# Approved GBIF Strategic Plan 2017-2021

## Background and Context

### *GBIF progress and challenges*

GBIF has a unique position among international organizations working within the field of biodiversity informatics.  It is the only body supported by national governments with a mandate for mobilization and management of data on all taxonomic groups.

Since it was set up in 2001, GBIF's Participant community has grown (as of July 2015) to include 54 countries and 39 organizations or economies, all signatories to the GBIF Memorandum of Understanding (MoU).  This global network, supported by a securely-established secretariat in Copenhagen, has mobilized more than half a billion data records and serves as a resource for a growing body of peer-reviewed research and policy applications, with the volume of use growing each year.

Hundreds of institutions have published their valuable data through the GBIF network, for free and open use worldwide. The volume of data and the number of contributing data publishers continues to grow, thanks to collaborative efforts by GBIF's community of Participant nodes promoting best practices and sharing skills to help overcome barriers to data sharing. In the last fifteen years, GBIF has been central to the development and adoption of workable and scalable technologies, as well as the incentives, for mobilizing, organizing and accessing very large volumes of biodiversity data.

GBIF has completed its early development phases and is now an operational infrastructure, recognized as the global aggregator for species occurrence data and as a leader in development of globally connected solutions for biodiversity information. GBIF's role is recognized in the context of the Convention on Biological Diversity (CBD), Intergovernmental Platform on Biodiversity and Ecosystem Services (IPBES), Group on Earth Observations Biodiversity Information Network (GEO BON) and other international activities to support biodiversity assessment and planning.

GBIF's achievements to date have been impressive (see Box 1 on progress relating to the 2012-2015 Strategic Plan). However as the network approaches its fourth funding period, from 2017 to 2021, key remaining challenges need to be addressed within a coherent strategy based on updated priorities and well-defined goals, understood by all GBIF's stakeholders. This document sets out such a strategy, aimed at putting GBIF on a secure footing and serving the needs of the global community into the next decade.

Among the critical challenges addressed by this Strategic Plan are:

- **WIDENING AND STRENGTHENING PARTICIPATION**: Many countries in all regions remain unconnected with GBIF and progress in establishing strong national capacity to mobilize and access relevant data varies significantly among existing participants. As a result, funding remains a challenge, with costs shared among a relatively small number of countries and disproportionate impact when one or more of these is unable to contribute at agreed levels. More conspicuously, data coverage is highly variable between regions and even between adjacent countries.
- **BUILDING TRUST IN DATA PRODUCTS:** Improvements are required in the quality and fitness-for-use of aggregated data and metadata within the network. More work is needed to ensure that all data are sufficiently documented and catalogued in ways that assist users in filtering according to their needs, and greater precision and accuracy is necessary particularly in representing taxonomy. Mechanisms and

incentives are needed to engage expert communities in validation and correction of this global data resource.

- **FILLING DATA GAPS AND BROADENING THE EVIDENCE BASE:** Opportunities exist to make significant advances both in the completeness and coverage of GBIF data and of the richness of available information. GBIF must identify and understand where existing data are inadequate to meet user needs and must prioritize effective responses to address these issues. All relevant sources of data must be incorporated, including sample-based data sets, ecogenomics and other molecular research, remote-sensing, literature records, local and regional checklists, and expert knowledge. These resources should be used to establish GBIF not only as a source of occurrence information but as an effective tool to discover and access data on species abundance and community composition, and related genetic data.
- **SCALING UP INFRASTRUCTURE:** Integrating growing volumes of data will bring new challenges in efficient storage, management, presentation and access of these data. GBIF will face related challenges as it engages with more countries, organizations and institutions and as its services become more mission-critical for many stakeholders. During 2017-2021, GBIF must accordingly continue to innovate and to review all processes to ensure smooth future growth.

## GBIF's place in the landscape

Information on the world's biodiversity is recognized to be an essential requirement to support research into species and ecosystems, to underpin conservation and sustainability goals and to support a wide range of spatial assessment and management activity. Delivering information systems to support these needs is a complex requirement, given the multidimensional complexity of biodiversity itself and the limited nature of all direct observations and measurements of biodiversity patterns.

As outlined in the *Global Biodiversity Informatics Outlook* (GBIO – www.biodiversityinformatics.org), coordinated activity is required across multiple levels of cooperation and thousands of institutions, organizations and researchers.  The GBIO presents a multi-tiered model, with adoption of a global approach to free and open sharing of data (CULTURE layer), broad-scale engagement of data holding institutions to share data in open and compatible formats (DATA layer), comprehensive services to catalogue and organize these data (EVIDENCE layer), and large-scale modelling activities to provide the best possible picture of actual distributions, community composition and ecosystem structure (UNDERSTANDING layer). The global community has the opportunity to address the challenges of each of these layers in parallel, with increased efficiency and shared benefits from advances in any aspect.

Coordinated activity across these layers will deliver the information tools needed to deliver an effective GEO Biodiversity Observation Network (GEO BON, the biodiversity component within the Global Earth Observation System of Systems (GEOSS)), and thereby offer data products open to community validation in support of the scientific assessments of the Intergovernmental Platform on Biodiversity and Ecosystem Services (IPBES), and addressing the needs of governments and MEAs including the Convention on Biological Diversity (CBD).

GBIF's mission relates to coordinated global effort in the CULTURE and DATA layers of the GBIO framework and to delivery of key components of the EVIDENCE layer. GBIF aims to deliver the most complete and best-organized integration and organization possible of all data items representing evidence of the occurrence of any species in time and space. In doing so, GBIF can serve both as a global resource for spatial modeling of biodiversity

patterns and as a tool for taxonomists, collection managers and others requiring information on and access to the world's natural history collections.

## Response

### *Strategic framework*

This Strategic Plan identifies a set of five inter-related priorities for GBIF activity during 2017-2021.  The overall goal is to increase the relevance and value of GBIF as a mechanism for delivering comprehensive information on the recorded occurrence of species around the world. Achieving this goal depends on expanding engagement into all countries and regions and on maximizing the value and benefits provided by GBIF for all key stakeholders, including data publishers, national and organizational participants, researchers, and governments and intergovernmental frameworks.

The first and over-riding priority is to **deliver relevant data** – to provide access to the data resources and the aggregated data products that are needed by biodiversity researchers of all types and by policymakers to support conservation and sustainable development.  This priority requires closer engagement with user communities both to establish present and future requirements and to collaborate in curating data to the highest possible standard.

The second priority – to **improve data quality** – and the third priority – to **fill data gaps** – support the first.  The greatest weaknesses in the data delivered through the GBIF network arise from the challenge of maintaining clarity and precision while integrating data from thousands of different institutions and projects and from the uneven nature of content mobilization with regard to taxa, time and space.  GBIF must focus on enhancing the clarity and completeness of published data and metadata and on improving handling of aggregated data to ensure that overall quality is improved and that researchers are able to assess the fitness of data for their use.  At the same time, it is important to assess and prioritize the various gaps and variations within GBIF data and to work with the global community to address these weaknesses.

The fourth priority – to **enhance biodiversity information infrastructure** – supports the previous priorities.  GBIF must enhance its cooperation with other biodiversity informatics activities to increase interoperability and interconnection between GBIF species occurrence data and the best available taxonomic information, sequences, species traits, literature, etc.  Progress in delivery of the GBIO framework will reinforce GBIF's work in all areas by enabling integration of complementary data and use of species-level information to validate, clean and augment GBIF data.

The final priority – to **empower global network** – underpins everything else.  GBIF is not simply a partnership focused on delivering a global data resource.  The goal must be to see the establishment of effective national biodiversity information facilities (BIFs) as activities and networks able to connect with all groups holding or requiring information on biodiversity and for these national BIFs to deliver value at the national level. Progress in any country will also deliver benefits to others by increasing available data and the comprehensiveness of the data resource and by contributing to the capacity and expertise of the global network.

These five priorities together represent a challenging but significant path forward for GBIF to increase its relevance for all stakeholders. Implementing this strategy will require coordinates activity throughout the GBIF network and engagement and collaboration with all other stakeholders in biodiversity information.

### *Priorities*

#### Deliver relevant data

*Ensure that GBIF delivers data in the form and completeness required to meet the highest-priority needs of science and, through science, society.*

a) Engage with expert communities to manage data to the highest quality possible

      i)   Engage taxonomic societies, professional researchers and amateur experts as partners in assessing and improving GBIF-mediated data

      ii)  Define criteria for determining data relevance (elements, completeness, etc.) for key applications and domains

      iii) Develop mechanisms for community curation of taxonomic, geographic or thematic sectors within GBIF-mediated data

b) Deliver well-organized and validated data to support key applications

      i)   Develop mechanisms to monitor and prioritize data needs from GBIF audiences

      ii)  Support supra-national research, conservation and sustainable use, particularly through CBD, IPBES and GEO BON

      iii) Provide data relevant to understanding and responding to the impacts of climate change on biodiversity

      iv) Ensure that GBIF serves as a virtual natural history collection to support taxonomic research

      v)   Assess and develop role of GBIF in support of Nagoya Protocol

### *Implementation*

- Secretariat: Coordinate working groups and other processes to refine requirements from user communities
- Secretariat: Articulate vision and strategy for continuing development of GBIF data products
- Participants: Establish working groups to capture requirements from additional user communities
- Participants: Publicise and advocate for GBIF with taxonomic societies, national agencies and international bodies

## Improve data quality

*Ensure that all data within the GBIF network are of the highest-possible quality and associated with clear indicators enabling users to assess their origin, relevance and usefulness for any application.*

a) Enhance automated data validation

      i)   Validate metadata for completeness and interpretability

      ii)  Validate individual data records

      iii) Validate data sets as a whole (patterns, etc.)

b) Implement tools for expert curation

      i)   Integrate and promote tools to annotate individual records and batches of records

      ii)  Enable publication of cleaned data ("reference data sets")

      iii) Improve communication with and support for data publishers to address issues

c) Provide clear quality indicators for all data

      i)   Improve metadata for all data sets explaining methods, provenance, transformations, etc. ("data stories")

      ii)  Provide metrics and filters for download of data subsets matching user criteria

      iii) Work with expert communities to develop reference data sets and/or filters for data suitable for key applications

### *Implementation*

- Secretariat: Implement extensible tools for automated data validation
- Secretariat: Incorporate requirements from user communities in data validation
- Secretariat: Develop mechanisms and tools for community data curation

- Secretariat: Incorporate quality indicators in user interfaces and data downloads
- Participants: Assist data publishers in improving data quality and responding to issues detected
- Participants: Develop extensions and enhancements to automated validation tools
- Participants: Promote use of GBIF as a repository for reference data sets

## Fill data gaps

*Prioritize and promote mobilization of new data resources which combine with existing resources to maximize the coverage, completeness and resolution of GBIF data, particularly with respect to taxonomy, geography and time.*

a) Expand checklists to cover all taxonomic groups
   i) Coordinate and deliver comprehensive solutions in partnership with Catalogue of Life and other partners
   ii) Target mobilization of checklists for remaining taxonomic groups
   iii) Mobilize national and regional species lists
b) Identify and prioritize gaps in spatial and temporal data
   i) Develop models and criteria to assess completeness and adequacy of existing data
   ii) Maintain "gap map" highlighting priority regions, taxa and time periods to fill gaps
c) Engage institutions and researchers with complementary data
   i) Develop mechanisms to integrate other classes of data providing information on species occurrence (checklists, expert maps, sample-based data, remote sensing, genomics, etc.)
   ii) Ensure that tools and documentation support mobilization of data streams from all relevant research areas and observation systems
   iii) Promote best practices for GBIF Nodes to identify, engage and support all national holders of relevant data
   iv) Develop mechanisms to support data sharing from researchers and institutions in countries which are not yet GBIF Participants
   v) Seek supplementary funding to support large-scale content mobilization

### *Implementation*

- Secretariat: Engage with other biodiversity informatics organisations to deliver a comprehensive checklist of the world's species
- Secretariat: Implement tools and interfaces for "gap map"
- Secretariat: Publicise priority needs to address gaps in mobilised data
- Secretariat: Coordinate engagement with potential data publishers
- Secretariat: Simplify and extend tools and processes for mobilising relevant data
- Participants: Mobilise national and regional species lists
- Participants: Engage holders of biodiversity data to publish through GBIF
- Participants: Secure supplementary funding for content mobulisation

## Enhance biodiversity information infrastructure

*Provide leadership, expertise and tools to support the integration of all biodiversity information as an interconnected digital knowledgebase.*

a) Coordinate vision and strengthen partnerships with major biodiversity informatics initiatives

      i)   Develop and promote shared architecture to enable all biodiversity knowledge to be integrated and managed as a linked digital resource ("biodiversity knowledge graph")

      ii)  Agree plans and secure funding to implement shared architecture in partnership with other major biodiversity informatics organizations, particularly around taxonomic and nomenclatural data

      iii) Seek funding with partner initiatives to deliver a stable and efficient network of interoperable infrastructures connecting and maintaining all biodiversity knowledge

b)  Promote standardization and common mechanisms for exchange of biodiversity data

      i)   Promote development and adoption of a comprehensive domain model, standards and vocabularies for biodiversity information through TDWG and other partnerships

      ii)  Provide leadership in development and adoption of standardized tools, data formats and vocabularies to enable mobilization and use of all biodiversity knowledge

c)  Provide stable and persistent data infrastructure to support research

      i)   Implement and promote models and tools for persistent sharing and use of open data, including publishing tools, repositories, data licensing models, persistent identification, citation and tracking tools (the components which form the "Culture" layer in the GBIO document)

      ii)  Contribute to wider initiatives, particularly in the context of the Research Data Alliance, to standardize management and use of open data in research

*Implementation*

- Secretariat: Develop complementary work plans with major biodiversity informatics initiatives
- Secretariat: Ensure all GBIF-compatible tools, interfaces and standards are clearly and adequately documented
- Participants: Publicise and advocate for open access to research data and for use of GBIF-compatible standards
- Participants: Engage with national and regional initiatives for long-term preservation and management of research data
- Participants: Secure funding for development of biodiversity informatics tools and infrastructure

**Empower global network**

*Ensure that governments, researchers and users are equipped and supported to share, improve and use data through the GBIF network, regardless of geography, language or institutional affiliation.*

a)  Remove barriers to participation

      i)   Engage with government stakeholders in all regions to expand national participation in GBIF

      ii)  Ensure that data holders everywhere have access to tools and support to publish and use data through GBIF

      iii) Evaluate and respond to needs for delivery of data and data summaries in additional formats and through new technologies

b)  Increase benefits associated with publishing biodiversity data

      i)   Promote best practices for citation and acknowledgment of data publishers

    ii)  Report on use of data accessed through GBIF network

    iii)  Enable data publishers and users to collaborate in correcting and improving data

c) Address capacity needs

    i)  Maintain current and complete guidance materials to support all types of data holder and data publisher

    ii)  Support capacity exchange and reuse of tools and expertise to accelerate establishment of national biodiversity information facilities

    iii)  Partner with other organizations, institutions and agencies to maximize benefits from investments in capacity development in all regions

    iv)  Promote inclusion of biodiversity informatics training as part of relevant university and workplace education

    v)  Seek supplementary funding to support development of national biodiversity information facilities

*Implementation*

- Secretariat: Coordinate engagement with governments and organisations not yet involved in GBIF

- Secretariat: Monitor and report on use of data through GBIF network

- Secretariat: Coordinate development and translation of information and training materials

- Secretariat: Coordinate development and deployment of tools to support data publishing and data access in all countries and regions

- Participants: Promote GBIF participation with governments and organisations not yet involved in GBIF

- Participants: Develop or translate information and training materials for use within the GBIF network

- Participants: Provide support and mentorship for new and developing GBIF Nodes