

Connecting biodiversity biobanks in the global infrastructure landscape

Gabi Droege^{1,*}, Jonas J. Astrin², Anton Güntsch¹, Andrew Bentley³, Carolina Corrales², Kim Labuschagne⁴, Gilberto Ocampo⁵, Wouter Addink⁶, Hannah J. Appiah-Madson⁷, Masanori Arita⁸, Heather Avrili⁹, Katharine Barker¹⁰, Gustavo A. Bravo¹¹, Elena Buzan^{12,13}, Ana Casino¹⁴, Guy Cochrane¹⁵, Manuela da Silva¹⁶, Dan Distel⁷, Marcus Ernst¹, Belen Escobari Vargas¹, Rosalia Falco Poulin⁷, Kasia Fantoni⁶, David Fichtmueller¹, Frank Oliver Glöckner^{17,18}, Peter Grobe², Quentin Groom¹⁹, Vikas Gupta¹⁵, Eva Häffner¹, Katherine D. Heinemann²⁰, Jörg Holetschek¹, Peter M. Hollingsworth²¹, Chris Hunter²², Christina Hvilsom²³, Sharif Islam⁶, Vishnukumar Balavenkataraman Kadhivelu¹⁵, D. İpek Kurtböke²⁴, Harris Lewin²⁵, Alice C. McHardy²⁶, Anna Mekarska²⁷, Rachel S. Meyer²⁸, Melania Muñoz-García²⁹, Carmen Paulmann²⁶, Joana Paupério¹⁵, Lyubomir Penev³⁰, Sujeevan Ratnasingham³¹, Dominik Röpert¹, Santiago R. Ron³², Oliver Ryder^{33,34}, Dmitry Schigel³⁵, Amber Hartman Scholz²⁹, Lynn M. Schriml³⁶, Boripat Siriaronrat^{33,37}, Laura Tilley¹⁴, Robert M. Waterhouse³⁸, Bernhard Misof² and Thomas Borsch¹

¹Botanic Garden and Botanical Museum Berlin, Freie Universität Berlin, Königin-Luise-Str. 6-8, 14195, Berlin, Germany

²Leibniz Institute for the Analysis of Biodiversity Change, Museum Koenig, Adenauerallee 127, 53113 Bonn, Germany

³Biodiversity Institute, University of Kansas, 1345 Jayhawk Boulevard, Lawrence, KS 66045, United States

⁴Foundational Biodiversity Science, South African National Biodiversity Institute (SANBI), 232 Boom Street, Pretoria 0001, South Africa

⁵Departamento de Biología, Centro de Ciencias Básicas, Universidad Autónoma de Aguascalientes, Aguascalientes 20100, Mexico

⁶Naturalis Biodiversity Center, Darwinweg 2, 2333 CR Leiden, The Netherlands

⁷Ocean Genome Legacy Center, Marine Science Center, Northeastern University, 430 Nahant Road, Nahant, MA 01908, United States

⁸National Institute of Genetics, 411-8540 Shizuoka, Japan

⁹Natural History Museum, Cromwell Road, London SW7 5BD, United Kingdom

¹⁰National Museum of Natural History, Smithsonian Institution, 10th and Constitution Ave. NW, DC 20013, United States

¹¹Centro de Colecciones y Gestión de Especies, Instituto Alexander von Humboldt, Bogotá, Colombia

¹²Faculty of Mathematics, Natural Sciences and Information Technologies, University of Primorska, Glagoljaška 8, 6000 Koper, Slovenia

¹³Faculty of Environmental Protection, Trg mladosti 7, 3320 Velenje, Slovenia

¹⁴Consortium of European Taxonomic Facilities (CETAF), Rue Vautier 29, 1000 Brussels, Belgium

¹⁵European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridgeshire CB10 1SD, United Kingdom

¹⁶Oswaldo Cruz Foundation (Fiocruz), Fiocruz Biobanks and Biological Collections Coordination, Rio de Janeiro, Brazil

¹⁷MARUM - Center for Marine Environmental Sciences, University of Bremen, Leobener Str. 8, 28359 Bremen-Horn-Lehe, Germany

¹⁸Alfred Wegener Institute - Helmholtz Centre for Polar- and Marine Research, Am Handelshafen 12, 27570 Bremerhaven, DE

¹⁹Meise Botanic Garden, Nieuwelaan 38, 1860 Meise, Belgium

²⁰IUCN SSC Center for Species Survival Biodiversity Banking, San Diego Zoo Wildlife Alliance, Conservation Science, 15600 San Pasqual Valley Rd., Escondido, CA 92027, United States

²¹Royal Botanic Garden Edinburgh, Edinburgh EH3 5LR, United Kingdom

²²Freelance Biocurator, Exmouth, United Kingdom

Received: 31 March 2026. **Revised:** 19 August 2026. **Accepted:** 20 August 2026

© The Author(s) 2026. Published by Oxford University Press on behalf of the American Institute of Biological Sciences. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

²³Centre for Research and Conservation, Royal Zoological Society of Antwerp, Koningin Astridplein 20-26, 2018 Antwerpen, Belgium

²⁴School of Science, Technology and Engineering, University of the Sunshine Coast, Maroochydore DC QLD 4558, Australia

²⁵Julie Ann Wrigley Global Futures Laboratory, Arizona State University, PO Box 872404, Tempe, AZ 85287, United States

²⁶Helmholtz Centre for Infection Research, Inhoffenstr. 7, 38124 Braunschweig, Germany

²⁷European Association of Zoos and Aquaria, Plantage Middenlaan 45, 1018DC Amsterdam, The Netherlands

²⁸Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, CA 95064, United States

²⁹Leibniz Institute DSMZ - German Collection of Microorganisms and Cell Cultures GmbH, Inhoffenstraße 7B, 38124 Braunschweig, Germany

³⁰Pensoft Publishers, Prof. Georgi Zlatarski Str, 12, 1700 Sofia, Bulgaria

³¹Centre for Biodiversity Genomics, University of Guelph, Guelph, Ontario N1G2W1, Canada

³²Museo de Zoología, Centro de Investigaciones de la Biodiversidad CIBIO, Facultad de Ciencias Exactas, Naturales y Ambientales, Pontificia Universidad Católica del Ecuador, Av. 12 de Octubre y Roca, Aptdo. 17-01-2184 Quito, Ecuador

³³IUCN SSC Animal Biobanking for Conservation Specialist Group, Gland, Switzerland

³⁴San Diego Zoo Wildlife Alliance, Escondido, CA 92027, United States

³⁵Global Biodiversity Information Facility (GBIF) Secretariat, Universitetsparken 15, DK-2100 Copenhagen Ø, Denmark

³⁶School of Medicine, Institute for Genome Sciences, University of Maryland, Baltimore, MD 21201, United States

³⁷Faculty of Environment and Resource Studies, Mahidol University, Nakhon Pathom 73170, Thailand

³⁸SIB Swiss Institute of Bioinformatics, Batiment Amphipole, UNIL-Sorge, 1015 Lausanne, Switzerland

*Corresponding author: g.droege@bo.berlin

Gabi Droege (g.droege@bo.berlin) is the head of GGBN's Technical Management Office as well as the head of the Collection Data Networking working group at the Center for Biodiversity Informatics and Collection Data Integration at Botanic Garden and Botanical Museum Berlin, Freie Universität Berlin in Berlin, Germany. Jonas Astrin (j.astrin@leibniz-lib.de) is the head of GGBN's Secretariat Office and Biobank Curator/section head at the LIB, Leibniz Institute for the Analysis of Biodiversity Change, Bonn, Germany. Anton Güntsch (a.guentsch@bo.berlin) is the director of the Center for Biodiversity Informatics and Collection Data Integration at Botanic Garden and Botanical Museum Berlin, Freie Universität Berlin in Berlin, Germany. Andrew Bentley (abentley@ku.edu) is a collections specialist and bioinformatician at the Biodiversity Institute, University of Kansas (KU), and a member of the Biodiversity Collections Network (BCoN) and International Partners for the Digital Extended Specimen (IPDES) steering committees. He also teaches in the KU Museum Studies Program. Carolina Corrales (c.corrales@leibniz-lib.de) is the program manager of the Global Genome Biodiversity Network and is based at the Leibniz Institute for the Analysis of Biodiversity Change, Bonn, Germany. Kim Labuschagne (k.labuschagne@sanbi.org.za) is the Biobank Curator at the South African National Biodiversity Institute (SANBI) Wildlife Biobank, Pretoria, South Africa. Gilberto Ocampo (gilberto.ocampo@edu.uaa.mx) is professor of botany, Department of Biology, Universidad Autónoma de Aguascalientes, Mexico. He specializes in botany, systematics, phylogenetics, and Floristics. He is head of the HUAA herbarium and the University's DNA Bank. He is currently vice-chair of the GGBN executive committee. Wouter Addink (wouter.addink@naturalis.nl) is coordinator of international e-infrastructures and data at Naturalis Biodiversity Center and Deputy Director for Data and Technology in DiSSCo (Distributed System of Scientific Collections). His main interests are in biodiversity informatics and FAIR, AI ready specimen data. Hannah J. Appiah-Madson (h.appiah-madson@northeastern.edu) is the Collections Manager for the Ocean Genome Legacy Genomic Resource Collection at Northeastern University, Nahant, Massachusetts, United States. Masanori Arita (arita@nig.ac.jp) is head of the DNA Data Bank of Japan and steering member of the International Nucleotide Sequence Database Collaboration (INSDC). He also works on metabolomics as a team director at the RIKEN Center for Sustainable Resource Science, Japan. Heather Avriil (h.avriil@nhm.ac.uk) is the Molecular Collections Manager at the Natural History Museum in London. Katherine Barker (BarkerK@si.edu) is affiliated to the National Museum of Natural History, Smithsonian Institution, Washington DC, USA. Gustavo A. Bravo (gbravo@humboldt.org.co) is the curator of Ornithological Collections at the Instituto Humboldt in Colombia and a current member of the GGBN Executive Committee. His research uses biological specimens as sources of phenotypic, genomic, ecological, and environmental data to address questions in evolution, ecology, and biodiversity conservation. Elena Buzan (elena.buzan@upr.si) is a vice dean and head of the Molecular Ecology Group at the Faculty of Mathematics, Natural Sciences and Information Technologies, University of Primorska, Slovenia. She is also a council member and former Partnership Officer of the European Reference Genome Atlas (ERGA) community. Ana Casino (ana.casino@cetaf.org) is the Executive Director of the Consortium of European Taxonomic Facilities (CETAF), the organization gathering more than 80 scientific collections-based research institutions across 23 European countries, working on biodiversity data. She is co-author of the CETAF Code of Conduct on Access and Benefit Sharing, as well as representative at the ABS Forum led by the European Commission. Guy Cochrane (cochrane@ebi.ac.uk) is the Scientific Head of the Global Biodata Coalition (GBC) and a faculty member of the European Molecular Biology Laboratory at the organization's European Bioinformatics Institute (EMBL-EBI), Hinxton, United Kingdom. Manuela da Silva (manuela.dasilva@fiocruz.br) is the General Coordinator of Biobanks and Biological Collections at Fiocruz and she represents academia in the Genetic Heritage Management Council (CGen/MMA), the Brazilian National Competent Authority for Access and Benefit-Sharing (ABS). Daniel Distel (d.distel@northeastern.edu) is a research professor in the Department of Marine and Environmental Science and the director of the Ocean Genome Legacy Center, a marine dedicated genome resource repository, at Northeastern University. Marcus Ernst (m.ernst@bo.berlin) is a researcher at the Center for Biodiversity Informatics and Collection Data Integration at Botanic Garden and Botanical Museum Berlin, Freie Universität Berlin in Berlin, Germany. Belen Escobari (b.escobari@bo.berlin) is the DNA Bank curator at Botanic Garden and Botanical Museum Berlin, Freie Universität Berlin in Berlin, Germany. Rosalia Falco Poulin (r.falco@northeastern.edu) is the lab manager for the Ocean Genome Legacy Genomic Resource Collection at Northeastern University, Nahant, Massachusetts, United States. Kasia Fantoni (kasia.fantoni@naturalis.nl) is the Community Manager of iBOL Europe and is based at Naturalis Biodiversity Center in Leiden, the Netherlands. David Fichtmueller (d.fichtmueller@bo.berlin) is a researcher at the Center for Biodiversity Informatics and Collection Data Integration at Botanic Garden and Botanical Museum Berlin, Freie Universität Berlin in Berlin, Germany. He is working on research data management and biodiversity data standardization. Frank Oliver Glöckner (frank.oliver.gloeckner@awi.de) is Professor for Earth System Data Science at the Department of Geosciences at the University of Bremen (MARUM). He is head of Data at the Computing and Data centre of the Alfred Wegener Institute Alfred Wegener Institute - Helmholtz Centre for Polar- and Marine Research in Bremerhaven. He is head of the information system PANGAEA at MARUM and AWI. He is speaker of the National Research Data Infrastructure consortium NFDI4Biodiversity. Peter Grobe (p.grobe@leibniz-lib.de) is Principal Investigator at the Centre for Biodiversity Knowledge Science, Leibniz Institute for the Analysis of Biodiversity Change, Bonn, Germany. Quentin Groom (quentin.groom@plantentuinmeise.be) is a biodiversity scientist working at Meise Botanic Garden, Belgium, on the intersection of taxonomy, ecology, conservation, and informatics, with a focus on turning scientific knowledge into practical tools for understanding and protecting biodiversity. Vikas Gupta (vikasg@ebi.ac.uk) is a software engineer at the European Molecular Biology Laboratory—European Bioinformatics Institute (EMBL-EBI), United Kingdom. He specializes in scientific data platforms, data visualization, bioinformatics applications, cloud-native architectures, and scalable data engineering solutions. Eva Häffner (e.haefner@bo.berlin) works as science policy coordinator at Botanic Garden and Botanical Museum Berlin, Freie Universität Berlin in Berlin, Germany. Katherine D. Heineman (kheineman@sdzwa.org) is the Conservation Program Leader for Biodiversity Banking at San Diego Zoo Wildlife Alliance. She manages the IUCN SSC Center for Species Survival Biodiversity Banking, which strives to build technical capacity in biobanking in biodiversity hotspots through training and knowledge sharing. Jörg Holetschek (j.holetschek@bo.berlin) is the biodiversity data networks coordinator at the Center for Biodiversity Informatics and Collection Data Integration at Botanic Garden and Botanical Museum Berlin, Freie Universität Berlin in Berlin, Germany. Peter M. Hollingsworth (phollingsworth@rbge.org.uk) is Director of Science and Deputy Keeper at the Royal Botanic Garden of Edinburgh. He is also the chief scientific officer of the International Barcode of Life (iBOL) Consortium, and a member of the iBOL Europe Coordination Team. Chris Hunter (dr.chris.i.hunter@gmail.com) is a metadata professional with 20+ years of experience spanning data curation, research data infrastructure, product ownership, and stakeholder engagement. Proven track record of leading cross-functional teams to deliver data platforms and services aligned with FAIR principles and open data standards. Christina Hvilsom (hvilsomgenes@gmail.com) is the founder and chair of the EAZA Biobank, leading one of the world's largest coordinated wildlife biobanking initiatives, connecting more than 450 zoos and aquaria. Her work spans the full pathway from biospecimens to conservation genomics and policy, transforming biological samples into actionable knowledge for species conservation. Sharif Islam (sharif.islam@naturalis.nl) is a senior data architect at Naturalis Biodiversity Center in Leiden, The Netherlands, where he works on biodiversity informatics, FAIR data, and architecture of sustainable research infrastructures. Vishnukumar Balavenkataraman Kadhirvelu (kadhirvelu@ebi.ac.uk) is a senior software engineer at the European Nucleotide Archive, European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, United Kingdom. D. İpek Kurtböke (IKurtbok@usc.edu.au) is affiliated with the University of the Sunshine Coast, School of Science, Technology and Engineering, Maroochydore DC, Australia. Harris Lewin (harris.lewin@asu.edu) is a research professor at Arizona State University and Distinguished Professor Emeritus of Evolution and Ecology at the University of California, Davis. He is co-founder of the Earth BioGenome Project and served as its Chair from 2018 to 2026. His research interests are in comparative genomics

and chromosome evolution. Alice C. McHardy (alice.mchardy@helmholtz-hzi.de) is head of the Department of Computational Biology for Infection Research at the Helmholtz Centre for Infection Research and a full professor at TU Braunschweig in Germany. Anna Mekarska (anna.mekarska@eaza.net) is the Coordinator of the European Association of Zoos and Aquaria (EAZA) Biobank, one of the world's largest coordinated wildlife biobanking initiatives, connecting more than 450 zoos and aquaria. She is managing the biobank operations, samples storage and release process, as well as data maintenance. Rachel Meyer (rameyer@ucsc.edu) is affiliated to the Dept. of Ecology and Evolutionary Biology, University of California, Santa Cruz, United States. Melania Muñoz-García (melania.munoz@dsmz.de) is the German Nagoya Protocol Hub project officer, Science Policy and Internationalisation Department, Leibniz Institute DSMZ - German Collection of Microorganisms and Cell Cultures. Carmen Paulmann (carmen.paulmann@helmholtz-hzi.de) is Project Manager for the NFDI4Microbiota consortium and located at the Helmholtz Centre for Infection Research (HZI), Braunschweig, Germany. Joana Paupério (joanap@ebi.ac.uk) is the biodiversity project manager at the European Nucleotide Archive, European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, United Kingdom. Lyubomir Penev (l.penev@pensoft.net) is a professor in ecology at the Bulgarian Academy of Sciences, Sofia and also a founder and Chief Executive Officer (CEO) of Pensoft Publishers. His main interests are in biodiversity informatics, semantic publishing and re-use of biodiversity data. Sujeevan Ratnasingham (sratnasi@uoguelph.ca) is affiliated to the Centre for Biodiversity Genomics, University of Guelph, Guelph, Canada. Dominik Röpert (d.roepert@bo.berlin) is the coordinator for internal information systems at the Center for Biodiversity Informatics and Collection Data Integration at Botanic Garden and Botanical Museum Berlin, Freie Universität Berlin in Berlin, Germany. Santiago R. Ron (santiago.r.ron@gmail.com) is Director of Centro de Investigaciones de la Biodiversidad at Pontificia Universidad Católica del Ecuador. He is also chair of the GGBN Executive Committee and Principal Professor at the School of Biological Sciences, Pontificia Universidad Católica del Ecuador. Oliver Ryder (oryder@sdzwa.org) guides the strategic development of efforts to develop and expand a global network of cryobanking facilities, especially for viable tissue culture cells as Co-chair of the Animal Biobanking for Conservation Specialist Group of SSC-IUCN. He is also an adjunct professor in the Department of Evolution, Behavior and Ecology, University of California, San Diego. Dmitry Schigel (dschigel@gbif.org) is lead scientific officer at the Secretariat of Global Biodiversity Information Facility, working with the science team to broaden scientific relevance and impact of world's biodiversity data for researchers, including through data cross-linking and discoverability of DNA-derived, sensor-based, and systematic survey data. Amber H. Scholz (amber.h.scholz@dsmz.de) is the head of the Science Policy and Internationalisation Department at the Leibniz Institute DSMZ - German Collection of Microorganisms and Cell Cultures. Lynn Schriml (lynn.schriml@gmail.com) is a professor at the University of Maryland School of Medicine, Institute for Genome Sciences. Dr. Schriml, as President of the Genomic Standards Consortium, leads coordination of minimal information standards (MInS) development for any genomic study, across environments and technologies. Boripat Siriaronrat (boripat.siriaronrat@mahidol.ac.th) is affiliated with (1) IUCN SSC Animal Biobanking for Conservation Specialist Group, Gland, Switzerland; (2) Faculty of Environment and Resource Studies, Mahidol University, Nakhon Pathom, Thailand. Laura Tilley (lauratillye.ljt@gmail.com) is the Projects Coordinator at the Consortium of European Taxonomic Facilities (CETAF), specifically involved in tasks where data collection, management and exploitation is required. She is the chair of the Earth Science Working Group of CETAF focused on geological collections. Robert M. Waterhouse (robert.waterhouse@sib.swiss) is director of the Environmental Bioinformatics group at the SIB Swiss Institute of Bioinformatics and a leader of the European Reference Genome Atlas (ERGA) community (former vice-chair, current chair). Bernhard Misof (b.misof@leibniz-lib.de) is the Director General of the Leibniz Institute for the Analysis of Biodiversity Change (Bonn & Hamburg), Germany, which hosts the GGBN Secretariat Office. Thomas Borsch (t.borsch@bo.berlin) is director of the Botanic Garden and Botanical Museum Berlin, Freie Universität Berlin, Germany and professor of plant systematics and plant geography. His research is on integrative taxonomy and phylogenomics, and the use of networked biological collections in the knowledge generation process.

Abstract

The Global Genome Biodiversity Network (GGBN) is an international consortium of biodiversity biobanks and collections that strives to standardize and exchange best practices, and to share information on molecular biodiversity resources held in repositories around the globe. GGBN's goal is to bridge the gap between biological material stored in these biobanks and the broader research community by establishing a searchable index of the world's genomic resources, openly available to the scientific community and all stakeholders. Here, we provide examples of how GGBN interacts synergistically with other biodiversity programs and initiatives to (1) connect researchers with physical materials linked to rich, highly structured contextual metadata and (2) strengthen the biobanking community.

Keywords GGBN, biodiversity biobanks, stable identifiers, data infrastructure, genetic resources.

Background

The global biodiversity data infrastructure landscape has evolved into a complex network of data, services, research, and collection infrastructures that support the documentation, sharing, analysis, and application of biodiversity knowledge worldwide. Each infrastructure specializes in certain data types, including specific biobank data. The growing demand for seamless data exchange relies on shared stable identifiers, metadata standards, and interoperability frameworks (Bingham et al. 2017, Schulman et al. 2021). Together, the connections between different data types, as well as between data and physical collection material, create the potential for a more comprehensive and multidimensional understanding of biodiversity. They also enable the creation of more robust, policy-relevant indicators for tracking biodiversity change, supporting global frameworks such as the Kunming-Montreal Global Biodiversity Framework (Convention on Biological Diversity 2022). Collaborations, including the Global Biodata Coalition (Hoborn et al. 2019, Global Biodata Coalition 2025) and the Alliance for Biodiversity Knowledge, have been formed to address these aspects and associated challenges.

Data infrastructures are part of a global, integrated, and open biodiversity data ecosystem aiming to make biodiversity and genomic data Findable, Accessible, Interoperable, and Reusable (FAIR) (Penev et al. 2024), where specimen data, DNA sequence data, metadata standards, and biodiversity knowledge are fully connected and openly accessible for research, conservation, and innovation (Letsch et al. 2025). Within this data ecosystem, the Global Genome Biodiversity

Network (GGBN) provides an information infrastructure for physical biological materials (e.g., DNA samples, tissues, environmental samples, cells, cultures, specimens, and others) used to generate molecular data or allow specific applications. Various other data infrastructures either serve as primary storage, such as the International Nucleotide Sequence Database Collaboration (INSDC) and the Barcode of Life Data Systems (BOLD) or as aggregators, such the Global Biodiversity Information Facility (GBIF) or the Distributed System of Scientific Collections (DiSSCo). Others focus on research data workflows and data life cycles, such as the National Research Data Infrastructure (NFDI) or the Biodiversity Community Integrated Knowledge Library (BiCIKL). To link and share data, they either develop standards (e.g., the Genomic Standards Consortium [GSC], DiSSCo, or INSDC) or adopt them, such as GBIF, BiCIKL, or NFDI. Together with GGBN, they form a robust and reliable global biodiversity data knowledge graph allowing complex linkages between biological material and data (see box 1), for example, the information that a DNA sample is derived from a tissue sample. In addition, a tissue sample is often related to a voucher specimen whose data can be found in GGBN, DiSSCo, and GBIF. Furthermore, sequences in INSDC or BOLD, as well as publications where the physical material has been used, can be linked to both the samples and the voucher specimen.

The key players in the overall global infrastructure landscape are described in more detail in box 2. Their connections with biodiversity biobanks and GGBN are shown in figure 1. An alphabetic list of abbreviations can be found at the very end.

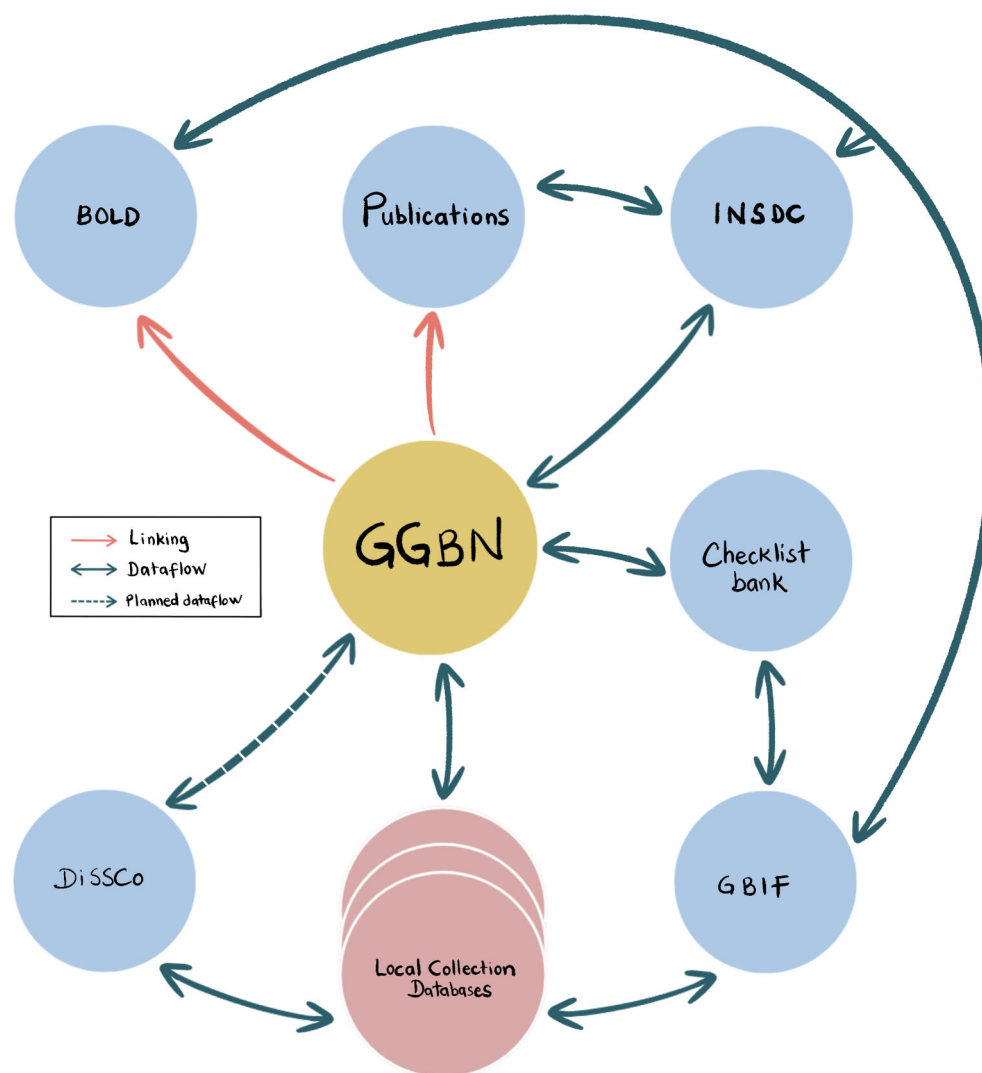


Figure 1 GGBN in the global landscape of biodiversity infrastructures. Local collection databases provide data to platforms such as GGBN, GBIF, and DiSSCo. Established or planned partnerships exist between GGBN and ENA (European Nucleotide Archive as part of the INSDC), DiSSCo, and the local collection databases (GGBN members). Links to publications and BOLD (Barcode of Life Data System) are currently only unidirectional. Like GBIF, GGBN uses certain checklists from Checklistbank as its taxonomic backbone. For example, the National Center for Biotechnology Information (NCBI) taxonomy (as part of INSDC) is available via Checklistbank. Furthermore, both BOLD and INSDC supply occurrence records to GBIF and INSDC, enabling data integration across these platforms. INSDC also maintains bidirectional links with publications, ensuring that sequence data and associated research outputs are interconnected and easily accessible.

Box 1 Benefits of collaborations between data infrastructures

Collaborations between biodiversity infrastructures such as GGBN, DiSSCo, NFDI, GBIF, INSDC, BOLD, GSC, and BiCIKL foster

- Seamless **interoperability** among different infrastructures
- Improved provenance, reproducibility, and compliance with **FAIR** (Findable, Accessible, Interoperable, and Reusable; Wilkinson et al. 2016) and **CARE** principles (Collective Benefit, Authority of Control, Responsibility, and Ethics; Carroll et al. 2020)
- Integration of **persistent identifiers** (PIDs) into workflows
- Harmonized **APIs** (Application Programming Interface), **data standards**, and identifiers
- Development of cloud-based, **cross-domain** analysis tools

Box 2 Data infrastructures and standards

The **International Nucleotide Sequence Database Collaboration (INSDC)** is a global collaboration of databases that capture, integrate, and archive nucleotide sequence information and associated annotations, building a comprehensive collection of data and enabling broad data sharing and reuse. INSDC members, DNA Data Bank of Japan (DDBJ) at the National Institute of Genetics, National Library of Medicine, National Center for Biotechnology Information (NLM-NCBI), and the European Nucleotide Archive (ENA) at the European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), provide the data resources (sequence data and annotations, sample, project and experimental metadata) and several tools and services that support management and sharing of information. The ENA Cross-Reference Search service is a powerful tool to enable links to external resources, such as GGBN, currently providing over three million

GGBN-ENA crossrefs. By using the relationships between biological material records stored in GGBN, not only can the direct physical resource be cited, but also all derivatives, for example, DNA, tissue, and specimen.

The **Barcode of Life Data System (BOLD)** is a global informatics platform supporting DNA barcoding by providing infrastructure for specimen registration, sequence data management, validation, and identification services (Ratnasingham et al. 2024). Since its launch in 2005, BOLD has become a critical resource for biodiversity genomics, linking specimen records, sequence data, and collection metadata through a structured, self-service system. With more than 21 million specimen records representing over one million species, BOLD supports applied monitoring projects and large-scale discovery of species. Its data model and workflows ensure that DNA barcode records are tied to voucher specimens, images, geospatial data, and taxonomic information, maximizing data utility and reproducibility. BOLD is unique in supporting the full research data lifecycle from pre-publication data capture, secure sharing with research partners and networks, and validation to deposition in INSDC databases.

The **Genomic Standards Consortium (GSC)** is a global community focused on making genomic data discoverable, reusable, and interoperable through the development of community-driven genomic metadata standards as methods of capturing and exchanging metadata (Eloe-Fadrosh et al. 2024). Established in 2005, the GSC coordinates the development of MiXS (Yilmaz et al. 2011), genomic checklists, and environmental extensions. The GSC's Compliance and Interoperability Working Group (CIG) promotes alignment among major biodiversity data standards by supporting the reuse of MiXS terms and fostering interoperability with frameworks such as the GGBN Data Standard and Darwin Core. In 2016, the GSC backed the "GGBN Data Standard Specification" project (Droege et al. 2016), which aimed to harmonize MiXS and GGBN attributes. Collaborative efforts are further advanced through the Global Biodiversity Working Group (GBWG), a joint interest group of TDWG (Biodiversity Information Standards) and the GSC, which coordinates cross-community standardization activities. The 2022 Memorandum of Understanding between TDWG and the GSC formalized the joint commitment to sustainable convergence of metadata standards, including alignment between Dwc and MiXS.

The **Distributed System of Scientific Collections (DiSSCo)** is a pan-European research infrastructure designed to digitize, standardize, and integrate natural science collections held in museums, herbaria, and related institutions. Its central objective is to create digital representations of specimens that are accessible through persistent identifiers and interoperable metadata standards (Addink and Theocharides 2024). By linking these digital specimens across infrastructures, DiSSCo enables bidirectional integration with external data sources such as GBIF, BOLD, ENA, and GGBN, thereby contributing to a machine-actionable biodiversity knowledge graph.

The **Global Biodiversity Information Facility (GBIF)** is an international network and data infrastructure that provides anyone, anywhere, free and open access to data about all types of life on Earth. GBIF.org and its related services currently deliver more than three billion occurrence records mobilized by thousands of data publishing institutions with the support of GBIF's distributed node network. GBIF is a Global Core Biodata Resource. Users can access GBIF-mediated data by filtering and downloading records directly through GBIF.org as well as through the robust GBIF API and through R and Python libraries. Digital Object Identifier (DOI)—based data citations enable

users to maintain the provenance and transparency of the data they access and use, allocating credit to data sources used in published research and policy.

The **National Research Data Infrastructure (NFDI)** is a German initiative, coordinated by 26 consortia that are funded by the DFG (German Research Foundation), to systematically manage and network research data across disciplines. The NFDI promotes open, FAIR, and sustainable research by developing community-driven standards, procedures, and cross-disciplinary metadata frameworks. Within this, NFDI4Biodiversity and NFDI4Microbiota each bring together around 50 partner institutions, including research centers, museums, and natural history societies, to mobilize biodiversity, environmental, microbiological, and microbiome-related data for research and conservation. They provide access to high-quality standardized data, tools, and services for data archiving, publishing, searching, and analysis, as well as metadata frameworks, curated databases, and comprehensive training opportunities required for cross-institutional data use. It is driven by the understanding that stakeholders in science, health, agriculture, policy, nature conservation, and landscape management require reliable data to make informed contributions to biodiversity conservation and health. The consortium collaborates with major European and international initiatives, including the European Life Sciences Infrastructure for Biological Information (ELIXIR), the European Open Science Cloud (EOSC), EMBL-EBI, and GSC. Notably, through the NFDI4Microbiota project, the German Collection of Microorganisms and Cell Cultures (DSMZ) has further developed the StrainInfo database (Lissin et al. 2025). Originally developed by the Belgian Coordinated Collections of Microorganisms (BCCM), StrainInfo now integrates all existing identifiers and information on microbial strains from culture collections, sequence databases, and scientific literature to enhance their findability, reusability, and traceability.

The **Biodiversity Community Integrated Knowledge Library (BiCIKL)**, a Horizon 2020 project (2021–2024), played a pivotal role in establishing methods, tools, standards, and workflows for linking biodiversity data between Research Infrastructures (RIs), including links between genomic and collection and taxonomic data (Penev et al. 2022). A special focus of BiCIKL was the extraction and liberation of data from millions of pages of published literature, making it accessible and reusable. BiCIKL delivered global-level access to data and tools for seamless linking and usage along the data life cycle: collecting specimens > extracting molecular sequences > identifying species > analyzing and publishing results > constructing a biodiversity knowledge graph > reusing data for new discoveries and addressing societal needs.

Biodiversity biobanks

Biodiversity biobanks collect, store, and curate biological materials (such as tissues, cultures, environmental samples, and biochemical isolates, such as DNA or RNA; see figure 2) along with their associated metadata under standardized conditions that maintain viability or molecular integrity (Hewitt and Watson 2013, International Organization for Standardization 2018, Astrin and Corrales 2023, International Society for Biological and Environmental Repositories 2023). The physical samples are stored in specialized facilities, often at ultra-low or cryogenic temperatures (Allison et al. 2021, Corrales and Astrin 2023, Reichel et al. 2026). Biodiversity biobanks support a broad spectrum of research fields, such as taxonomy, genetics, evolution, ecology,

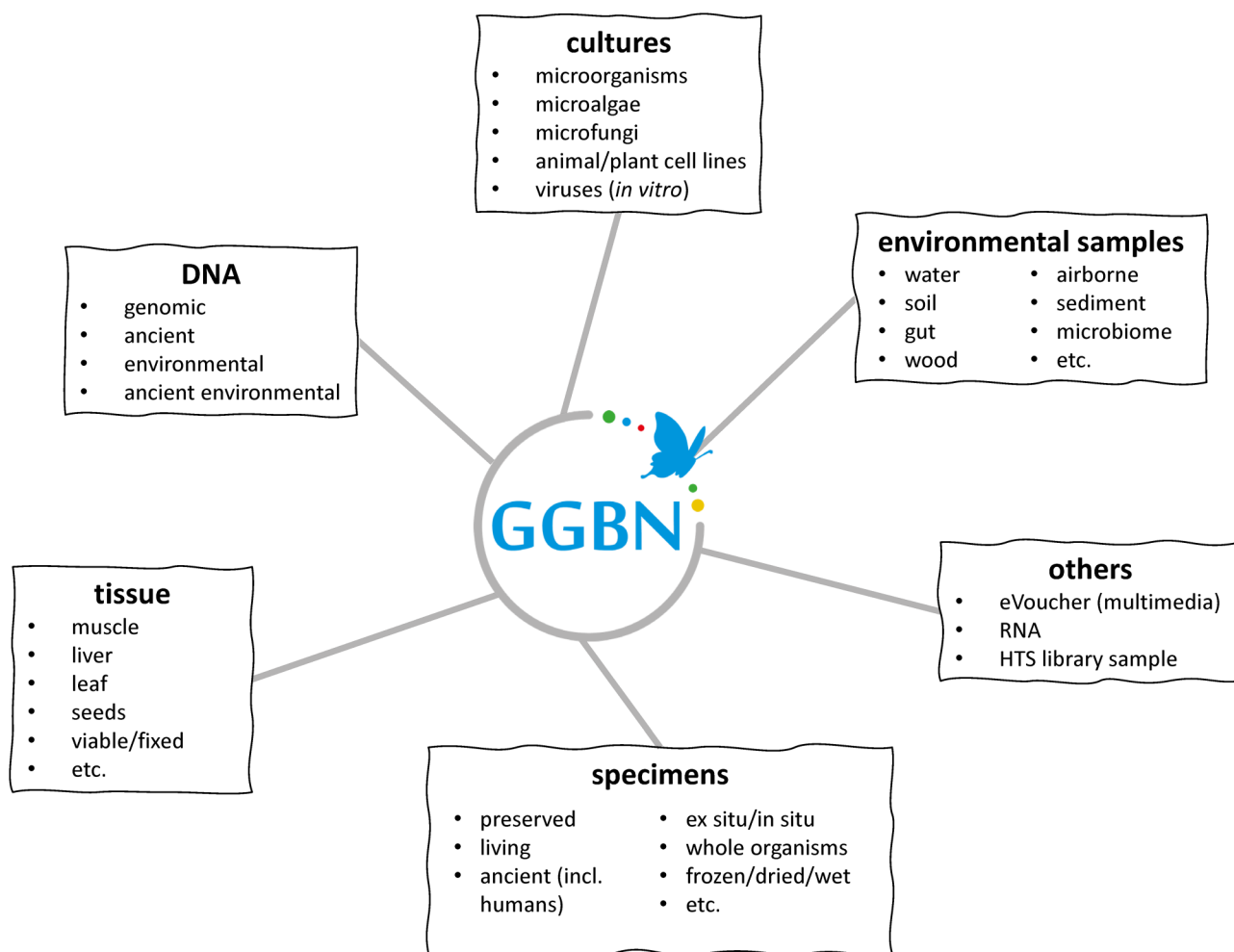


Figure 2 The diversity of biological material in GGBN's collections is categorized and described through recommended vocabularies, shown here along with examples. As a network of biodiversity biobanks, GGBN does not include human genetic resources, except for ancient human remains and organisms found in or on humans. Fossils are out of scope due to the complete degradation of DNA.

and biotechnology, as well as a wide range of applications, including biodiversity conservation and monitoring, breeding, artificial reproductive technologies, and the enforcement of biodiversity protection legislation. In addition to maintaining specific storage conditions for their physical samples, biodiversity biobanks developed sophisticated databases to manage the metadata associated with these samples, enabling published scientific results to be linked to the corresponding biological material, such as voucher specimens and sequence information. This is essential for reproducibility (Yeates et al. 2016, Buckner et al. 2021, Renner et al. 2024). Moreover, biobanks establish permanent links with preserved specimens and cultures, which provide authoritative taxonomic and morphological references, whereas other links associate tissues, environmental samples, or DNA extracts with DNA sequence data used for genomic and phylogenetic analyses (Schindel and Cook 2018, Lendemer et al. 2020). Physical samples preserved in biodiversity biobanks can be used for multiple molecular analyses, thereby enabling the integration of research data to become additive regarding the same specimen (Kilian et al. 2015).

DNA sequence and other molecular information is increasingly relevant for discovering and delimiting species, reconstructing the Tree of Life, health research, monitoring biodiversity, enforcing regulations (e.g., molecular identification of samples of organisms poten-

tially falling under the Convention on International Trade in Endangered Species of Wild Fauna and Flora [CITES]), assessing intraspecific genetic variation, and supporting a broad spectrum of applications in bioeconomy (including biotechnology, breeding, and bioprospecting). Beyond serving as a source for scientific information, banked samples, often particularly the viable ones (e.g., seeds, frozen cell cultures, and gametes), are also used in conservation and other applications. Consequently, biological material organized in biodiversity biobanks and its associated data are a critical but underrepresented component of biodiversity infrastructures and data systems. Biodiversity biobanks, therefore, also constitute a major element in a collectomics approach, which aims to digitize and interconnect materials preserved in collections with associated information (e.g., publications and genetic sequences), as well as with infrastructures such as GBIF and INSDC, thereby supporting efforts to address global challenges (Sigwart et al. 2025).

GGBN's mission and achievements

The Global Genome Biodiversity Network (GGBN) (www.ggbn.org/), often also referred to as the Global Biodiversity Biobank Network, was

founded in 2011 as an international consortium of biodiversity repositories holding biological materials, preserved at the molecular level (Droege et al. 2014). Since its inception, GGBN has grown to 121 member organizations from over 40 countries worldwide and has evolved into a highly interconnected, community-driven, and collaborative network (figure 3).

The Executive Committee is at the core of GGBN's governance. It functions as the primary decision-making body and is responsible for setting strategy, policy, direction, and budget for GGBN; developing the network's work plans; fundraising; reporting; and overseeing membership applications. Additionally, the Executive Committee establishes dedicated task forces (box 3) to complete high-priority tasks for GGBN's development.

Box 3 GGBN Task Forces

GGBN operates a **dynamic task force structure** that enables the effective addressing of current technical and policy-related topics. These include:

- **Data Standards and Data Access for Biological Material:** Establishes guidelines for standardizing and sharing data for biological material and streamlines and implements data standards
- **Policies Related to Management and Stewardship of Biological Material:** Promotes Access and Benefit-Sharing (ABS) by raising awareness, developing best practices, and assessing implementation
- **Biobank Procedures:** Develops repository workflows and studies sample preservation methods, leading to best practices and guidelines
- **Document Library:** Curates the documents in the GGBN Document Library
- **Communications and Outreach:** Recruits new members, engages global stakeholders, and prepares outreach materials

GGBN's day-to-day operations are supported by two offices for a fixed term of 5 years. The Secretariat Office handles overall coordination, administration, communications, and member support. It was hosted by the Smithsonian National Museum of Natural History until the end of 2025, when the Leibniz Institute for the Analysis of Biodiversity Change, Museum Koenig Bonn, assumed this role until the end of the term. The Technical Management Office manages the GGBN Data Portal, Document Library, and the Registry and provides technical support to members. The Technical Management Office is hosted by the Botanic Garden and Botanical Museum Berlin and, since 2021, is funded by the German Federal Ministry of Research, Technology, and Space (BMFTR).

As of August 2026, 45 members are contributing data to the GGBN portal, totaling over 7 million biological material records. This count is rapidly rising (figure 4). The strong increase of data in 2019 and constant growth in membership reflects the success of GGBN's conference series, including two regional workshops in 2017 (Asia and North America), webinars (e.g., during the pandemic), and the GGBN-Global Genome Initiative (GGI) Awards Program (sponsored by the Smithsonian National Museum of Natural History from 2017 to 2022), all of which helped to mobilize data, as well as two joint conferences (one in 2016 with the Society for the Preservation of Natural History Collections, SPNHC, and one in 2025 with the International Society

for Biological and Environmental Repositories, ISBER). Another significant milestone was the publication of the GGBN Data Standard in 2016 (Droege et al. 2016), which facilitates data sharing through GGBN's data portal.

The Global South is home to many of the world's most biodiverse regions; however, the presence of GGBN—and of biodiversity biobanks—in these areas is still limited. GGBN is working to enhance international visibility and encourage active participation from institutions in the Global South by (1) organizing biannual conferences in the region (e.g., Mexico 2023, South Africa 2025) and (2) inviting key collections and speakers from the region. Currently, four representatives of GGBN's Executive Committee are based in the Global South, including chair and vice chair. Furthermore, over the years the GGBN-GGI Awards Program has been an important tool to empower collections from low- or middle-income countries, many in the Global South.

GGBN's collections span the entire Tree of Life. Animal samples comprise the largest contribution to GGBN's data portal, both in terms of the number of taxa represented and biological material records (figure 5).

The samples within GGBN are not equally suited to all applications, and not all of them are stored in liquid nitrogen or -80°C freezers. A very large proportion of samples contain DNA fragments spanning several tens of thousands of base pairs. However, a growing number of samples now contain high-molecular-weight DNA, as illustrated by the reference genome sequencing projects featured in this article (both the voucher specimens from these initiatives and newly collected materials intended to support future genome sequencing efforts). Additionally, viable biological materials, particularly cell cultures, are becoming increasingly relevant within GGBN, opening completely new pathways to conservation and research.

All samples held by GGBN members represent valuable resources for the scientific community, particularly when information on sample quality is available. The GGBN data standard supports detailed metadata descriptions, and the use of standardized vocabulary for sample preservation methods is strongly recommended. Despite this, only 18% of GGBN records currently include preservation information through the GGBN portal. In practice, researchers frequently request this data directly from collection managers and curators. GGBN is actively working to increase the proportion of records with detailed preservation metadata online.

Technical GGBN infrastructure

The GGBN infrastructure is built entirely on open-source technologies and consists of four components. First, at its core is the GGBN Data Portal, which provides a unified, global access point to biodiversity biobanking resources held by GGBN members, and unique linkages between molecular data and the underlying biological material. With the release of a new version of the GGBN Data Portal in 2025, a new system of HTTP-based (Hypertext Transfer Protocol) globally unique and persistent identifiers (GUIDs) was introduced for all object types processed within GGBN. Participating biobanks do not need to change or replace their established local GUID systems; instead, a central service synchronizes local IDs with the unified GGBN identifiers. In addition, a mechanism was implemented to redirect calls to GGBN identifiers to the corresponding object metadata in the form of HTML (Hypertext Markup Language) pages or machine-readable RDF (Resource Description Format) representations depend-

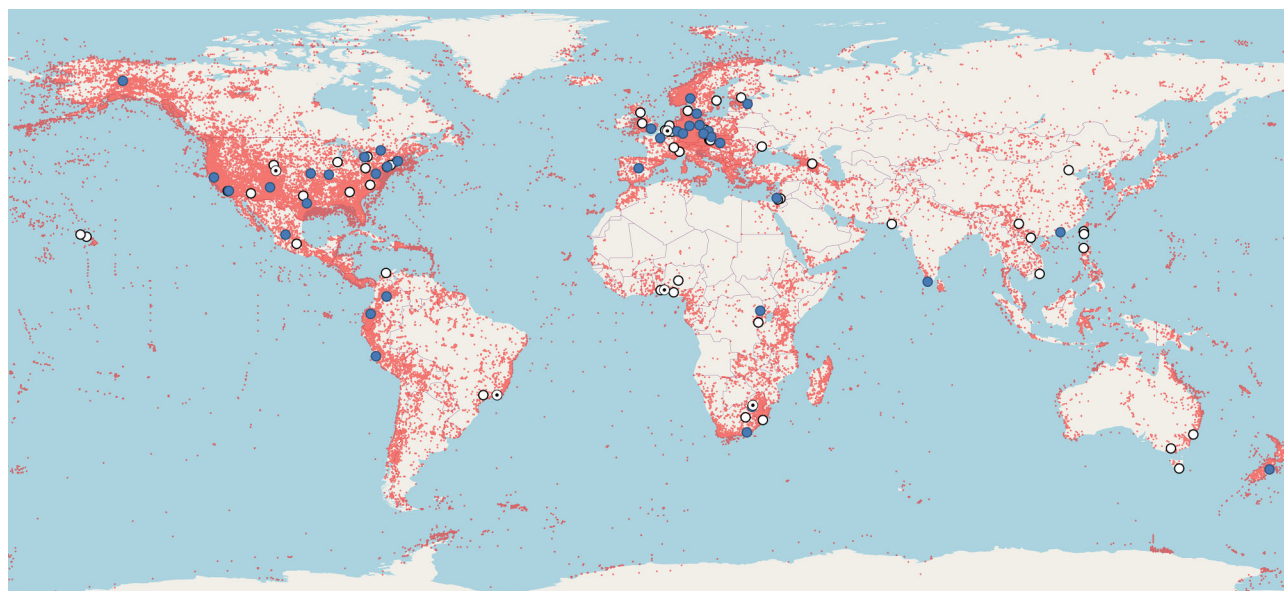


Figure 3 GGBN members and records of biological material available on the GGBN Data Portal. This map shows sample locations (small dots) for ca. 5.4 million georeferenced biological material records, such as tissues, DNA extracts, cultures, specimens, and environmental samples, as well as all 121 GGBN members (larger circles). The latter are classified in three groups: data-providing members (filled circles, $n = 45$), non-data-providing members (empty circles, $n = 71$), and consortia (circles with dot, $n = 5$) (see also www.ggbn.org/members). Map created with QGIS and OpenStreetMap.

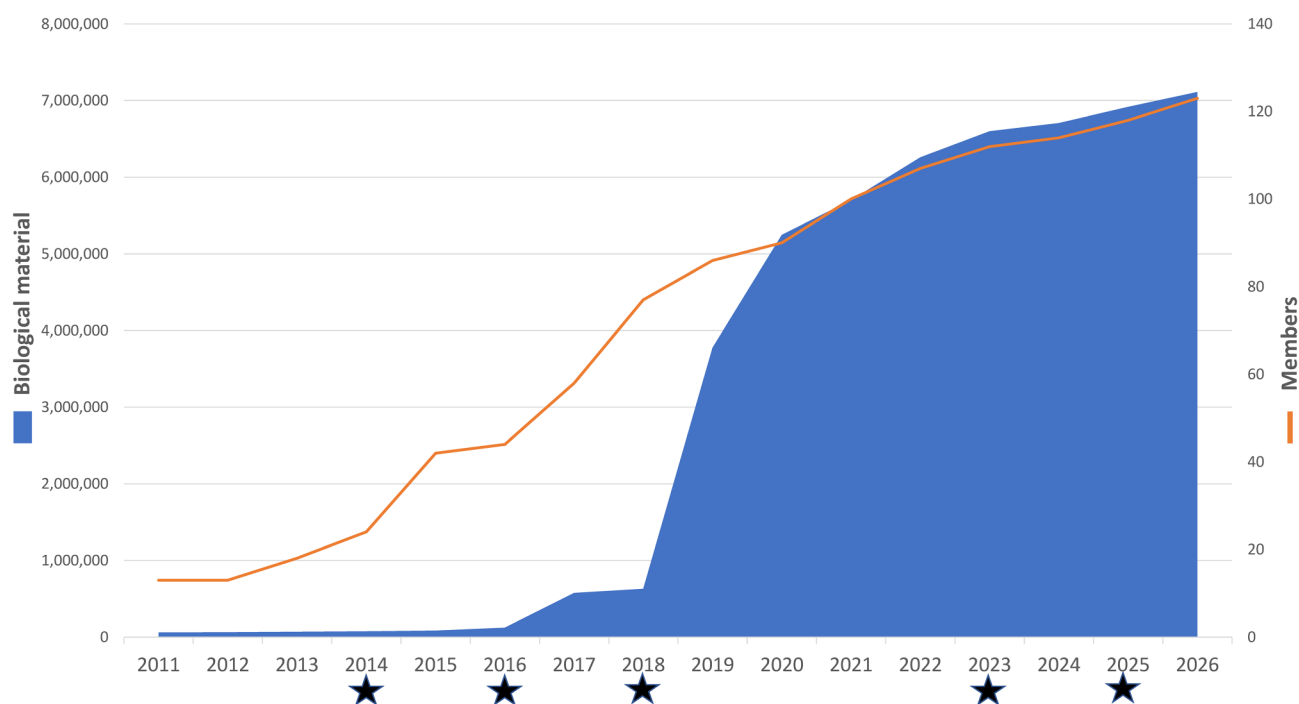


Figure 4 Growth in members and biological material records available on the GGBN Data Portal over time. Starting with 60,000 records and 13 members in 2011, GGBN currently (August 2026) has 121 members and 7.1 million biological material records shared through its data portal. Stars = GGBN conferences. The significant increase in data between 2018 and 2020 is based on the successful establishment of the conference series and the publication of the GGBN Data Standard, which enabled large collections or consortia such as Arctos and Guelph to make their data available through GGBN.

ing on whether the Uniform Resource Identifier (URI) was called by a human or a machine (content negotiation). This new consistent identifier layer creates the foundation for cross-linking between GGBN and other biodiversity informatics infrastructures, including literature (figure 1).

As part of the development of the new portal, a first example of such cross-linking was established between GGBN and ENA (figure 6). It is now possible to link sequences in ENA (as part of INSDC) to the corresponding objects in GGBN, and vice versa. Linking to ENA further enhances GGBN data. Figure 6 demonstrates how associated

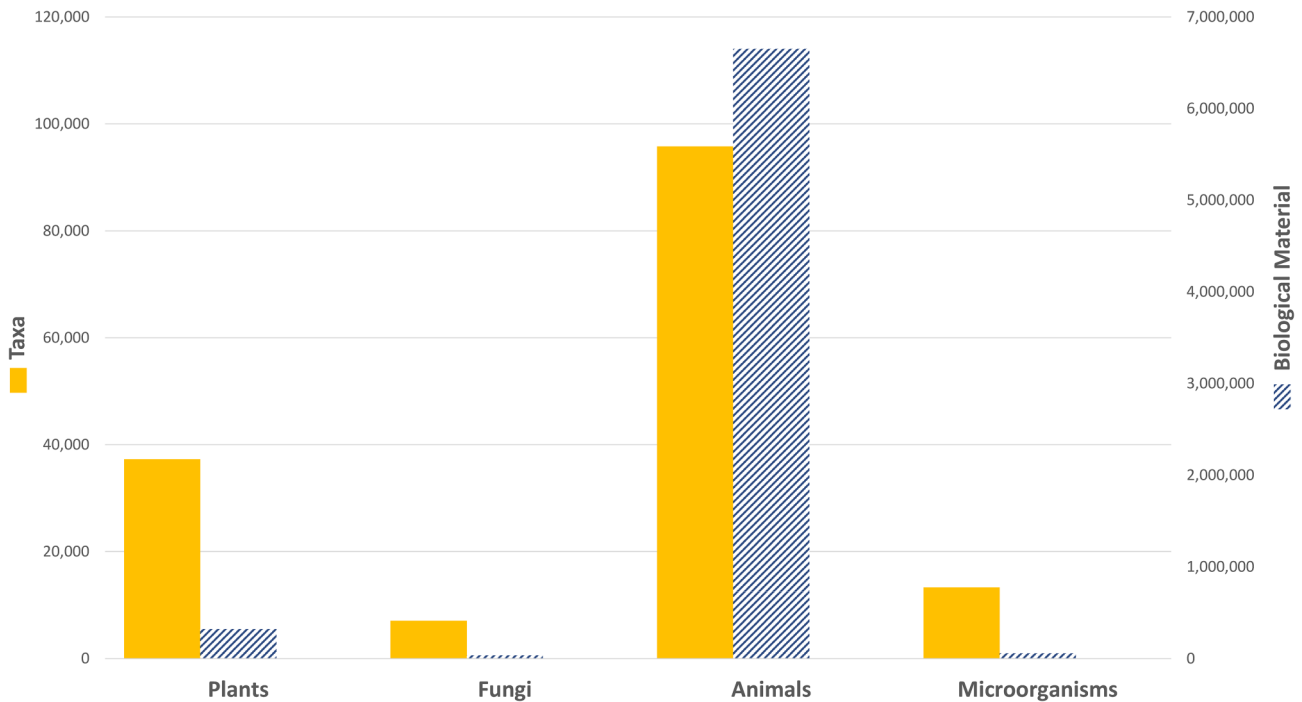


Figure 5 Representation of major taxonomic groups on the GGBN Data Portal. Ca. 151,000 taxa are represented on GGBN's data portal; animals, comprising about 95% of all biological material records, are the largest major taxonomic group, followed by plants, microorganisms, and fungi. Different scales are used to combine this information.

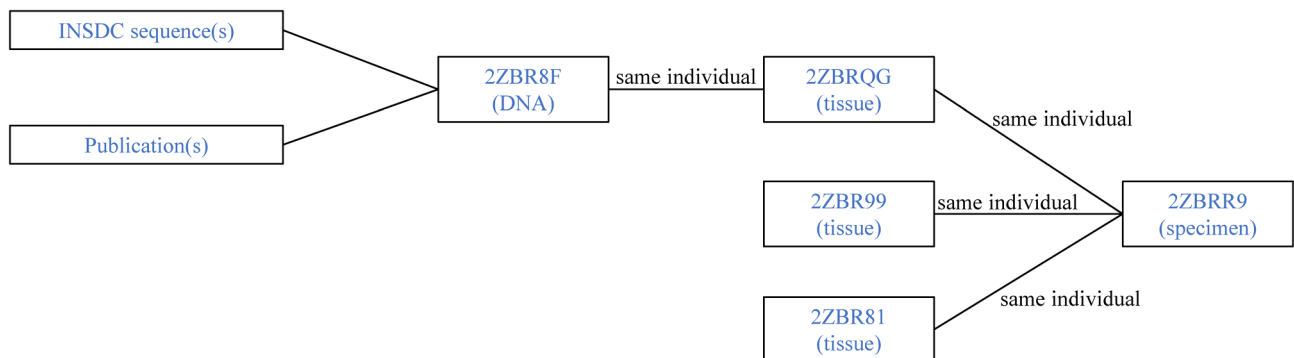


Figure 6 Hierarchical relationships between a physical DNA sample (GGBN object <https://id.ggbn.org/2ZBR8F>), other tissue samples of the same specimen, and related resources, such as sequences and publications. The graph is taken from GGBN's identifier landing pages. The example shows how through the use of GGBN identifiers, INSDC sequences and publications are associated with three material entities (2ZBR8F (DNA), 2ZBRQG (tissue), 2ZBRR9 (specimen)). Two additional tissues (2ZBR99 and 2ZBR81) are related to the same specimen but have no association with DNA, sequences or publications. The relationship between materials is defined by the data-providing collection (e.g., "same individual").

publications are identified by connecting ENA records with publications linked to GGBN identifiers, based on the known relationship between the two datasets. Building on this concept, cross-linking between GGBN and further data infrastructures (e.g., DiSSCo) is planned.

This newly conceived core infrastructure of the network, developed with German federal (BMFTR) support, makes an essential contribution to the interoperability of GGBN data within the biodiversity data infrastructure landscape, and provides a crucial foundation for processing the continuously growing data volumes in the field of biodiversity biobanks.

The second data infrastructure component is the member registry, which provides information about each of GGBN's 121 member collections, together with the corresponding data. Third, the Document Library provides access to over 1300 documents, including presen-

tations from GGBN conferences and other events, reports, and published and gray literature related to biobanking. The fourth component is a wiki that complements the document library and provides general information about GGBN (e.g., Terms of Reference, How to become a member). All four resources provide unique biobank-related information for the community and are available on the GGBN web portal.

How the scientific community uses GGBN

GGBN provides a trusted hub for data sharing and integration, knowledge exchange, and a physical distributed biobank infrastructure (see figure 7).

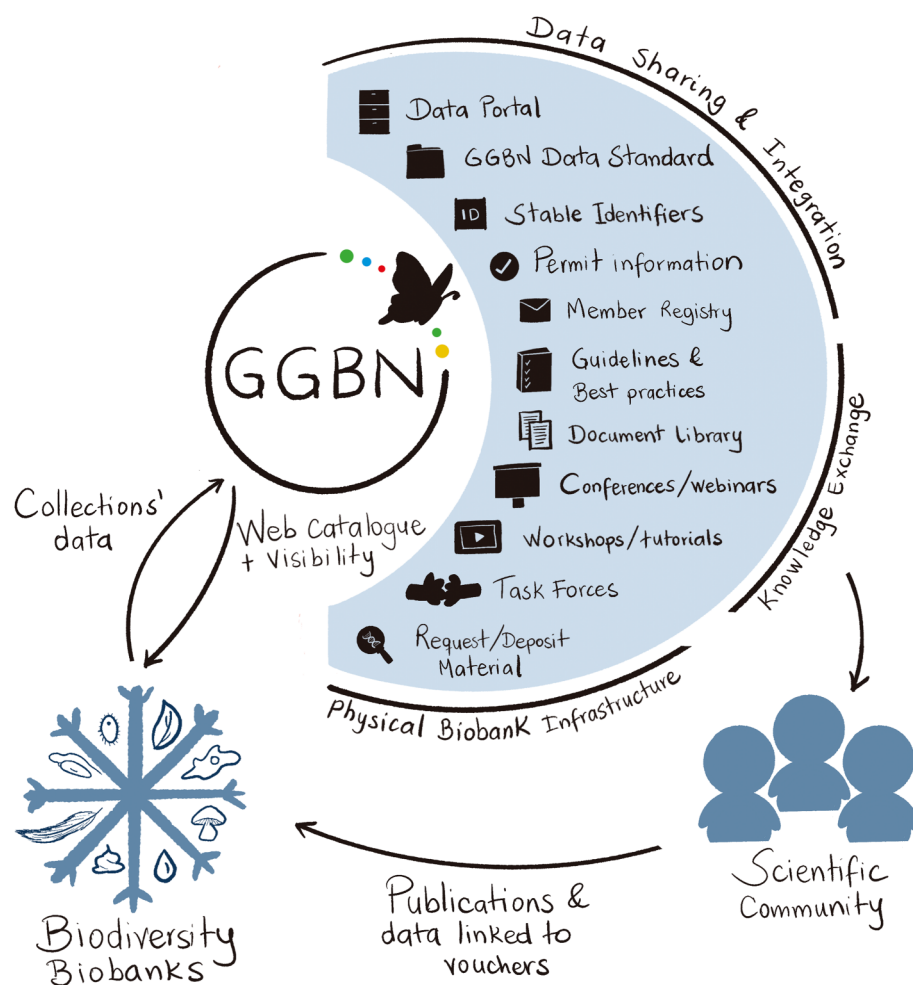


Figure 7 GGBN's core activities: Data sharing and integration, knowledge exchange, and physical biobank infrastructure. This diagram illustrates the three core activities of GGBN that underpin its role in advancing biodiversity research and conservation. The figure emphasizes GGBN's unique contributions to the scientific community through its structured approach to data sharing and integration, community-driven knowledge exchange, and physical infrastructure of member biobanks.

In particular, GGBN serves the following functions for the community:

Data sharing and integration

Through its data portal, GGBN mirrors and integrates biodiversity biobanks held by its members, offering centralized, citable access to biological material and associated data. A common persistent identifier infrastructure enables genomic DNA, tissue samples, voucher specimens, and other collection objects to be uniquely identified, cited, and linked to sequence data, literature, permits, and related resources across institutions. This supports both individual end users as well as large platforms and networks (e.g., INSDC, BOLD, DiSSCo, NFDI, European Reference Genome Atlas [ERGA], Earth BioGenome Project [EBP], Consortium of European Taxonomic Facilities [CETAF]), while providing essential attribution to member collections as well as promoting the Extended Specimen concept (Lendemer et al. 2020), which integrates genetic and morphological data. GGBN member collections benefit from global visibility through the GGBN data portal, serving researchers with a single point of access to biobanked materials held across multiple partner institutions. This shared approach enhances research resource accessibility, reduces the need to search across indi-

vidual institutional portals, and strengthens collaboration across the biodiversity biobanking community.

The network has a strong cross-domain approach, aiming to bring together a diverse range of communities, collection types, and stakeholders that have an interest in biobanking for different purposes such as research, species conservation, or applications in forestry and agriculture. By taking into account their knowledge and requirements, GGBN has developed an infrastructure that facilitates data sharing across many biobanking domains. Targeted outreach toward key stakeholders, such as the European Association of Zoos and Aquaria (EAZA), as well as networks and organizations such as the Union for Conservation of Nature (IUCN) Species Survival Commission (SSC), Animal Biobanking for Conservation Specialist Group (ABCSG), and the World Federation for Culture Collections (WFCC), has led to partnerships and collaborations, thereby enhancing visibility and access to biological material of so far underrepresented domains, including microorganisms and fungi, and information that can be used for the development of conservation strategies. (For a description of the mentioned partners, see box 4.) To support these efforts, GGBN members receive guidance on aligning internal databases with GGBN data standards, whereas collaborations with open-access collection man-

agement systems (e.g., Specify, Arctos, Symbiota) further help mobilize data.

The GGBN Data Standard supplements existing frameworks such as the Access to Biological Collection Data (ABCD) standard for exchanging collection and observation data (Holetschek et al. 2012) and Darwin Core (DwC), the most widely used standard for sharing biodiversity occurrence data (Wieczorek et al. 2012) by introducing specialized terms for laboratory data (e.g., extraction or preparation method and date, DNA concentration) and preservation data (e.g., storage temperature and medium), as well as information on associated permits and loan availability, which are key elements mostly missing in collection databases before the introduction of the GGBN Data Standard. The GGBN Data Standard constitutes the basis of the GGBN data portal (Droege et al. 2016) and enables biobank data management in compliance with the Nagoya Protocol. It integrates terms from the MIxS (Minimal Information about any (x) Sequence) standard (Yilmaz et al. 2011) and DwC, and contributes terms to GBIF's data mobilization of DNA-derived data and Metabarcoding Data Programme. Organizations such as ABCSG and WFCC benefit from collaborating with GGBN to include new collection types and terms. Those organizations can adopt the GGBN Data Standard, which enhances interoperability with other platforms, such as GBIF, and their samples can gain more visibility and discoverability through the GGBN data portal, particularly benefiting underrepresented collections or taxa. In addition, conservation practitioners get improved access and a transparent and reliable framework for requesting genetic resources through the network.

For example, collection curators use the GGBN Data Standard to record standardized metadata and maintain explicit links between genomic samples, their voucher specimens, and derived sequence data. Persistent identifiers ensure that these relationships remain traceable across institutional databases and international repositories; this makes genomic samples easier to discover, reuse, and integrate into global research infrastructures while supporting FAIR and reproducible research. In addition, collaborating with GGBN improves sample access for conservation practitioners through a well-established framework that complies with international legislation.

Knowledge exchange

By convening curators, scientists, and data managers from a broad spectrum of organizations (e.g., museums, botanic gardens, zoos, research institutions, and universities), GGBN promotes transdisciplinary dialogue that fosters innovation and capacity-building and creates a dynamic forum for the sharing of expertise, best practices, and emerging solutions in biodiversity biobanking. Regular workshops, webinars, and global conferences offer members and partners opportunities to collectively address challenges such as data standardization, sample preservation, ethical and legal considerations, and the practicalities of implementing new technologies or workflows. The topics for these may be proposed by the Executive Committee, Task Forces, or by individual members, ensuring that GGBN's activities align with the network's strategic goals and heeding new challenges faced by the collections. These activities not only support the technical integration outlined in the previous section, but also help shape a shared understanding of the responsibilities and best practices required for effective biobank stewardship worldwide. GGBN's knowledge exchange thus complements its data sharing activities and

strengthens global cooperation, ensuring that the value of biobanked resources is maximized for research, conservation, and education.

A key aspect of GGBN's knowledge-sharing efforts is its support for policy compliance, particularly regarding international ABS regulations, through the GGBN Policies Task Force, which provides guidance and practical support on ABS-related challenges to GGBN members and the broader biodiversity research and biobanking community. To meet international obligations, researchers, biological collections, and biobanks that routinely receive, preserve, and distribute biological material globally must understand the Nagoya Protocol and its ABS principles. In simple terms, ABS means obtaining permission before accessing and utilizing biological material and ensuring that benefits are shared with the country where those were originally collected. GGBN supports FAIR data sharing by making information about biological material discoverable and interoperable, while supporting compliance with ABS requirements by linking sample records to relevant ABS-related restrictions and documentation, where available. This enables researchers to discover collections and understand the conditions associated with their use, although recognizing that access to and utilization of the underlying biological material remain subject to permit conditions and applicable ABS requirements under the national legislation of provider countries.

As a best practice, collections should store Nagoya Protocol-relevant information and documentation together with biological material, as it facilitates future transfer and responsible utilization. The GGBN Policies Task Force has facilitated this process by making support material freely available on the GGBN Data Portal. These resources include guidance for researchers and curators, as well as compilation of best practices, codes of conduct, and learning tools that could be tailored and used by institutions beyond GGBN.

ABS compliance is further strengthened through GGBN by integrating standardized terms for loans and permits into its data framework. For example, the GGBN Data Standard includes fields for recording and tracking permits under the Nagoya Protocol, ensuring institutions can document and share material responsibly, a critical step for global biodiversity research.

Moreover, ABS goes beyond genetic resources and the Nagoya Protocol. Countries began claiming benefit-sharing from the use of genetic sequences available in open-access databases, also known as digital sequence information (DSI). GGBN's Policies Task Force, in collaboration with the DSI Scientific Network, has actively engaged in international negotiations under the Convention on Biological Diversity (CBD), developing information materials such as policy briefs and peer-reviewed papers, and coordinating outreach with country delegates (all available at www.dsiscientificnetwork.org).

Physical biobank infrastructure

GGBN's member collections constitute the physical foundation for depositing, managing, accessing, and providing biological material, in accordance with agreed-upon quality standards and common policies. GGBN biobanks serve as critical sources of reference materials for international projects such as ERGA and iBOL (International Barcode of Life), ensuring high-quality genomic or high-throughput sequencing.

Stakeholders

GGBN's success relies on both its members and collaborative relationships with a diverse array of stakeholders across the biodiversity data

landscape. These stakeholders can be broadly categorized into several groups: professional networks and community groups (such as ABCSG, CETAF, EAZA, or WFCC) that connect dedicated organizations and specialized experts around collaborative research, conservation, and biobanking goals; sequencing initiatives and consortia (including EBP, ERGA, and iBOL Europe) that drive large-scale genomic sequencing efforts; and data infrastructures and standards described in the background chapter above and in box 2 that compile and provide access to biodiversity data at global scales while developing data standards and technical frameworks. Together with GGBN, these stakeholders enhance an interconnected ecosystem that facilitates data sharing, standardization, and scientific collaboration across domains and collections, taxonomic groups, and geographic regions.

Professional networks and community groups

Professional networks and community groups are committed to preserving biodiversity by maintaining high-quality biological collections, whether they focus on animals (ABCSG, EAZA), microorganisms (WFCC), or all biodiversity (CETAF) (box 4). Many of their members are also GGBN members, forming an integral part of GGBN's physical biobank infrastructure and expertise. Each network and community brings together institutions and experts to share resources, knowledge, and capabilities, recognizing that biobanked samples and well-documented collections are essential for genomics and molecular research, taxonomy and systematics, conservation planning and management, and biotechnology and applied science. These organizations rely on and actively support global access to collections, contributing to and benefiting from FAIR data principles and data linkages provided by GGBN's infrastructure. Across all communities, the establishment and adoption of standardized protocols, high-quality sample preservation, proper documentation, and ethical and legal compliance (e.g., ABS) are recognized as fundamental to enabling effective and responsible use of biological materials. Each of these professional networks and communities is vital for GGBN's cross-domain approach, encouraging active participation from their members, thereby promoting greater visibility and availability of, for example, underrepresented taxonomic groups, such as microorganisms, and associated data.

Box 4 Professional networks and community groups

The World Federation for Culture Collections (WFCC) serves as a global authority on all matters related to culture collections such as the operation and management as well as addressing issues in a wider context such as the importance of (i) standardization and best practices guidelines, (ii) networking, capacity building, and education, (iii) postal, quarantine and safety regulations, (iv) intellectual property, patent and commercialization, (v) access, policies, and legal frameworks, and (vi) sustainability of endangered collections. WFCC interacts with different global organizations to promote the importance of culture collections with emphasis placed on the contributions and impact they make on science, biotechnology, agriculture, health, education, and society. Microbial culture collections have been providing biological resources to many different sectors, as well as offering preservation services to microbiologists working in different fields. Historically, GGBN has been dominated by zoological and botanical biobanks. For instance, WFCC members such as the German Collection of Microorganisms and Cell Cultures (DSMZ) now share fungal

and bacterial cultures through GGBN's portal—expanding its reach beyond traditional animal and plant collections and supporting studies in environmental microbiology. The WFCC-GGBN partnership, for instance, enables access to microbial cultures for conservation efforts.

The Animal Biobanking for Conservation Specialist Group (ABCSG) was formed by the International Union for Conservation of Nature (IUCN) Species Survival Commission (SSC) in 2022. It aims to unite a global community of practitioners that work on animal biobanking for conservation purposes. The ABCSG currently has four Task Forces. Three of these are committed to different specializations in biobanking technologies: (i) viable cell cultures, (ii) reproductive cells, and (iii) blood fractions and frozen tissues. The fourth Task Force on Justice, Equity, Diversity, and Inclusion focuses on reshaping biobanking practices consistent with ABS practices that properly acknowledge the FAIR and CARE principles. The ABCSG also recognizes the need for a framework to ensure that information on samples is available to conservation practitioners, and the role that GGBN can play to support this. Moreover, the IUCN SSC has also created the Microbial Conservation Specialist Group, which aims to integrate microbial biodiversity into conservation frameworks (Gilbert et al. 2025).

The European Association of Zoos and Aquaria (EAZA) Biobank is a centralized, long-term repository for high-quality samples from the animals held in its member institutions and wild counterparts. Uniquely, in partnership with Species360, the EAZA Biobank is integrated in the globally largest database for managed populations, the Zoological Information Management System (ZIMS), enabling direct linkage between sample data and the animal's existing records in ZIMS, its demographic data, and health records. The EAZA Biobank represents a strategic shift from fragmented sample storage to a coordinated, scientifically robust system that underpins modern zoo population management, supporting conservation and research. Looking ahead, EAZA Biobank plans to integrate its data more thoroughly with GGBN to increase its contribution to making samples and research results available for species conservation.

The Consortium of European Taxonomic Facilities (CETAF) is Europe's network of biological and geological collections-holding research institutions. CETAF acts as a catalyst for integrative taxonomic research, enhancing species description and data linkage. The collaborative character of modern taxonomy has always called for reliable documentation and the possibility to revisit material and literature for the validation of existing and the addition of new knowledge. The advent of molecular methods in taxonomy has enormously sped up and refined taxonomic knowledge generation, whereas advances progressively made under different initiatives, such as TETTRIs (Transforming European Taxonomy through Training, Research and Innovations), ensure that taxonomy relies on correct standards and protocols, and is supported by innovative methods and tools.

Sequencing initiatives

Sequencing initiatives are large-scale, long-term scientific programs that collect or access biological specimens and sequence their DNA or RNA to create genetic reference databases for biodiversity. iBOL Europe focuses on DNA barcodes, ERGA creates complete reference genomes for European species, and EBP strives to create a comprehensive genomic library of all known eukaryotic species on Earth. All of these initiatives depend on reliable specimens and associated metadata deposited in repositories and made available, for example, through GGBN, to ensure that genetic information is trustworthy and

can be verified. Their shared goal is to create accessible, scientifically robust genetic resources that support, for example, conservation, biomonitoring, taxonomy, or evolutionary research. Sequencing initiatives are important users of and contributors to biobank collections (Astrin et al. 2013), as exemplified for GGBN by the projects highlighted in box 5.

Box 5 Sequencing initiatives

The **Earth BioGenome Project (EBP)** aims to sequence the genomes of ~1.8 million named eukaryotic species in 10 years, a “biology moonshot.” The project is nearly halfway toward meeting the Phase I goal of sequencing 10,000 species, including one representative species of half of the ~10,000 eukaryotic families. Looking ahead, EBP envisions deeper integration between GGBN’s sample registries and genome project tracking systems, enabling end-to-end traceability from field collection to genome assembly.

The **European Reference Genome Atlas (ERGA)** is Europe’s regional node of the Earth BioGenome Project with a mission to coordinate efforts across projects and researchers working on producing and utilizing reference genomes for European eukaryotic biodiversity. ERGA is committed to guiding members to follow standardized and harmonized procedures and practices to ensure that genomes are accompanied by rich metadata and associated reference materials, including traditional morphological vouchers, digital images, and molecular vouchers such as tissues and/or nucleotide extractions. Moving forward, ERGA aims to continually adapt its procedures in line with technological advancements in the field while strengthening connections with stakeholders and end-users to ensure that reference genome resources are both scientifically robust and broadly accessible.

iBOL Europe is the European node of the International Barcode of Life Consortium (iBOL). It was established in 2021 to coordinate the growing DNA barcoding community across Europe by connecting existing networks, scientists, and projects exploring biodiversity using DNA barcoding. The iBOL project aims to deliver DNA-based species identification and discovery, unravel biological interactions, and enable high-throughput DNA biomonitoring. A key part of this work for iBOL Europe is to produce, curate, and provide access to a complete DNA barcode reference library of European eukaryotic biodiversity. A crucial element to enable this goal is the availability of well-identified specimens, with associated vouchers and suitable metadata. Its mission aligns closely with GGBN, which provides centralized access to genomic data and encourages high-quality standards.

Conclusions

The Global Genome Biodiversity Network has become a pivotal force in the global biodiversity collections and data infrastructure landscape by unifying biodiversity biobanks—repositories of biological materials such as tissues, cells, DNA, environmental samples, cultures, and specimens—to enhance genomic research, conservation, health, and many other disciplines. Through its data portal, standardized frameworks (e.g., the GGBN Data Standard), and collaborations with networks such as WFCC or ERGA, GGBN has enabled the centralized discovery and sharing of over 7 million biological material records across GGBN member institutions. Its identifier system and

cross-infrastructure linkages with platforms such as ENA have enhanced interoperability, whereas the GGBN-GGI Awards Program has expanded representation of poorly covered regions and taxa. By prioritizing ABS compliance and fostering partnerships across disciplines, GGBN has not only improved data-driven biodiversity research processes but also reinforced the ethical and equitable use of biological resources globally.

Looking forward, GGBN’s success hinges on three pillars:

1. **Standardization:** Deepening alignment of the GGBN Data Standard with evolving frameworks such as Darwin Core, the Darwin Core Data Package, and MIxS. This standardization ensures seamless interoperability with global systems while incorporating emerging standards to reflect the dynamic landscape of biodiversity informatics.
2. **Integration:** Scaling the ENA linking model to platforms such as GBIF and DiSSCo, leveraging GGBN’s unified identifier system to create transparent connections between biological material, sequence data, and metadata. This integration ensures traceability of derivation histories and supports reproducibility, with open standardized interfaces enabling full integration into the global biodiversity knowledge graph.
3. **Collaboration:** Fostering knowledge sharing and collaboration among biobanks and institutions through GGBN conferences, webinars, and the document library. GGBN relies on an engaged global community, as reflected in its member-driven governance. Topics such as data management, ABS, and sample preservation are addressed collectively, helping shape and adapt best practices.

It is commonly recognized that further biodiversity biobanks should be established in the Global South to counteract biodiversity loss (Angeles and Catap 2022, Streit et al. 2024). To achieve this goal, prioritizing capacity-building in institutions in that region through programs such as the GGBN-GGI Awards is only logical, but several challenges remain to be addressed, including limited governmental research funding, infrastructure constraints, limited capacity building, and sequencing costs that often exceed those in wealthier nations, along with the need to directly connect genomic research with societal and conservation priorities (Corrales et al. 2023, Sánchez-Pacheco et al. 2025). To further strengthen the representation of the Global South within GGBN, it is planned to reach out to various genomic initiatives for potential partnerships. In Latin America, these include the Colombian Genomic Network, the 1000 Genomes Project from Chile, the Neotropical Biogenomes Latin American Network, and the Genomics of Brazilian Biodiversity consortium. East and Southeast Asia will be another key strategic focus over the coming years, with the 2027 GGBN meeting in Asia planned to serve as a catalyst for new partnerships and for the advancement of biodiversity biobanking across the region.

A key function of the GGBN network is its role as a mediator at the interface between physical sample and specimen collections, and the sequence information derived from them. As DNA-derived data constitutes a rapidly growing segment of biodiversity data, the systematic linking of global biodiversity data with the world’s biobank data will remain essential for achieving both FAIR and CARE principles. This linkage is also critical for enabling efficient, transparent, and trusted data exchange across disciplinary, institutional, and national communities involved in, for example, biodiversity and environmental research, health, and conservation.

However, during its establishment phase, GGBN had a strong focus on rapidly expanding the network and increasing the volume of data made available through the GGBN Data Portal. To date, only limited structured analyses have been conducted of data access patterns, associated requests for material, or the composition of the user community (e.g., GGBN members versus external users and their respective research communities). In the next phase of development, we intend to place a stronger emphasis on the user perspective, with the aim of better understanding the needs of different user communities and using these insights to guide the future evolution of the network.

Regarding physical molecular and viable collections, the increased deposition of samples in the distributed physical infrastructure operated by the GGBN members should be encouraged, particularly for large-scale initiatives, in order to ensure long-term storage and sustained access to—and use of—both data and samples. This must be achieved in full compliance with the Nagoya Protocol and applicable legal and regulatory frameworks. Accordingly, a continued core focus of GGBN will be the international coordination and documentation of shared procedures and standards for the storage, handling, and documentation of biological material, as well as for the consistent provision of associated metadata. This includes, in particular (but not exclusively), the development and dissemination of best practices for documenting and providing genome-quality information, which is especially critical when samples are used for long-read DNA sequencing and other emerging high-resolution genomic approaches.

The operation of the GGBN network will continue to require a broad and diverse range of activities, including community coordination, capacity building, standards development and documentation, software development, service operation, and help desk support. These activities are essential to maintaining interoperability, ensuring data quality, and supporting the growing, increasingly diverse user and biobank communities. Although some of these efforts can be supported through in-kind contributions from GGBN members and, to a limited extent, through membership fees, the stable functioning of the technical and administrative offices will continue to depend on substantial financial support from research funders. Such support is essential in recognition of the key role of the GGBN network in linking biodiversity data beyond the core domains of its partner institutions and in enabling cross-domain integration at a global scale.

List of abbreviations

ABCSG: Animal Biobanking for Conservation Specialist Group
ABCD: Access to Biological Collection Data
ABS: Access and Benefit-Sharing
API: Application Programming Interface
BCCM: Belgian Coordinated Collections of Microorganisms
BiCIKL: Biodiversity Community Integrated Knowledge Library
BMFTR: German Federal Ministry of Research, Technology and Space
BOLD: Barcode of Life Data System
CARE: Collective Benefit, Authority of Control, Responsibility, and Ethics
CBD: Convention on Biological Diversity
CETAF: Consortium of European Taxonomic Facilities
CITES: Convention on International Trade in Endangered Species of Wild Fauna and Flora
CIG: Compliance and Interoperability Working Group
DDBJ: DNA Data Bank of Japan

DFG: Deutsche Forschungsgemeinschaft (German Research Foundation)
DiSSCo: Distributed System of Scientific Collections
DNA: Deoxyribonucleic Acid
DOI: Digital Object Identifier
DSI: Digital Sequence Information
DSMZ: German Collection of Microorganisms and Cell Cultures
DwC: Darwin Core
EAZA: European Association of Zoos and Aquaria
EBP: Earth BioGenome Project
ELIXIR: European Life-sciences Infrastructure for biological Information
EMBL-EBI: European Molecular Biology Laboratory, European Bioinformatics Institute
ENA: European Nucleotide Archive
EOSC: European Open Science Cloud
ERGA: European Reference Genome Atlas
FAIR: Findable, Accessible, Interoperable, and Reusable
GBIF: Global Biodiversity Information Facility
GBWG: Global Biodiversity Working Group
GGBN: Global Genome Biodiversity Network
GGI: Global Genome Initiative
GSC: Genomic Standards Consortium
GUID: Globally Unique Identifier
HTML: HyperText Markup Language
HTTP: HyperText Transfer Protocol
iBOL: International Barcode of Life
ID: Identifier
INSDC: International Nucleotide Sequence Database Collaboration
ISBER: International Society for Biological and Environmental Repositories
IUCN: International Union for Conservation of Nature
MIxS: Minimum Information about any—x Sequence
NFDI: Nationale Forschungsdateninfrastruktur e.V. (National Research Data Infrastructure)
NLM-NCBI: National Library of Medicine, National Center for Biotechnology Information
PID: Persistent Identifier
RDF: Resource Description Format
RNA: Ribonucleic acid
RI: Research Infrastructure
SPNHC: Society for the Preservation of Natural History Collections
SSC: Species Survival Commission
TDWG: Biodiversity Information Standards
TETTRIs: Transforming European Taxonomy through Training, Research and Innovations
URI: Uniform Resource Identifier
WFCC: World Federation for Culture Collections
ZIMS: Zoological Information Management System

Acknowledgments

The work of GD, AG, DR, JH and TB was supported by the German Federal Ministry of Research, Technology, and Space (BMFTR) under grant agreement no. 03LC2020A. We thank Francesco Cerio for implementing the content negotiation system for the stable identifiers. GGBN's work is a community effort and would not be possible without the constant contribution of and support from GGBN members and partners. In particular we are very grateful to the South African National Biodiversity

Institute (SANBI), the Biodiversity Biobanks South Africa (BBSA) Project, International Society for Biological & Environmental Repositories (ISBER), and the Medical Biorepositories of South Africa (MBirSA) for providing support to the 2025 Joint Biobanking Conference and to the Universidad Autónoma de Aguascalientes for hosting the 2023 GGBN conference. We would also like to thank the organizers and contributors of the ELIXIR BioHackathon Europe 2023 for providing the collaborative environment that enabled the initial development of the GGBN-ENA cross-linking.

Data availability statement

Data are available in a repository and can be accessed using a unique identifier other than a DOI. The data underlying figure 6 of this article are available in the Global Genome Biodiversity Network at <https://www.ggbn.org>, and can be accessed with, <https://id.ggbn.org/2ZBRQG>, <https://id.ggbn.org/2ZBR99>, <https://id.ggbn.org/2ZBR81>.

Author contributions

Gabi Droege (Conceptualization, Funding acquisition, Project administration, Software, Supervision, Visualization, Writing—original draft, Writing—review & editing), Jonas Astrin (Conceptualization, Funding acquisition, Project administration, Supervision, Writing—original draft, Writing—review & editing), Anton Güntsch (Conceptualization, Funding acquisition, Project administration, Supervision, Visualization, Writing—original draft, Writing—review & editing), Andrew C Bentley (Writing—original draft, Writing—review & editing), Carolina Corrales (Funding acquisition, Project administration, Visualization, Writing—review & editing), Kim Labuschagne (Writing—original draft, Writing—review & editing), Gilberto Ocampo (Writing—original draft, Writing—review & editing), Wouter Addink (Writing—original draft), Hannah J. Appiah-Madson (Writing—original draft), Masanori Arita (Writing—original draft), Heather Avrili (Writing—original draft, Writing—review & editing), Katharine Barker (Project administration, Writing—original draft), Gustavo A. Bravo (Writing—original draft), Elena Buzan (Writing—original draft), Ana Casino (Writing—original draft), Guy Cochrane (Writing—original draft), Manuela da Silva (Writing—original draft), Dan Distel (Writing—original draft, Writing—review & editing), Marcus Ernst (Writing—original draft), Belen Escobari Vargas (Writing—original draft), Rosalia Falco Poulin (Writing—original draft), Kasia Fantoni (Writing—original draft), David Fichtmueller (Visualization, Writing—original draft, Writing—review & editing), Frank Oliver Glöckner (Writing—original draft), Peter Grobe (Writing—original draft), Quentin Groom (Writing—original draft), Vikas Gupta (Writing—original draft), Eva Häffner (Visualization, Writing—original draft, Writing—review & editing), Katherine D. Heinemann (Writing—original draft), Jörg Holetschek (Software, Writing—original draft), Peter M. Hollingsworth (Writing—original draft), Chris Hunter (Writing—original draft), Christina Hvilsom (Writing—original draft, Writing—review & editing), Sharif Islam (Writing—original draft), Vishnukumar Balavenkataraman Kadhivelu (Writing—original draft), D. Ipek Kurtböke (Writing—original draft, Writing—review & editing), Harris Lewin (Writing—original draft), Alice Carolyn McHardy (Writing—original draft), Anna Mekarska (Writing—original draft), Rachel Meyer (Writing—original draft), Melania Muñoz-García (Writing—original draft, Writing—review & editing), Carmen Paulmann (Writing—original draft), Joana Paupério (Funding acquisition, Writing—original draft, Writing—review & editing), Lyubomir Penev (Writing—original draft), Sujeevan Ratnasingham (Writing—original draft), Dominik Röpert

(Writing—original draft), Santiago R. Ron (Writing—original draft), Oliver A. Ryder (Writing—original draft), Dmitry Schigel (Writing—original draft), Amber Hartman Scholz (Writing—original draft), Lynn Marie Schriml (Writing—original draft), Boripat Siriaronrat (Writing—original draft), Laura Tilley (Writing—original draft), Robert Waterhouse (Writing—original draft), Bernhard Misof (Funding acquisition, Writing—original draft), and Thomas Borsch (Conceptualization, Funding acquisition, Supervision, Writing—original draft, Writing—review & editing)

Funding

German Federal Ministry of Research, Technology, and Space (BMFTR) under grant agreement no. 03LC2020A.

References cited

- Addink W, Theocharides S. 2024. The future of referencing specimens is near: Cite the digital specimen DOI. *Biodiversity Information Science and Standards* 8: e137534. <https://doi.org/10.3897/biss.8.137534>
- Allison MJ, Round JM, Bergman LC, Mirabzadeh A, Allen H. 2021. The effect of silica desiccation under different storage conditions on filter-immobilized environmental DNA. *BMC Research Notes* 14: 106. <https://doi.org/10.1186/s13104-021-05530-x>
- Angeles NAC, Catap ES. 2023. Challenges on the development of biodiversity biobanks: The living archives of biodiversity. *Biopreservation and Biobanking* 21: 5–13. <https://doi.org/10.1089/bio.2021.0127>
- Astrin JJ, Corrales C. 2023. Preface. Pages 5–6 in Corrales C, Astrin JJ eds., *Biodiversity Biobanking—A Handbook on Protocols and Practices*. Pensoft. <https://doi.org/10.3897/ab.e101876>
- Astrin JJ, Zhou X, Misof B. 2013. The importance of biobanking in molecular taxonomy, with proposed definitions for vouchers in a molecular context. *ZooKeys* 365: 67–70. <https://doi.org/10.3897/zookeys.365.5875>
- Bingham H, Doudin M, Weatherdon L, Despot-Belmonte K, Wetzel F. 2017. The biodiversity informatics landscape: Elements, connections and opportunities. *Research Ideas and Outcomes* 3: e14059. <https://doi.org/10.3897/rio.3.e14059>
- Buckner JC, Sanders RC, Faircloth BC, Chakrabarty P. 2021. The critical importance of vouchers in genomics. *Elife* 10: e68264. <https://doi.org/10.7554/eLife.68264>
- Carroll SR, Garba I, Figueroa-Rodríguez OL, Holbrook J, Lovett R. 2020. The CARE principles for indigenous data governance. *Data Science Journal* 19: 43. <https://doi.org/10.5334/dsj-2020-043>
- Convention on Biological Diversity. 2022. *Biodiversity Framework: Draft Decision Submitted by the President*. CBD Secretariat. Report No. CBD/COP/15/L.25. www.cbd.int/article/cop15-final-text-kunming-montreal-gbf-221222
- Corrales C, Astrin JJ, eds. 2023. *Biodiversity Biobanking—A Handbook on Protocols and Practices*. Pensoft Publishers. <https://doi.org/10.3897/ab.e101876>
- Corrales C, Luciano S, Astrin JJ. 2023. Biodiversity biobanks: A landscape analysis. *ARPHA Preprints* 4: e103105. <https://doi.org/10.3897/arphpreprints.e103105>
- Droege G, Barker K, Astrin JJ, Bartels P, Butler C. 2014. The Global Genome Biodiversity Network (GGBN) data portal. *Nucleic Acids Research* 42: D607–D612. <https://doi.org/10.1093/nar/gkt928>
- Droege G, Barker K, Seberg O, Coddington J, Benson E. 2016. The Global Genome Biodiversity Network (GGBN) data standard specification. *Database* 2016: baw125. <https://doi.org/10.1093/database/baw125>

- Eloe-Fadrosh EA, et al. 2024. A practical approach to using the genomic standards consortium MIXS reporting standard for comparative genomics and metagenomics. *Methods in Molecular Biology* 2802: 587–609. https://doi.org/10.1007/978-1-0716-3838-5_20
- Gilbert JA, Peixoto RS, Scholz AH, Dominguez Bello MG, Korsten L. 2025. Launching the IUCN Microbial Conservation Specialist Group as a global safeguard for microbial biodiversity. *Nature Microbiology* 10: 2359–2360. <https://doi.org/10.1038/s41564-025-02113-5>
- Global Biodata Coalition. 2025. *Working Cooperatively for Global Biodata Resource Sustainability: A Global Biodata Coalition White Paper*. Zenodo. <https://doi.org/10.5281/zenodo.14727224>
- Hewitt R, Watson P. 2013. Defining Biobank. *Biopreservation and Biobanking* 11: 309–315. <https://doi.org/10.1089/bio.2013.0042>
- Hoborn D, Baptiste B, Copas K, Guralnick R, Hahn A. 2019. Connecting data and expertise: A new alliance for biodiversity knowledge. *Biodiversity Data Journal* 7: e33679. <https://doi.org/10.3897/BDJ.7.e33679>
- Holtschek J, Dröge G, Güntsch A, Berendsohn WB. 2012. The ABCD of rich data access to natural history collections. *Plant Biosystems—An International Journal Dealing with all Aspects of Plant Biology* 146: 771–779.
- International Organization for Standardization. 2018. *Biotechnology. Biobanking. General Requirements for Biobanking*, 36. Geneva, Switzerland: ISO. Standard No.: ISO 20387:2018.
- International Society for Biological and Environmental Repositories. 2023. *ISBER Best Practices: Recommendations for Repositories*. 5th ed. www.isber.org/page/BPR
- Kilian N, Henning T, Plitzner P, Müller A, Güntsch A. 2015. Sample data processing in an additive and reproducible taxonomic workflow by using character data persistently linked to preserved individual specimens. *Database* 2015: bav094. <https://doi.org/10.1093/database/bav094>
- Lendemer J, Thiers B, Monfils AK, Zaspel J, Ellwood ER. 2020. The extended specimen network: A strategy to enhance US biodiversity collections, promote research and education. *Bioscience*: 70: 23–30. <https://doi.org/10.1093/biosci/biz140>
- Letsch H, Greve C, Hundsdoerfer AK, Irisarri I, Moore JM. 2025. Type genomics: A framework for integrating genomic data into biodiversity and taxonomic research. *Systematic Biology*: 74: 1029–1044. <https://doi.org/10.1093/sysbio/syaf040>
- Lissin A, Schober I, Witte JF, Lüken H, Podstawka A. 2025. StrainInfo—the central database for linked microbial strain identifiers. *Database* 2025: baaf059. <https://doi.org/10.1093/database/baaf059>
- Penev L, Groom Q, Casino A, Barov B. 2024. Uniting FAIR data through interlinked, machine-actionable infrastructures. *Research Ideas and Outcomes* 10: e126588. <https://doi.org/10.3897/rio.10.e126588>
- Penev L, Koureas D, Groom Q, Lanfear J, Agosti D. 2022. Biodiversity Community Integrated Knowledge Library (BiCIKL). *Research Ideas and Outcomes* 8: e81136. <https://doi.org/10.3897/rio.8.e81136>
- Ratnasingham S. 2024. BOLD v4: A centralized bioinformatics platform for DNA-based biodiversity data. Pages 403–441 in DeSalle R, ed. *DNA Barcoding: Methods and Protocols*. Springer US. https://doi.org/10.1007/978-1-0716-3581-0_26
- Reichel K, Pohjoismäki J, Astrin JJ, Böhne A, Bortoluzzi C. 2026. From permits to samples—addressing key challenges for high-quality reference genome generation in Europe. *Molecular Ecology Resources* 26: e70100. <https://doi.org/10.1111/1755-0998.70100>
- Renner SS, Scherz MD, Schoch CL, Gottschling M, Vences M. 2024. Improving the gold standard in NCBI GenBank and related databases: DNA sequences from type specimens and type strains. *Systematic Biology* 73: 486–494. <https://doi.org/10.1093/sysbio/syad068>
- Sánchez-Pacheco SJ, Crawford A, Pulido-Santacruz P. 2025. Colombia: Epic biodiversity, struggling genomics. *Acta Zoologica Lilloana* 69: 68.
- Schindel DE, Cook JA. 2018. The next generation of natural history collections. *PLoS Biology* 16: e2006125. <https://doi.org/10.1371/journal.pbio.2006125>
- Schulman L, Lahti K, Piirainen E, Heikkinen M, Raitio O. 2021. The Finnish biodiversity information facility as a best-practice model for biodiversity data infrastructures. *Scientific Data* 8: 137. <https://doi.org/10.1038/s41597-021-00919-6>
- Sigward JD, Schleuning M, Brandt A, Pfenninger M, Saedi H. 2025. Collectomics—towards a new framework to integrate museum collections to address global challenges. *Natural History Collections and Museomics* 2: 1–20. <https://doi.org/10.3897/nhcm.2.148855>
- Streit DP, Jr, Zhang T, Paredes E, eds. 2024. *Cryobiology for South American Neotropical Fish Species*. CRC Press. <https://doi.org/10.1201/9781003269441>
- Wieczorek J, Bloom D, Guralnick R, Blum S, Döring M. 2012. Darwin core: An evolving community-developed biodiversity data standard. *PLoS ONE*: 7: e29715. <https://doi.org/10.1371/journal.pone.0029715>
- Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M. 2016. The FAIR guiding principles for scientific data management and stewardship. *Scientific Data* 3: 160018. <https://doi.org/10.1038/sdata.2016.18>
- Yeates DK, Zwick A, Mikheyev AS. 2016. Museums are biobanks: Unlocking the genetic potential of the three billion specimens in the world's biological collections. *Current Opinion in Insect Science* 18: 83–88. <https://doi.org/10.1016/j.cois.2016.09.009>
- Yilmaz P, Kottmann R, Field D, Knight R, Cole JR. 2011. Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIXS) specifications. *Nature Biotechnology* 29: 415–420. <https://doi.org/10.1038/nbt.1823>