biodiversity losses and conservation responses in the Anthropocene.** *Science.* 2017; **356**: 270–275.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Kissling WD, et al.: **Building essential biodiversity variables (EBVs) of species distribution and abundance at a global scale.** *Biol. Rev.* 2018; **93**: 600–625.
[Publisher Full Text](#)
- Lewin HA, et al.: **Earth BioGenome Project: Sequencing life for the future of life.** *Proc. Natl. Acad. Sci.* 2018; **115**: 4325–4333.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Lewin HA, et al.: **The Earth BioGenome Project 2020: Starting the clock.** *Proc. Natl. Acad. Sci.* 2022; **119**: e2115635118.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Marselle MR, et al.: **Pathways linking biodiversity to human health: A conceptual framework.** *Environ. Int.* 2021; **150**: 106420.
[Publisher Full Text](#)
- Mora C, Tittensor DP, Adl S, et al.: **How Many Species Are There on Earth and in the Ocean?** Mace, GM, editor. *PLoS Biol.* 2011; **9**: e1001127.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Moura MR, Jetz W: **Shortfalls and opportunities in terrestrial vertebrate species discovery.** *Nat. Ecol. Evol.* 2021; **5**: 631–639.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Nature.: **Biodiversity faces its make-or-break year, and research will be key.** *Nature.* 2022; **601**: 298–298.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Nic Lughadha E, et al.: **Extinction risk and threats to plants and fungi.** *PLANTS PEOPLE PLANET.* 2020; **2**: 389–408.
[Publisher Full Text](#)
- van Ommen KE, et al.: **ARISE: Building an infrastructure for species recognition and biodiversity monitoring in the Netherlands.** *Biodivers. Inf. Sci. Stand.* 2022; **6**: e93613.
[Publisher Full Text](#)
- Page R: **Towards a biodiversity knowledge graph.** *Res. Ideas Outcomes.* 2016; **2**: e8767.
[Publisher Full Text](#)
- Page RDM: **BioNames: linking taxonomy, texts, and trees.** *PeerJ.* 2013; **1**: e190.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Penev L, et al.: **Towards Interlinked FAIR Biodiversity Knowledge: The BiCikl perspective.** *Biodivers. Inf. Sci. Stand.* 2021; **5**: e74233.
[Publisher Full Text](#)
- Rahman MM, Burian A, Creedy TJ, et al.: **DNA-based assessment of environmental degradation in an unknown fauna: The freshwater macroinvertebrates of the Indo-Burmese hotspot.** *J. Appl. Ecol.* 2022; **59**: 1644–1658.
[Publisher Full Text](#)
- Robinson WD, Peres CA: **Editorial: Benchmarking Biodiversity in an Era of Rapid Change.** *Front. Ecol. Evol.* 2021; **9**: 810287.
[Publisher Full Text](#)
- Segelbacher G, et al.: **New developments in the field of genomic technologies and their relevance to conservation management.** *Conserv. Genet.* 2022; **23**: 217–242.
[Publisher Full Text](#)
- Smith V, et al.: **Research Infrastructure Contact Zones: a framework and dataset to characterise the activities of major biodiversity informatics initiatives.** *Biodiversity Data Journal.* 2022; **10**.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Theissing K, et al.: **How genomics can help biodiversity conservation.** *Trends Genet.* 2023. S0168952523000203.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Turvey ST, Cress JJ: **Extinction in the Anthropocene.** *Curr. Biol.* 2019; **29**: R982–R986.
[PubMed Abstract](#) | [Publisher Full Text](#)
- UN: **THE 17 GOALS | Sustainable Development.** 2015. (Accessed April 7, 2023).
[Reference Source](#)
- da Veiga Leprevost F, et al.: **BioContainers: an open-source and community-driven framework for software standardization** Valencia, A, editor. *Bioinformatics.* 2017; **33**: 2580–2582.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Waterhouse RM, et al.: **Recommendations for connecting molecular sequence and biodiversity research infrastructures through ELIXIR.** *F1000Res.* 2022; **10**: 1238.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Waterhouse RM: **Extended Data 1: Biodiversity RIs & Projects.** [Dataset]. *figshare.* 2023.
[Publisher Full Text](#)
- Wezel A, et al.: **Agroecological principles and elements and their implications for transitioning to sustainable food systems. A review.** *Agron. Sustain. Dev.* 2020; **40**: 40.
[Publisher Full Text](#)
- WWF: *The Living Planet Report 2022 – Building a nature-positive society.* Almond REA, Grooten M, Juffe Bignoli D, et al., editors. WWF; [Reference Source](#) 2022.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research