

## GBIF Implementation Plan 2017-2021 and Annual Work Programme 2018

### Overview

This document serves both as the 2018 revision of the GBIF Implementation Plan for the period 2017-2021 and also as the GBIF Annual Work Programme 2018. Both elements are based upon the [GBIF Strategic Plan for 2017-2021](#).

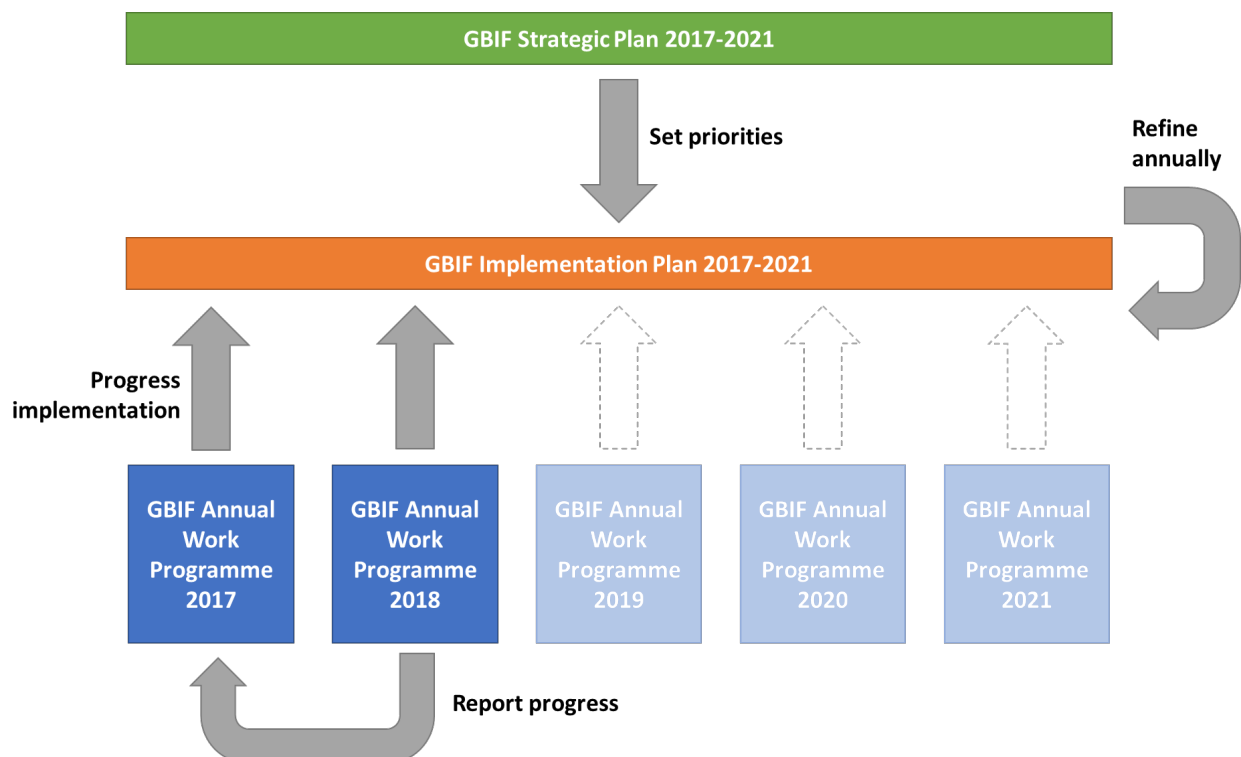
The relationship between these documents is represented by the following graphic.

The Strategic Plan (agreed in 2015 at GB22) established five **priorities** for GBIF's work.

The Implementation Plan presented here in this document identifies 22 **activities** which GBIF must undertake to address these priorities during the five-year period (and a set of specific **tasks** under each activity).

Each year, GBIF reviews progress against these activities and available resources (including Secretariat staff and other uses of core and supplementary funding, as well as resources committed by GBIF Participants). Using this information, an Annual Work Programme is defined for the following year. Each Annual Work Programme identifies a prioritised set of tasks for work during the year. Annual Work Programmes in the second and subsequent years also report on progress against the plan.

The Implementation Plan is reviewed and refined each year.



The Implementation Plan is presented here with updates reflecting progress during 2017 and planned activities for 2018 (the 2018 Annual Work Programme).

The Annual Work Programme 2018 is summarized below. The revised Implementation Plan begins on page 8.

## GBIF Annual Work Programme 2018

Information on activities planned for 2018 is included under each of the Activity sections included in the Implementation Plan, beginning on page 13.

For convenience, this information has also been collated on page 4 as an overview of the items for which approval is sought from the Governing Board at GB24.

### Financial situation

At GB23, a budget for 2017 was agreed based on Participant expectations regarding their annual contributions. The expectation at that time was that income for the year would be much lower than the agreed target, at around €2,635,626. As a result, a relatively limited work programme was agreed for 2017

In fact, although even by August 2017, only around 81% of the budgeted contributions for 2017 have been received, GBIF's financial situation has improved significantly. There are several reasons for this:

- Canada, Poland and Switzerland have become Voting Participants and together contribute around €450,000 to the budget.
- The UK has contributed around €110,000 more than predicted at GB23.
- Argentina and Mauretania made payments totaling €26,500 for contributions delayed from previous years.
- The European Commission gave approval for the GBIF Secretariat to offset a greater number of staff hours using funds from the BID programme, representing an additional €50,000 in income for 2017.

Additionally, Sweden, the US and SBNO have made supplementary contributions totaling €77,000 to cover activities which were originally excluded from the 2017 Work Programme owing to lack of funds (Ebbe Nielsen Challenge, Young Researchers Awards and Global Nodes Meeting).

As a result of these factors, and careful expenditure in a number of areas, GBIF expects to end 2017 with a balance of around €1,125,000, around €775,000 higher than projected at GB23.

As a result of this, we propose to expand expenditure in a number of key areas during 2018. It remains prudent, however, to retain a higher balance than the agreed minimum of €250,000, owing to continuing uncertainties around future contributions from several participants.

### Secretariat staffing

Staff changes in recent years have enabled the Secretariat to maintain a steady level of service with lower salary costs than in the past. The Secretariat is now structured as four teams, as follows:

- Participation and Engagement (six staff, plus CAS visiting scientist)<sup>1</sup>
  - Direct support and guidance for GBIF participants and Nodes
  - Capacity enhancement activities
  - Documentation and communication
- Data Products (four staff)
  - Technical support and helpdesk for data publishers and users
  - Data quality and fitness-for-use

---

<sup>1</sup> The Chinese Academy of Science has funded Maofang Luo spending a year working with the Participation and Engagement team

- Informatics (six or seven staff)<sup>2</sup>
  - Software development
  - Operation of GBIF IT infrastructure
  - Standards
- Administration (five staff, plus student support)<sup>3</sup>
  - Support for Governing Board and Committee activities
  - Administration of calls for proposals and contract processes
  - Organisation of meetings and travel
  - Financial management

The leader of each of these teams reports directly to the Executive Secretary.

## Planned expenditure

Funding is reserved in the Work Programme 2018 for the following activities:

- **Regional Nodes Meetings: €60,000** (Revised from €30,000 following GB24)

Funds to provide partial support for organisation of regional nodes meetings. See: *Activity 1f - Plan implementation*.

- **Global Biodiversity Informatics Outlook follow-on workshop: €50,000**

Funds (requiring co-funding from other sources) to organise a follow-on workshop to the 2012 Global Biodiversity Informatics Conference (GBIC) event. GBIC led to the development of the Global Biodiversity Informatics Outlook (GBIO), which provided a high-level vision for major required components in a globally interconnected biodiversity data infrastructure. The goal of the new workshop will be to develop agreement, including a governance strategy and funding options, for a project office to assist biodiversity data infrastructures and projects in contributing to a joined-up solution. See: *Activity 1f - Plan implementation*.

- **Capacity Enhancement Support Programme (CESP) call: €80,000**

Funds for the 2018 call for projects to support the development and activity of GBIF Participant nodes. See: *Activity 1g - Coordinate resources*.

- **Deliver Names Infrastructure: €80,000**

Funds to support the GBIF component (with partial funding from NLBIF) of the second year of joint development with Catalogue of Life, funded by NLBIF. See: *Activity 2b - Deliver names infrastructure*.

- **Five-year Review of GBIF Progress: €30,000**

Funds to support an external review of GBIF operations – the last such reviews took place more than five years ago. See: *Activity 5d - Assess impact*.

---

<sup>2</sup> In recent years, the Informatics this team has included six developers. This has allowed for steady maintenance of the systems but has left a significant legacy of deferred or delayed work on data quality, data presentation, and tools for data publishers, users and nodes. Accordingly, we plan to expand this team by one member.

<sup>3</sup> The Administration team lost a staff member during the last year. This position has not been filled, but the Secretariat is making use of a part-time student worker to supplement capacity.

The complete budget proposal for 2018 may be summarised as follows (revised following GB24, see GBIF Draft 2017 Budget for more detail):

Opening balance	€1,126K
Participant contributions <sup>4</sup>	€3,000K
Secretariat supplementary income	€140K
<b>Total available funds</b>	<b>€4,266K</b>
Secretariat salaries	€2,401K
Governance costs	€101K
Operating costs <sup>5</sup>	€553K
Work programme expenditure	€300K
<b>Total expenditure</b>	<b>€3,345K</b>
<b>Closing balance</b>	<b>€921K</b>

## Overview of 2018 Work Items

This following is a summary of all 2018 Work Items proposed in the 2018 update of the *Implementation Plan* below. It is intended to summarise the work items for which approval is sought from the Governing Board at GB24. Accordingly, it does not include the additional commitments made and reported by Participants. These are described in the Activity sections within the body of the document below.

### Activity 1a - Focus on people

- Enable users to associate ORCID IDs with their GBIF.org accounts and to identify with which datasets they are associated. This will enable individuals to track citations through both the ORCID and DOI systems.
- Allow users to link their information to events and projects and declare their areas of expertise (e.g. Darwin Core, IPT, taxonomic groups) within GBIF activities.
- Retire the GBIF community site and replace with a discussion forum integrated into GBIF.org and with appropriate use of alternative platforms (e.g. Facebook, LinkedIn).
- Review current use of digital badge system in BID, align with skills training offered by partner networks, and expand to more generic application in capacity programmes.

### Activity 1b - Strengthen skills

- Adapt current BID data mobilization training modules into a more generic curriculum able to be picked up in future programmes by external trainers (including a proposed BIFA workshop in 2018), and including a larger element of self-instruction.
- Extend BID training modules to incorporate key skills around data application and use, with first demonstration at the BID Africa workshop in February 2018
- Complete set of explainer texts in multiple languages, as well as video explainers on key aspects of GBIF participation

<sup>4</sup> Including €23,000 from Copenhagen University towards Secretariat IT costs.

<sup>5</sup> Including banking costs.

- Develop BID helpdesk and mentoring processes for more general application across the network, including a mechanism to assign specific ‘focal point’ mentors to all new nodes and projects

#### **Activity 1c - Equip Participant nodes**

- Review and further development of nodes guidance materials, including additional material on engaging citizen science, private sector, research and policy communities
- Retire the unmaintained community site of GBIF and replace with a more modern solution. A review of the functionality desired and needed will be conducted with the community, and one or more solutions put in place, including forums for direct discussion and helpdesk on GBIF.org and promotion and use of already commonly used tools such as Facebook groups.
- Review use of national self-assessment tool with a view to possible enhancements and guidance on future use
- Continue webinars and adapt based on feedback after first year, expand to wider audience including e.g. publisher and user communities
- Development of a role for members of the GBIF network and interested researchers to serve as ‘GBIF science ambassadors’ to promote understanding and use of GBIF data and services

#### **Activity 1d - Equip data publishers**

- Review of the IPT product and conduct a consultation with community on necessary changes beyond maintenance and bug fixing.
- Maintain cloud based IPT installations (updates, backups, helpdesk) and deploy further instances as required.

#### **Activity 1e - Expand national participation**

- Further development of multilingual content on GBIF.org to support wider participation
- Annually-updated review of global participation based on regional and sub-regional representation, to support engagement strategies to be led by existing GBIF regions.

#### **Activity 1f - Plan implementation**

- Develop plans and coordinate funding for a workshop to develop a shared plan for a project office for coordinating biodiversity informatics activity and to develop draft roadmaps for collaborative delivery of some of the GBIO components.
- Organise regional Nodes meetings - €60,000 (revised following GB24) reserved in budget to support organisation of these meetings in regions lacking necessary funds

#### **Activity 1g - Coordinate resources**

- Build on BID and BIFA programmes to seek additional funding for aligned capacity enhancement activity
- Operate the 2018 Capacity Enhancement Support Programme call with an allocated budget of €80,000.
- Operate the 2018 Ebbe Nielsen Challenge and Young Researchers Awards, with a funding of €30,000 from the Governance allocation within the 2018 Budget.

#### **Activity 2a - Modernize data standards**

- GBIF will participate in the joint meeting of TDWG and SPNHC (Society for the Preservation of Natural History Collections) to be held in Dunedin, New Zealand (25 August – 2 September 2018).
- Implementation of recommendations from TDWG Biodiversity Data Quality Interest Group.

**Activity 2b - Deliver names infrastructure**

- Deliver public prototype and solicit community feedback on comprehensive nomenclature and checklist using clearing-house model.

**Activity 2c - Catalogue collections**

- Develop joint action plan with GRSciColl for joint delivery and maintenance of the catalogue of collections.

**Activity 3a - Identify priority gaps**

- Synthesize outcomes of recent activities (including Ebbe Nielsen Challenge 2016) to visualize and measure data gaps and to offer data mobilization targets
- Encourage additional research communities to provide feedback on possible improvements to GBIF data management and presentation to meet their needs – this may be delivered following the fitness-for-use group model.

**Activity 3b - Expand data streams**

- Work with BHL to incorporate literature-based occurrence data.
- Work with BOLD and others to incorporate sequence-based occurrence data.

**Activity 3c - Engage data holders**

- Engagement with data holder communities relies on effective national scale coordination. The GBIF Secretariat will revise guidance for Participant Nodes on engagement with citizen science groups, private sector environmental impact consultants, etc.
- Material will be added to GBIF.org to summarize guidance and best practice relating to data mobilization from environmental impact assessment. The GBIF Secretariat will continue to engage with the International Association of Impact Assessment, World Bank, International Finance Corporation, World Economic Forum, etc. to promote open data publication through GBIF.

**Activity 3d - Rescue datasets**

- Establish a repository and processes for hosting, curating, publishing and endorsing datasets suggested for rescue.

**Activity 3e - Liaise with journals**

- Develop communications with journals around GBIF DOI-based data management, data downloads and data citation policy. Encourage journals, editors and authors to practise responsible data citation.

**Activity 4a - Ensure data persistence**

- Establish ongoing detection for future orphan data sets and processes to ensure persistent hosting and management.

**Activity 4b - Assess data quality**

- Engage with TDWG Data Quality group to define and document the rules and flagging required to assess data quality.
- Engage with the ALA team specifically, and others where possible, to ensure that systems treat data consistently, and use the same software libraries. The GBIF / ALA ingestion processes should become the reference implementation for a TDWG standard.
- Accumulate and present dataset quality statement metrics for the dataset, publisher and national level on GBIF.org to help direct all parties to address data issues and prioritise effort.
- Evaluate options to incorporate expert review as an element in GBIF data management.

**Activity 4c - Enable data curation**

- Connect more systems to receive direct annotations from GBIF.org users. Symbiota and Specify communities have expressed interest.

**Activity 5a - Engage academia**

- A BID workshop will be held in Africa during February 2018 to assist projects with engaging users in applied use of mobilized data
- New brochures will be produced to highlight the relevance of GBIF for researchers (multiple languages).

**Activity 5b - Document needs**

- Address recommendations from 2015-2016 data fitness for use groups.

**Activity 5c - Support biodiversity assessment**

- Explore possible enhancements to GBIF data management to optimize access to data in formats which are appropriate for constructing species distribution and population abundance EBVs.
- Further discussion is required to integrate GBIF data into IUCN Red List processes.

**Activity 5d - Assess impact**

- Extend collection of GBIF use cases to include grey literature and other non-research applications of data
- Commission value proposition study – extra funding will be required.
- Contract external review of GBIF operations and deliverables - €30,000 allocated.

## GBIF Implementation Plan 2017-2021

The GBIF **Strategic Plan** 2017-2021 presents five broad overlapping priorities for the GBIF network. Each of these priorities sets a direction and guides activities for the coming period. This document here presents an **Implementation Plan** to enable the GBIF community as a whole to make significant progress in each priority area. It establishes the framework for GBIF's **Annual Work Programmes** over the period of the Strategic Plan.

The following notes provide context for the activities outlined here.

### Delivery

Some activities in this plan will be led or coordinated by GBIF Secretariat staff, using core or supplementary funds. Other activities may be led or coordinated by GBIF Participants with particular interest or expertise in the areas in question.

Combining energies across the whole network within the framework of this plan will enable GBIF to progress the set of activities identified in this document more rapidly. Each Annual Work Programme in the period 2017 to 2021 will build on progress in past years and will focus available resources on remaining priorities identified in this document.

### Structure of GBIF

GBIF operates on three primary levels:

- **Global** – GBIF is an international activity and works at the global level to increase standardization and adoption of best practices, to facilitate sharing of expertise and resources and to offer an integrated global dataset and associated informatics infrastructure.
- **Participant** – GBIF Participants, including the nodes that they establish, have primary responsibility for engaging stakeholder communities, mobilizing and curating data, providing support and training, and ensuring that national researchers and agencies gain maximum benefit from GBIF. GBIF regional networks provide a framework for national and organizational stakeholders to collaborate more closely, particularly when addressing challenges and opportunities they share.
- **Data Publisher** – Data-holding institutions, agencies and individual researchers are the foundations on which GBIF depends. GBIF has a responsibility to ensure that data publishers have access to the tools and support that they need and that they receive recognition for their work and contributions.

This Implementation Plan seeks to address needs at all three of these levels while aiming to reinforce the value of GBIF for stakeholders at each level. It is highly desirable to increase coordination and feedback between these levels. The 2018 version of the document includes reports from a few GBIF Participants on their 2017 activities and 2018 plans. It would be most beneficial to receive inputs from a wider range of Participants to give a clear picture of everything occurring in the network and to assist the Secretariat and all Participants with better alignment and reuse.

A particular requirement is for the GBIF network to maintain and coordinate software to support each of these levels and to ensure that the publication, integration, management and use of data are as efficient as possible.

At the global level, GBIF focuses on delivery of [GBIF.org](http://GBIF.org) as a software platform to support both human users and tools (via web services). Many of GBIF's goals depend on being able to deliver a fully integrated view of all data from all sources in the best possible forms.

At the data publisher level, GBIF maintains the Integrated Publishing Toolkit (IPT) as its default tool for sharing GBIF-compatible data. Ongoing enhancements are required to



increase both the usability of this tool and the value that its services provide to data publishers. At the same time, GBIF seeks to support GBIF-compatible data publishing via other tools such as collection management systems and citizen science platforms.

At the Participant level, several efforts have been made to deliver generic portal tools to assist node managers in their work and to deliver value at the national level. The most significant roles for such a platform would be to:

- Enable national- or regional-level integration and curation of datasets
- Enhance data records to reflect national administrative units and national species lists
- Provide tools that meet the requirements of national or regional researchers, policymakers and the general public

Over the last few years, a growing number of GBIF Participants have collaborated to leverage Australia's investment in the *Atlas of Living Australia* (ALA) and to customize these tools to deliver their own portals. The Implementation Plan responds to this with a focus on increasing community development and support to deliver a reusable and sustainable portal toolkit based on the ALA tools. This approach not only benefits Participants already committed to using these tools, but also provides a solution for rapid adoption by other Participants at national, regional, or even thematic levels. In addition, Belgium has indicated an interest in exploring the potential for an alternative lightweight solution based on repository tools such as [CKAN](#).

The vision for the technical aspects underlying this Implementation Plan is that the components developed and offered by the GBIF network at each scale should combine to deliver a robust solution to support GBIF data management and that together they should continue to evolve to support a leading international data infrastructure. One aspect of this will be to ensure that all users can share and reuse common components such as data validation tests and that GBIF promotes and fosters wider activity around open projects.

GBIF is more than just a technical network, and this Implementation Plan also addresses the capacity enhancement, networking, communication and reporting activities necessary to advance GBIF's work in these areas. Since it is impossible for GBIF to support all aspects of its network using core funds, this plan focuses on resources which will support the work of GBIF Participants and data holders everywhere, coordinating the skills, expertise and investment of the whole network to maximize benefits to all stakeholders.

### **Timeline**

None of the priorities, or the more specific goals identified for each priority, can be achieved in a single year; each will require coordinated effort across multiple annual work programmes. This document describes the set of activities currently recognized as important to deliver the Strategic Plan over the five-year period. It will be revised and expanded as required each year to address changes in the landscape in which GBIF operates.

This Implementation Plan was presented for approval by the Governing Board at GB23. It was accompanied by the Annual Work Programme for 2017. Both components were approved at the meeting.

Additional commitments from GBIF Participants or others to contribute or deliver particular elements will be welcomed throughout the five-year implementation period.

### **Resources**

GBIF has a range of resources for use in delivering its programme of activity, including core funding from annual Participant contributions, supplementary funds (including [BID](#) and [BIFA](#) and funds received for activities within other projects), and investments by GBIF Participants.

These resources combine as follows:

- **Core Funding**
  - Supports GBIF Governance structures (Governing Board, committees)
  - Supports staffing and operation of GBIF Secretariat to coordinate global activity, including:
    - Delivery of global-level IT infrastructure (GBIF.org)
    - Coordination of delivery of software components to support Participants and Data Publishers
    - Coordination of capacity enhancement activity and information materials for Participants, Data Publishers and Users
    - Coordination of supplementary-funded activity and Participant investments in shared GBIF activity
    - Monitoring and reporting usage and benefits
    - Administrative support for GBIF activities
    - Communication, outreach and partnership activities
  - Limited funds to support workshops, capacity enhancement and developments to expand or enhance GBIF
- **Supplementary Funding**
  - Additional funds to support workshops, capacity enhancement and developments to expand or enhance GBIF
  - Funding to support Participants and Data Publishers with content mobilisation or development of national Biodiversity Information Facilities
- **Participant Investments**
  - Operating GBIF nodes
  - Engaging with and supporting data publishers and users
  - Curating data from Data Publishers
  - Additional funds or staffing to support workshops, capacity enhancement and developments to expand or enhance GBIF

In past years, Annual Work Programmes have been developed to reflect the use of Core Funding, but this Implementation Plan is intended to offer a framework for GBIF Participants and other stakeholders to take leadership in delivering some of the identified activities and tasks. Where possible, Participants are encouraged to seek funds or allocate staff time to enable progress in areas beyond those for which Core Funding is adequate.

All such commitments reported by GBIF participants have been included within the 2017 Annual Work Programme and will be reported as part of GBIF's work during the year. Additional Participant commitments will be most welcome at any stage in the implementation of the 2017 Annual Work Programme and for inclusion in subsequent Annual Work Programmes. Examples of such commitments include (but are not limited to):

- Organization of workshops to develop required new standards or best practice recommendations
- Capacity enhancement and training at regional or global levels, in particular for workshops and projects in regions which are not addressed by current Supplementary Funding sources
- Development of tools or software components to advance GBIF's work
- Representation of GBIF in international fora
- Recruitment of additional GBIF Participants

Please advise the Secretariat at the earliest possible opportunity of any such proposed commitments. A form is made available for this purpose.

### **Implementation Plan structure**

The Implementation Plan is structured around the five priorities identified in the GBIF Strategic Plan 2017-2021. These are presented here in reverse order from their sequence in the Strategic Plan, and numbered accordingly. This reordering allows the Implementation Plan to place its initial focus squarely on the global network of GBIF Participants and data publishers, building on these foundations to address in turn GBIF's informatics, content mobilization, curation and delivered products.

For each of the five priorities, the plan presents a series of numbered **Activities**. These are the major areas where GBIF needs to develop further in the coming period. A rationale and implementation approach is presented for each Activity.

Based on the implementation approach, each Activity includes a series of **Tasks**. These Tasks are the items that the Annual Work Programmes need to address. They are presented here as a table, at the end of each section describing the Activity.

## Overview of Priorities and Activities

<b>Priority 1 – Empower Global Network</b> .....	<b>13</b>
Activity 1a - Focus on people .....	13
Activity 1b - Strengthen skills .....	15
Activity 1c - Equip nodes .....	18
Activity 1d - Equip data publishers .....	20
Activity 1e - Expand national participation.....	22
Activity 1f - Plan implementation .....	24
Activity 1g - Coordinate resources .....	26
<b>Priority 2 – Enhance Biodiversity Information Infrastructure</b> .....	<b>29</b>
Activity 2a - Modernize data standards .....	29
Activity 2b - Deliver names infrastructure .....	32
Activity 2c - Catalogue collections .....	35
<b>Priority 3 – Fill Data Gaps</b> .....	<b>37</b>
Activity 3a - Identify priority gaps.....	37
Activity 3b - Expand data streams .....	39
Activity 3c - Engage data holders.....	42
Activity 3d - Rescue datasets .....	44
Activity 3e - Liaise with journals .....	46
<b>Priority 4 – Improve Data Quality</b> .....	<b>48</b>
Activity 4a - Ensure data persistence .....	48
Activity 4b - Assess data quality .....	50
Activity 4c - Enable data curation .....	52
<b>Priority 5 – Deliver Relevant Data</b> .....	<b>54</b>
Activity 5a - Engage academia.....	54
Activity 5b - Document needs.....	57
Activity 5c - Support biodiversity assessment .....	59
Activity 5d - Assess impact.....	61

## Priority 1 – 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.”*

### Activity 1a - Focus on people

#### Rationale

GBIF is the result of work by thousands of people in agencies and institutions worldwide. This network’s long-term sustainability depends on demonstrating the value of such contributions and justifying continued investment of effort.

The GBIF Secretariat can enhance efforts to develop capacity within the network and build an effective distributed help desk system that acknowledges and showcases relevant skills and experience that people across the network possess. Users of GBIF data products would also benefit from showcasing the network’s indispensable human resources and their impact on assessing and improving biodiversity data.

For these reasons, the next round of improvements to GBIF.org should enhance the network’s capacity to serve as a hub for the GBIF community by identifying its active contributors, integrating information on their relevant GBIF activities, and supporting broader knowledge exchange and skill development.

#### Approach

The key task is to enhance GBIF.org to connect and display information on its contributors and other users of the site. This may include contact details (subject to individual control), formal roles (e.g. within participant delegations; in connection to publishing/improving datasets; based on training experience) and online participation in help desk discussions. Whether through the use of existing social or commercial platforms or through custom web application development, GBIF.org should absorb activity currently handled through the separate GBIF community site. GBIF should also build on experience during 2015-2016 in using digital badges to identify skills held by individuals. Such community functions will support the operation of GBIF Task Groups, externally funded projects like BID, and engagement of expert groups to curate GBIF data.

#### Tasks

- 1.a.i Develop mechanisms to support and reflect the skills, expertise and experience of individual and organizational contributions to the GBIF network (including revision of identity management system and integration of ORCID identifiers)
- 1.a.ii Consolidate use of digital badges to record skills and contributions

#### Progress

Capacity enhancement activities through the BID programme continued to develop mechanisms to recognize and communicate the roles and skills of individuals involved in the GBIF community. In particular, project teams have access through the BID grantseekers’ site to a growing pool of mentors who volunteer their services and provide information about their areas of expertise; and two further capacity enhancement workshops, in the Caribbean and Pacific regions, have awarded digital badges to trainees recognizing basic and advanced skills in data mobilization.

The new GBIF.org website has improved the linking of multiple roles that individuals hold in GBIF governance (e.g. national delegations, nodes, regional representatives), demonstrating how individual profile information could be further developed in future.

The new site also links published datasets with research publications that cite data downloads through use of Digital Object Identifiers (DOIs), opening up further opportunities

to link individual roles in data mobilization with the re-use of data in research, for example by associating GBIF user accounts with ORCIDs (see also activity 5d, Assess impact).

**GBIF UK** has implemented Data Sharing Badges during 2016 and 2017 for all data contributors to the National Biodiversity Network (NBN).

### 2018 Work Items

- Enable users to associate ORCIDs with their GBIF.org accounts and to identify with which datasets they are associated. This will enable individuals to track citations through both the ORCID and DOI systems.
- Allow users to link their information to events and projects and declare their areas of expertise (e.g. Darwin Core, IPT, taxonomic groups) within GBIF activities.
- Retire the GBIF community site and replace with a discussion forum integrated into GBIF.org and with appropriate use of alternative platforms (e.g. Facebook, LinkedIn).
- Review current use of digital badge system in BID, align with skills training offered by partner networks, and expand to more generic application in capacity programmes.

### See Also

- [1b. Strengthen skills](#) - help desk and training activities build on community portal functions
- [1c. Equip nodes](#) – GBIF.org should highlight role of node managers and staff
- [1d. Equip data publishers](#) – proper credit should be given for data publishers
- [3c. Engage data holders](#) – stronger community support and help desk to support data publishing
- [4c. Enable data curation](#) – a community structure is required before GBIF can support engagement with experts as data curators
- [5d. Assess impact](#) – citation of data use in research literature will link directly to individual data publishing roles through use of DOIs and ORCIDs

## Activity 1b - Strengthen skills

### Rationale

The strengthening of personal skills through international collaboration has been one of the great successes of the GBIF global network. During this implementation period, GBIF must build on this past experience (including the support mechanisms developed for the BID programme) to reinforce efficient training and capacity enhancement across the network. Central to this is the development of a collaborative help desk capability and the alignment of relevant aspects of national training initiatives with a global curriculum to facilitate direct reuse of resources.

### Approach

To strength key skills, GBIF needs to develop and maintain a comprehensive set of clear reference information and training materials that support all GBIF audiences. During 2016, the Secretariat has been upgrading documentation to deliver this comprehensive resource, first via concise explanatory text materials (including numerous translations by network members into French, Portuguese, Spanish, Chinese, Japanese and Russian). By the end of 2016, some of these text materials will be supported by short videos that present key functional and operational aspects of GBIF. The aim is to deliver a structured set of short, clear, current and complete documents that enable GBIF stakeholders to gain an understanding of relevant aspects of GBIF's work quickly, including formal Participation, establishing a national or thematic node, planning digitization, mobilizing data, discovering and using data, citing data, among others. This work complements significant activity among Participants to produce information and training materials relevant to particular audiences.

GBIF will organize all these materials to form a curriculum that builds and supports the skills and activities identified in GBIF's self-assessment tools for nodes and data holders. A core set of these resources should provide new stakeholders with an effective introduction to GBIF's structure, roles, approaches and solutions. Other more technical resources will provide the additional detail needed by those working in specific areas like digitization, data publishing and use of GBIF-mediated data. The Secretariat will coordinate ongoing maintenance and updates to this curriculum with contributions and translations from the network.

### Tasks

- 1.b.i Develop collaborative help desk capability – including discussion groups, support materials and FAQs
- 1.b.ii Articulate training and capacity enhancement activities around a single, global curriculum for GBIF
- 1.b.iii Deliver explainer texts and videos to support self-directed instruction and learning

### Progress

Capacity enhancement activities through the BID programme have enabled the Secretariat to develop well-functioning collaborative helpdesk services incorporating a combination of direct support from Secretariat staff and brokering access to mentors through a community mailing list.

In addition, a set of training modules in the basic skills of data mobilization, created for the first BID capacity enhancement workshop in Africa in 2016, has been adapted and re-used for subsequent workshops in the Caribbean and Pacific in 2017, using an e-learning platform provided by GBIF Spain. These modules, including multi-lingual elements where appropriate, will be further applied at a second BID Africa workshop in December 2017.

The new GBIF website provides clearer introductory text explaining the key elements of GBIF participation including use of data standards, data publication and online hosting of datasets, with links to more detailed resources.

**GBIF France, GBIF Spain, GBIF Argentina** and **VertNet** contributed capacity for training and mentoring activities under the BID programme.

**GBIF Spain** has established a new e-learning platform (<http://elearning.gbif.es/>), which hosts the training modules for the GBIF BID programme and is expected also to support the needs of other GBIF Participants.

**GBIF Spain** and **GBIF Colombia** have prepared an animated promotional video on GBIF data sharing (as part of the CESP project “Facilitating the engagement of Ibero-American community”<sup>6</sup>) – this will be premiered at GB24. Additional short videos showcase experiences in both countries in managing and publishing biodiversity data.

### 2018 work items

- Adapt current BID data mobilization training modules into a more generic curriculum able to be picked up in future programmes by external trainers (including a proposed BIFA workshop in 2018), and including a larger element of self-instruction.
- Extend BID training modules to incorporate key skills around data application and use, with first demonstration at the BID Africa workshop in February 2018
- Complete set of explainer texts in multiple languages, as well as video explainers on key aspects of GBIF participation
- Develop BID helpdesk and mentoring processes for more general application across the network, including a mechanism to assign specific ‘focal point’ mentors to all new nodes and projects
- **GBIF Spain** will work with **GBIF Argentina** to make training materials from Argentina accessible through the Spanish e-learning platform
- **GBIF Spain** and **GBIF Colombia** will disseminate their promotional videos and assess impact
- **GBIF UK** has recruited a Technical and Data Support Officer and expects to be able to contribute to the collaborative GBIF help desk.

### See Also

- [1a. Focus on people](#) - acknowledgement of skills and contributions in the area of training and capacity
- [1c. Equip nodes](#) - capacity enhancement and support directed to nodes teams
- [1d. Equip data publishers](#) – documentation and help desk must support data publishing
- [1e. Expand national participation](#) – improved information materials will assist with communicating benefits of national participation
- [3c. Engage data holders](#) – improved information materials will assist with engagement of new data holders
- [4b. Assess data quality](#) – GBIF data quality approaches must be clear and well documented
- [4c. Enable data curation](#) – as GBIF seeks to engage experts to curate data, processes and tools must be clear and well documented

---

<sup>6</sup> <http://www.gbif.org/project/2016-ibero-american-engagement>



- [5a. Engage academia](#) – broader engagement with academic bodies, universities and libraries will require clear standalone information resources

## Activity 1c - Equip Participant nodes

### Rationale

By coordinating national, regional and thematic networks, Participant nodes play an essential role in helping GBIF engage the broadest possible community of institutions, initiatives and individuals engaged in biodiversity informatics. GBIF must provide learning materials and tools to support nodes efficiently as they work to mobilize biodiversity data, promote the reuse of available data and support users by improving data management and quality. Preferred approaches enable any one node to invest in developing tools and capacity that others can easily leverage for the benefit of the whole community. The skills and experience of the node managers and other team members are recognized as uniquely valuable in helping new Participants establish their nodes and allowing the community to develop together.

### Approach

GBIF promotes capacity self-assessment as a tool to help nodes with planning and tracking progress at all stages in the development of national biodiversity information facilities. The results of these assessments can also identify important capacity needs to address with learning materials, tools and collaborative projects. The reuse and adaptation of the Atlas of Living Australia's open-source software tools by a growing community of nodes between 2013 and 2016 is an example of successful collaboration based on the mutual needs of the countries involved. Additional support for this international partnership will allow still more nodes to benefit from the collective resources and expertise to build websites that provide biodiversity data and services to a range of national and regional user communities. The nodes' teams will also engage in collaborative helpdesk activities to enable them to share knowledge across the scientific and technical domains handled by the nodes.

### Tasks

- 1.c.i Perform ongoing nodes capacity self-assessment
- 1.c.ii Build reusable national/regional platform tools centred on ALA software
- 1.c.iii Develop online collaboration through GBIF.org helpdesk to assist and mentor node managers
- 1.c.iv Explore alternative options for lightweight national portal solutions, e.g. CKAN

### Progress

The national capacity self-assessment tool has helped some regions to assess the relative strengths and weaknesses of national participant nodes, and has provided baseline information for national projects within the BID programme from which to measure the impact of funded activities.

The collaboration around re-purposing of the Atlas of Living Australia portal tools has continued to expand as a significant focus of node activity. This has included the launch of the National Biodiversity Network (NBN) Atlas platform in the UK, as well as the selection of three separate projects under the 2017 Capacity Enhancement Support Programme related to the Living Atlases collaboration. This will include developing installations of the tool in Canada and Benin, as well as a workshop in 2018 on advanced application of the ALA tools involving participants from 10 countries. Interest in this collaboration has also come from the Caribbean, Asia and the Pacific.

To improve coordination and communication across the network, the Secretariat initiated a series of bimonthly webinars aimed at node managers and staff. These included four editions between January and August 2017, with the last two held at two separate times daily to accommodate different time zones. Secretariat staff have used the webinars to provide updates on key developments, canvass views and answer questions from nodes, and each

webinar has also included guest presentations from selected nodes. The webinars have been well-received by the Nodes Steering Group, and feedback has been universally positive.

**GBIF Spain** is developing a portal based on the ALA software for **GBIF Andorra**, planning an ALA portal for Catalonia, and advising **GBIF Uruguay**, **GBIF Chile** and **GBIF Luxembourg** on the technology. **GBIF Spain** is also assisting **GBIF Chile** and **GBIF Ecuador** with data publication.

### 2018 Work Items

- Review and further development of nodes guidance materials, including additional material on engaging citizen science, private sector, research and policy communities
- Retire the unmaintained community site of GBIF and replace with a more modern solution. A review of the functionality desired and needed will be conducted with the community, and one or more solutions put in place, including forums for direct discussion and helpdesk on GBIF.org and promotion and use of already commonly used tools such as Facebook groups.
- Review use of national self-assessment tool with a view to possible enhancements and guidance on future use
- Continue webinars and adapt based on feedback after first year, expand to wider audience including e.g. publisher and user communities
- Development of a role for members of the GBIF network and interested researchers to serve as 'GBIF science ambassadors' to promote understanding and use of GBIF data and services
- **Many GBIF Participants** will continue to work and seek funds for wider deployment of the ALA software for national portals

### See Also

- [1a. Focus on people](#) - acknowledgement of roles and contributions of those engaged in the nodes' teams
- [1b. Strengthen skills](#) - resources to support the learning of members of the nodes' teams that can also be adapted for reuse within their networks
- [1e. Expand national participation](#) – better tooling and support for nodes is an incentive for broadened participation
- [1f. Plan implementation](#) – implementing this plan requires strong nodes to take leadership and build national capacity
- [1g. Coordinate resources](#) - programmes supporting the development of fundable collaborative projects and capacity exchange and between the nodes
- [2c. Catalogue collections](#) – nodes are central to understanding and cataloguing national collections
- [3c. Engage data holders](#) – nodes lead in engagement with data holders, including collections, citizen science groups, consultants, universities, government agencies, etc.
- [4c. Enable data curation](#) – data curation mechanisms need to build on nodes capacity and engage with national bodies
- [5a. Engage academia](#) – nodes are the primary interface between GBIF and national academic bodies and universities
- [5b. Document needs](#) – nodes contribute to understanding data requirements and best practices and may take leadership in new fitness-for-use investigations
- [5d. Assess impact](#) – understanding and presenting GBIF's impact will support nodes in demonstrating the value of engagement and data publishing

## Activity 1d - Equip data publishers

### Rationale

Data publishers are an essential component of the GBIF network as they share their content through the common infrastructure. More than 800 data publishers actively distribute datasets through GBIF.org, and their ranks increase steadily. Publishers from different parts of the world often face unique challenges, though common themes emerge. These problems range from lack of data publishing experience or skills, lack of equipment, language barriers, difficulties in managing data hosting facilities, and the inability to publish high-quality data or curate data into the future. The Integrated Publishing Toolkit (IPT) requires ongoing improvements and enhancements, including the establishment of hosted instances that reduce the technical burden on data publishers.

### Approach

Following the model of the self-assessment tool for node managers, the Secretariat has developed a self-assessment tool for data publishers as part of the support for the BID programme, which will guide the work with collection managers and other data holders to assess and prioritize areas for capacity enhancement or investment. The Secretariat already operates instances of the IPT that data publishers lacking their own infrastructure may use, and Participants are encouraged to deploy instances of the IPT or other GBIF-compatible data publishing software to support data holding institutions. Planned enhancements to the IPT will simplify publishing pathways using spreadsheet templates as an alternative for the less advanced data publishers. GBIF will improve reporting to data publishers on both quality aspects of their data and uses of data documented through download DOI citations.

### Tasks

- 1.d.i Promote and support capacity self-assessment for data holders
- 1.d.ii Promote publication of collection metadata
- 1.d.iii Simplify data publication pathways (spreadsheet-level publishing)
- 1.d.iv Manage IPT feature upgrades
- 1.d.v Operate hosted IPT infrastructure
- 1.d.vi Consolidate NSG-led endorsement process
- 1.d.vii Develop online collaboration through GBIF.org helpdesk to assist and mentor data publishers
- 1.d.viii Provide clear online reporting of the use of data for data publishers
- 1.d.ix Promote data management plans as key tool for data publishers (Added 2018)

### Progress

Data holding institutions, including BID project partners, have been encouraged to use the self-assessment tool for data holders<sup>7</sup> to evaluate and plan mobilization activities.

The new version of GBIF.org includes enhancements to simplify the process and increase usability for new data publishers.

The GBIF Secretariat now operates a cloud-based installation of the IPT<sup>8</sup>, particularly to support BID projects. During the remainder of 2017 and 2018, the Secretariat will explore the best approach to expand such support on behalf of other data publishers. The goal is to

---

<sup>7</sup> <http://www.gbif.org/resource/82785>

<sup>8</sup> <https://cloud.gbif.org/bid>

ensure that all data holders have an appropriate channel for sharing their data, without undercutting existing Nodes activities and endorsement processes.

The IPT underwent a maintenance and bug fix release (2.3.4) in March 2017.

Data publishers can now use a service in GBIF.org to discover publications citing use of their data. The Secretariat actively monitors and documents journal publications, with the result that linkages can be made between citations of DOIs for GBIF downloads (in place since 2016) in publications and the contributing datasets.

Development is already under way to offer a revised data validation service. Users will be able to submit a dataset through GBIF.org or via the API and carry out a full interpretation of the data before registering and publishing a new or updates dataset. The API layer development is complete and the user interface will be finished during Q4 2017. This service will also be integrated within the IPT, and could be used as part of a broader pre-publication peer review process.

**GBIF Spain** organized a data publishing workshop in April 2017, including a whole day on metadata publishing and data papers. **GBIF Spain** also plans to deliver a service reporting to data publishers on usage of their datasets.

### 2018 Work Items

- Review of the IPT product and conduct a consultation with community on necessary changes beyond maintenance and bug fixing.
- Maintain cloud based IPT installations (updates, backups, helpdesk) and deploy further instances as required.
- **GBIF Spain** will repeat its workshop from April 2017

### See Also

- [1a. Focus on people](#) – GBIF.org needs to give credit to data publishers and support them through the help desk
- [1b. Strengthen skills](#) – documentation and help desk for data publishing
- [2a. Modernize data standards](#) – improvements are required to clarify and simplify key aspects of representing biodiversity data
- [2b. Deliver names infrastructure](#) – data publishers would benefit from a more robust and comprehensive taxonomic framework
- [2c. Catalogue collections](#) – cataloguing collections will highlight opportunities to digitize collections and mobilize their data
- [3a. Identify priority gaps](#) – identified gaps may support efforts to mobilize data
- [3b. Expand data streams](#) – GBIF should ensure tools and documentation support all relevant sources of data
- [3c. Engage data holders](#) – tools and documentation will assist with convincing data holders to publish data
- [3d. Rescue datasets](#) – community assistance is a mechanism to support data publishing
- [3e. Liaise with journals](#) – better visibility for data in research publications
- [4a. Ensure data persistence](#) – assistance with long-term management of data
- [4b. Assess data quality](#) – includes feedback to publishers on potential improvements
- [4c. Enable data curation](#) – opportunities for broader involvement in improving data
- [5d. Assess impact](#) – visibility for the important role of data publishers

## Activity 1e - Expand national participation

### Rationale

GBIF's national membership has remained largely static in recent years. Active participation in the network is confined to a limited number of regions, with the most dynamic activity in node collaboration and data publication focused in Western Europe, North America, Latin America, Oceania, increasingly sub-Saharan Africa and limited parts of Asia. Thus, significant parts of the world have little or no direct involvement in GBIF's activities, which poses a risk to the long-term credibility of GBIF as a global collaborative network. Secretariat staff and existing Participants constantly engage potential new members through ad hoc contacts, but the situation calls for a more strategic approach to expand membership.

### Approach

The Secretariat will compile and maintain an up-to-date register of the status of GBIF engagement and relevant contacts in all countries, providing the Governing Board with an annual review that analyses progress and remaining gaps in membership across each region. Following the development of a coordinated engagement plan for each region, Participants and regional groups with strong contacts to relevant agencies in other governments may take the lead in engagement. The Secretariat will continue to develop strategic engagement through the CBD and other intergovernmental networks, as well as through the BID programme and other capacity-related activities. Adapting the GBIF.org interface to support key content in multiple languages will seek to address language barriers to national expansion, with targeted translation carried out both through external contracts and in-kind services provided by Participants.

### Tasks

- 1.e.i Ensure GBIF.org interface and key content supported in relevant languages
- 1.e.ii Improve coordination and mechanisms for engaging new countries
- 1.e.iii Provide Governing Board with annual review and progress assessment for GBIF status and engagement in all countries

### Progress

Switzerland, Canada, Poland and Central African Republic have moved from Associate to Voting Participant status, and Mali has joined GBIF as an Associate Participant. The BID programme has generated new interest in joining GBIF among countries in Africa, the Caribbean and the Pacific Islands regions, and an extended regional meeting in Asia has helped engagement in the region with specific interest shown from Vietnam and Bhutan in joining GBIF. Development of an active community of data publishing institutions in Russia has led to significant data mobilization from a country that has represented a major gap in coverage, as well as potential collaborations in several Former Soviet Union countries.

The new website provides an opportunity to include multilingual version of key content that will support further engagement with unrepresented regions.

**GBIF Argentina, GBIF Belgium, GBIF France, GBIF Portugal, GBIF Spain, GBIF Togo** and **VertNet** all contributed translations for training materials used in the 2017 BID workshops.

**GBIF Australia** contributed presentations as part of the Asia and Pacific regional meetings.

### 2018 Work Items

- Further development of multilingual content on GBIF.org to support wider participation
- Annually-updated review of global participation based on regional and sub-regional representation, to support engagement strategies to be led by existing GBIF regions.

## See Also

- [1b. Strengthen skills](#) – clearer documentation, tools and basic information will assist with engaging national stakeholders
- [1c. Equip nodes](#) – support for establishing nodes will reinforce case for engagement
- [1f. Plan implementation](#) – all GBIF stakeholders to work together to plan and support engagement with additional countries
- [1g. Coordinate resources](#) – Secretariat to coordinate activity by GBIF Participants to engage national governments
- [5a. Engage academia](#) – increased national participation will enable linkages with additional academic communities
- [5d. Assess impact](#) – demonstrating GBIF's impact will support approaches to new countries

## Activity 1f - Plan implementation

### Rationale

Implementing this plan hinges on effective coordination with the plans of individual GBIF national and organizational Participants. GBIF must also ensure that node activities are well recognized and integrated with other biodiversity research and informatics initiatives at national, regional and organizational levels.

### Approach

To make national-scale implementation more central across the GBIF network, the Secretariat will consult with the Governing Board on models for unifying aspects of the work of the Nodes Committee into Governing Board meetings. The Secretariat also seeks funds to increase the scope of nodes meetings to include work to develop proposals for fundable activities aligned with this Implementation Plan, particularly by addressing regional priorities. The GBIF Secretariat will also engage organizational Participants to develop specific memoranda of cooperation in key areas and to identify and support alignments with this Implementation Plan and opportunities for joint organization of meetings and workshops.

### Tasks

- 1.f.i Develop better integration between nodes activities and Governing Board processes
- 1.f.ii Expand regional meetings to serve as GBIF community Implementation Planning conferences and opportunities to develop fundable programmes of work
- 1.f.iii Increase alignment between GBIF work programmes and activities of each GBIF Organisational Participant and Affiliate
- 1.f.iv Explore increased alignment and joint organisation of workshops with other networks such as TDWG, SPNHC, GEO BON, RDA, etc.

### Progress

This Implementation Plan is a first attempt to include information on the work programmes of individual GBIF Participants as part of an integrated view of the work of the GBIF community. Participants are encouraged to provide information at an appropriate level of detail on their planned activities and to update on these in future years.

The GBIF Secretariat has organized a series of webinars during 2017 to provide better communication with its Participant Nodes. Webinars have taken place on four dates and are now repeated early and late in the day in Copenhagen to simplify access for all parts of the GBIF network. These webinars are used to update GBIF stakeholders on progress and developments. All presentations are archived and available online.

GBIF regional meetings play an important role in helping Participants to coordinate their activities and develop shared vision for projects and engagement with other countries. Regional Nodes meetings have taken place during 2017 in Asia (with funding from Japan under the BIFA programme), Europe and North America. The Asian meeting in particular included representatives from countries which are not yet GBIF Participants.

Additionally, the BID-funded kick-off meetings for the Caribbean and the Pacific in July 2017 offered an opportunity to explore regional-scale engagement. In each case, existing GBIF Participants contributed to the meetings.

A workshop on Exploring Synergies and Sustainability for Biodiversity Information Systems was held at the Senckenberg Research Institute, Frankfurt, Germany. One of the recommendations arising from this workshop was to explore ways for GBIF to evolve to deliver a project office supporting the coordination of international biodiversity informatics activities and projects, and to develop the vision represented by the Global Biodiversity



Informatics Outlook (GBIO)<sup>9</sup>. A possible model for this would be for the GBIF Secretariat or a GBIF Participant to host two dedicated staff with responsibility for working with international organizations and projects to document existing best practices, develop roadmaps and delivery plans for convergent development of each of the GBIO components, and to coordinate activities to secure the required funding. An additional role could be to provide pre-submission comments on project proposals to encourage reuse of existing tools, standards and best practices and to support better alignment of projects within a public framework of GBIO roadmaps.

Clearly, GBIF does not currently have the required funding to include such a role within its mission. A possible starting-point would be to secure funding for a workshop to build on the 2012 Global Biodiversity Informatics Conference (GBIC). This conference could focus on developing a vision and funding strategy for establishing such a project office. It could also include working streams to develop draft exemplar roadmaps for two or three of the GBIO components. Such a conference could serve to develop a community vision and to create a suitable plan against which to secure the required funding.

An allocation of €50,000 towards organization of this workshop is included in the 2018 budget. The GBIF Secretariat would be pleased to hear from any Participants interested in contributing to the organization or costs for such a conference, or to hosting a GBIF Project Office for Architecture and Standards.

### 2018 Work Items

- Develop plans and coordinate funding for a workshop to develop a shared plan for a project office for coordinating biodiversity informatics activity and to develop draft roadmaps for collaborative delivery of some of the GBIO components.
- Organise regional Nodes meetings - €60,000 (revised following GB24) reserved in budget to support organisation of these meetings in regions lacking necessary funds
- **GBIF Spain** aims to develop close coordination with LifeWatch at the national level.

### See Also

- [1c. Equip nodes](#) – reinforce mechanisms for nodes planning and implementation
- [1e. Expand national participation](#) – work together to seek new Participants
- [1g. Coordinate resources](#) – shared planning supports better alignment of resources
- [2a. Modernize data standards](#) – aligning plans with other informatics activities depends on clearer data models
- [2b. Deliver names infrastructure](#) – requires joint plans with other projects and organisations
- [2c. Catalogue collections](#) – requires joint plans with other projects and organisations
- [3a. Identify priority gaps](#) – prioritisation is a key stage in planning implementation
- [3b. Expand data streams](#) – requires broad shared strategy
- [5b. Document needs](#) – needs should help to determine implementation
- [5c. Support biodiversity assessment](#) – requires joint plans with other projects and organisations

---

<sup>9</sup> <http://www.gbif.org/resource/80859>

## Activity 1g - Coordinate resources

### Rationale

Among its other roles, the GBIF Secretariat coordinates efforts to expand the scale and scope of GBIF activity beyond the levels achievable using only annual core Participant contributions. GBIF Participants commit to establish and operate nodes which serve as significant centres for GBIF activity. Some nodes have sufficient resources to contribute skills and developments which advance GBIF's work, while others may require external support to become fully active. A limited amount of funding has been allocated each year under the GBIF work programme to support capacity enhancement for GBIF nodes. GBIF or individual Participants may also secure supplementary funds to contribute to particular areas of GBIF work. Improved coordination of these various resources will assist GBIF to advance more rapidly at all scales.

### Approach

This Implementation Plan itself provides a framework for organising information on Participant and supplementary fund resources to complement GBIF's core funding. The GBIF Secretariat seeks information from all Participants on planned activities and commitments which may help to advance the work of other Participants or GBIF globally – examples may include funding for workshops; committed resources to develop tools, standards or best practices; mentoring actions; etc. These will be recorded as part of the overall GBIF Work Programme and updates will be presented in the GBIF annual report. A shared Implementation Plan also offers the opportunity for GBIF (either the Secretariat or Participants) to develop concept notes to seek supplementary funding to target currently unfunded or underfunded areas. The annual Capacity Enhancement Support Programme budget includes only limited funds, but existing supplementary funds (in particular BID and BIFA) align with these funds and leverage CESP tools and processes. Future supplementary funds should follow a similar model.

### Tasks

- 1.g.i Secretariat to coordinate participant-led activities aligned with Work Programme (process and models to support participant commitments)
- 1.g.ii Development of concept notes/proposals for supplementary funding at regional or global scales
- 1.g.iii Operate annual Capacity Enhancement Support Programme (CESP) with core funds, aligned with supplementary investments
- 1.g.iv Ensure that BID programme and other supplementary funds align with and reinforce CESP and GBIF community, online resources and infrastructure
- 1.g.v Align Ebbe Nielsen Challenge and Young Researcher Awards with GBIF priorities

### Progress

GBIF continues to operate with core funding levels at the bottom of the range considered viable by the Governing Board, and a relatively small proportion of contributions are received each year by the specified target date (31 March). By early August 2017, only around 81% of the committed annual contributions for 2017 had been received.

This pattern of payments, along with restricted financial reserves, places GBIF under continual risk that cuts to planned activities may become necessary, and significantly limits our ability to offer leadership and accelerate delivery.

Supplementary funding and alignment of distributed activity through this shared Implementation Plan have therefore been critical for enabling GBIF to progress its mission.

The Biodiversity Informatics for Development (BID)<sup>10</sup> programme, funded by the European Union, and the Biodiversity Information Fund for Asia (BIFA)<sup>11</sup> programme, funded by the Ministry of Environment of the Government of Japan, have both served significantly to expand the scope of GBIF's capacity enhancement activities. The GBIF Secretariat has used this opportunity to develop reusable templates and processes which have streamlined calls for proposals and contract development and, with assistance from volunteer trainers and mentors, to run repeatable training workshops in multiple regions. Relevant staff activities have been funded directly via these programmes. These advances not only serve the BID and BIFA programmes themselves, but also allow GBIF to operate its own Capacity Enhancement Support Programme (CESP) calls more efficiently. As a result, GBIF now offers an even more cost-effective framework for capacity enhancement activity using supplementary funds. The GBIF Secretariat is ready to work with Participants to seek further funding for such aligned programmes.

The funding provided by NLBIF for improvements to the names infrastructure offered by GBIF and the Catalogue of Life, has also included partial payment for a GBIF developer to work on this activity.

The Chinese Academy of Science has funded a year's work by Maofang Luo as visiting scientist working with Secretariat Participation team. She has contributed significantly to engagement activities within the Asian region.

Participants at GB23 identified their support for the Ebbe Nielsen Challenge, the Young Researchers Award and the biannual Global Nodes Meeting, which had all been identified as exceeding the 2017 core budget. We gratefully acknowledge the contributions received from the following Participants and collaborators to enable these three activities to take place:

1. The Swedish Research Council supplied €20,000 to support the 2017 Ebbe Nielsen Challenge
2. Stichting ter Bevordering van Natuurwetenschappelijk Onderzoek (SBNO) supplied €10,000 to support the 2017 Young Researchers Awards, with in-kind support from Pensoft Publishers through waiving of publication charges.
3. The US National Science Foundation made €50,000 available to support organization and attendance at the 2017 Global Nodes Meeting.
4. Finland waived the €10,000 normally supplied by GBIF to support the organization of the Governing Board Meeting.

In 2017, the Ebbe Nielsen Challenge has targeted the mobilisation of currently incompatible data held within online repositories. This is a direct contribution to the development of *Activity 3d - Rescue datasets* within this Implementation Plan.

Representatives from **GBIF Argentina**, **GBIF Colombia**, **GBIF Costa Rica**, **GBIF Norway**, **GBIF South Africa**, the **Chinese Academy of Science** and **VertNet**, and a large number of external experts, contributed to the assessment and evaluation processes for BID funding calls.

### 2018 Work Items

- Build on BID and BIFA programmes to seek additional funding for aligned capacity enhancement activity
- Operate the 2018 Capacity Enhancement Support Programme call with an allocated budget of €80,000.

---

<sup>10</sup> <http://www.gbif.org/programme/bid>

<sup>11</sup> <http://www.gbif.org/programme/bifa>

- Operate the 2018 Ebbe Nielsen Challenge and Young Researchers Awards, with a funding of €30,000 from the Governance allocation within the 2018 Budget.
- **GBIF Spain** is seeking long-term funding to strengthen the international community around the use of ALA software (currently around 20 countries).

#### See Also

- [1c. Equip nodes](#) – nodes are the primary focus for use of nationally aligned GBIF activity
- [1e. Expand national participation](#) – coordinate GBIF resources to seek new country Participants
- [1f. Plan implementation](#) – plans underpin coordinated activity
- [5d. Assess impact](#) – understanding impact will assist with future coordination

## Priority 2 – Enhance Biodiversity Information Infrastructure

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

### Activity 2a - Modernize data standards

#### Rationale

The GBIF network participants are able to reliably exchange data thanks to their adherence to a set of standards. As GBIF looks to grow in capability, enable exchange of richer content and improve the quality of data, the standards must be revised and evolve accordingly. Current standards adopted by GBIF are not yet adequate to accommodate the needs expressed by many potential and existing data publishers. Weaknesses in the model have led to ambiguous or over-complex data representations and unclear documentation, leading to difficulties in data integration and use. The main issues relate to uncertainties around the use of Darwin Core record types, the basisOfRecord element, and the use of Core and Extension vocabularies. Reviewing and updating the core domain model, tightening up the vocabularies and documentation and adopting more robust exchange standards will result in an easier to use, and a wider reaching GBIF data exchange network.

#### Approach

GBIF will work with TDWG and other key stakeholders to review existing solutions for a common domain model, working towards agreement on a model to adopt with key partners. This conceptual model should cover the main components of biodiversity information (the domain “classes” such as Specimen, Collection, TaxonName, TaxonConcept, Publication, Sequence) and document the mandatory and recommended properties expected for each component and the vocabularies that should control the properties. A review of existing vocabularies and their current uses will be undertaken and revisions and new vocabularies will be proposed where necessary. A revision of the Darwin Core Archive mechanism and supporting tools, such as the publishing toolkit (IPT) and the data validator, will be undertaken to accommodate the richer content model and the new recommendations from the W3C CSV on the Web working group. GBIF should continue discussions with other key global biodiversity data infrastructures to develop comprehensive catalogues to support discovery and normalisation of instances of the most critical domain classes (particularly TaxonName, TaxonConcept, Collection, Specimen, TaxonOccurrence).

In addition to completing this knowledge graph, GBIF should be equipped to link between people, datasets, cited use and funding agencies through the correct attribution chains using e.g. Digital Object Identifiers (DOIs) and Open Researcher and Contributor ID (ORCID) as potential mechanisms.

#### Tasks

- 2.a.i Promote development of a shared domain model for sharing and linking all components of biodiversity information
- 2.a.ii Lead a review of the Darwin Core vocabulary and associated extensions to ensure consistency and full alignment with a shared domain model
- 2.a.iii Explore opportunities to increase accessibility of biodiversity data through evolution of Darwin Core Archive formats to W3C CSV on the Web formats
- 2.a.iv Explore models to enable GBIF and other biodiversity infrastructures to deliver comprehensive global catalogues of instances of key data classes
- 2.a.v Improve management of trait data of relevance to GBIF

## Progress

Multiple engagements during 2017 highlighted the need for GBIF to continue to play a leading role in the development and advancement of data standards and community coordination for biodiversity informatics. In addition to those discussed under *Activity 2b - Deliver names infrastructure* and *Activity 2c - Catalogue collections*, these included:

- Distributed System of Scientific Collections (DiSSCo)<sup>12</sup> – a European consortium is proposing a new long-term research infrastructure, based on natural history collections. The GBIF Secretariat contributed to the development of the Technical Design for this proposal. GBIF and other key infrastructures would develop to offer integrative cataloguing services to enable the collections together to deliver a combined linked open data graph.
- Workshop on Exploring Synergies and Sustainability for Biodiversity Information Systems at Senckenberg Research Institute, Frankfurt, Germany – Senckenberg hosted this workshop to explore current challenges and opportunities for increased coordination and alignment between biodiversity informatics projects.
- Linking Environmental Data and Samples workshop<sup>13</sup>, Canberra, Australia – this workshop opened up discussion relating to common metadata for both geological and biological samples, use of persistent resolvable identifiers for collection samples, and representation of provenance in data derived from samples. In particular, the workshop indicated opportunities for GBIF and the biodiversity informatics community to learn and benefit from progress made by geological scientists around the International Geo Sample Number (IGSN) as a standard identifier for their samples.
- Sampling event data and Humboldt Core – the Darwin Core sampling event standard allows GBIF rapidly to expand the information accessible from many field-based data sets. Ideally, however, complementary standards would be in place to support cataloguing of sampling protocols and capture of essential metadata about these in a structured, semantically-enabled standard. A proposal is in place to develop such a metadata standard (Humboldt Core)<sup>14</sup>. GBIF has been involved in a number of discussions, particularly with Map of Life and in the context of GEO BON, around progressing this concept.
- TDWG Biodiversity Data Quality Interest Group<sup>15</sup> – this group met in Canberra during 2017 to advance standardization of data quality checks for Darwin Core data and to explore more consistent implementation of such checks by biodiversity data repositories and tools. GBIF will work during the rest of 2017 and 2018 to implement some of these ideas, and to increase consistency between GBIF data handling and data handling by e.g. ALA and iDigBio.

It is difficult for GBIF to provide the kind of coordinated and detailed input that is required in all of these processes. Secretariat staff members participate in key workshops, but currently all Secretariat teams are fully occupied with existing non-standards activities and cannot dedicate the necessary time to advance each of these to production-ready standards.

## 2018 Work Items

- GBIF will participate in the joint meeting of TDWG and SPNHC (Society for the Preservation of Natural History Collections) to be held in Dunedin, New Zealand (25 August – 2 September 2018).
- Implementation of recommendations from TDWG Biodiversity Data Quality Interest Group.

---

<sup>12</sup> <http://dissco.eu/>

<sup>13</sup> <https://csiro-enviro-informatics.github.io/environmental-data-symposium-2017/>

<sup>14</sup> <http://onlinelibrary.wiley.com/doi/10.1111/ecog.02942/abstract>

<sup>15</sup> <https://github.com/tdwg/bdq>

- **GBIF UK** is carrying out a review of Darwin Core, particularly in the context of national particulars, such as standard map references, and will contribute to ongoing review of the standards.

#### See Also

- [1d. Equip data publishers](#) – improvements to standards can simplify and improve tools and documentation
- [1f. Plan implementation](#) – changes will require coordinated activity with TDWG and other partners
- [2b. Deliver names infrastructure](#) – should reflect broader standards model
- [2c. Catalogue collections](#) – should reflect broader standards model
- [3b. Expand data streams](#) – expansion to new categories of occurrence and sample data should offer clarity around provenance and evidence type
- [4b. Assess data quality](#) – improvements will simplify and enhance quality processing
- [4c. Enable data curation](#) – data curation should build on clear data standards
- [5c. Support biodiversity assessment](#) – depends on clear standards for underlying evidence

## Activity 2b - Deliver names infrastructure

### Rationale

The most significant challenge to improving the quality of aggregated occurrence data is the continuing need for a comprehensive checklist of known species, and even for a comprehensive list of published scientific names. Interpreting and mapping names depends on the quality and completeness of these resources. Even in cases where names in occurrence records are incorrect or misspelled, better names infrastructure can assist by increasing confidence that fuzzy match algorithms or human intervention is required. Delivering these resources is the focus of a number of GBIF Participants and other stakeholders, including the Catalogue of Life partnership, WoRMS, nomenclators (IPNI, Index Fungorum, ZooBank) and many national, regional or taxonomic databases. A comprehensive resource for scientific names and taxon concepts organized at least as a workable reference classification (but with support for additional classifications as appropriate) would also benefit other infrastructures, including Encyclopedia of Life, Biodiversity Heritage Library, Barcode of Life and GBIF nodes, and improve interoperability between data from these infrastructures. It would also be beneficial to accommodate vernacular names, informal names for undescribed species and other identifiers such as Barcode Index Numbers.

### Approach

GBIF and many other partners have worked on this challenge and much progress has been made, but we are still far from a comprehensive shared solution. GBIF has been in discussion with Catalogue of Life, EOL, BHL, BOLD Systems, nomenclators and others about pooling resources to deliver the best possible complete nomenclator and catalogue of all species, along with improved tools to enable the taxonomic community to own and maintain these resources more effectively. The challenges are not primarily informatics issues. The most important requirement is to understand the constraints and needs of existing content holders and the features that are required from an infrastructure that can be embraced by the majority of taxonomists. The solution must build on existing initiatives and give sufficient credit and benefit back to those who have invested in developing data. It must be flexible enough to accommodate existing well-managed datasets without disrupting their activity and to accommodate more open mechanisms to support wide community input for taxa which need more work. In the longer term, it should support evolution towards ownership of curation responsibilities by international taxonomic societies or other bodies recognized by researchers for each group. The infrastructure should include processes to review and interpret unrecognized name strings found by GBIF and others in aggregated data. Once these requirements have been resolved, implementation must rapidly follow to offer these resources as open public datasets for use by all.

The Netherlands has coordinated a significant commitment for 2017 and 2018, led by NLBIF and including resources both from Species 2000 and Naturalis. This funding will enable GBIF and partners to direct significant effort to this area over the period.

### Tasks

- 2.b.i Partner with other biodiversity informatics initiatives and taxonomic database holders to plan and deliver a comprehensive nomenclatural dataset and working consensus classification for all life
- 2.b.ii Promote publication of species checklists through GBIF network
- 2.b.iii Explore potential use of checklists to assist with data validation or derive augmented data products
- 2.b.iv Explore integration of Linnaean nomenclature of formally described taxa with provisional names and species hypotheses and OTU naming



## Progress

The project funded by **GBIF Netherlands** (NLBIF) started with a workshop in Leiden<sup>16</sup> in November 2016, attended by representatives of the Catalogue of Life partners, holders of nomenclatural data, and a selection of national and regional checklist activities. Participants unanimously agreed that renewed focus is required on rapid delivery of a comprehensive nomenclator and Catalogue of Life.

Following the workshop, NLBIF recruited Olaf Bánki to act as project manager for the activity and provided partial funding for the GBIF Secretariat to contract Markus Döring to continue work on GBIF taxonomic data management.

In July 2017, a project plan was agreed by the GBIF Secretariat, Catalogue of Life Team and NLBIF, and communicated with the major nomenclatural data holders. The plan aims to re-engineer both the GBIF taxonomic backbone and the Catalogue of Life as a shared clearing-house for species names and concepts, with GBIF ultimately assuming responsibility to host this shared infrastructure. This model will allow Catalogue of Life to offer a comprehensive working checklist of all species, linked to the nomenclators as primary references for all species names, and for GBIF and other infrastructures to benefit from a consistent and shared resource for interpreting all scientific names.

Ideally, the solution to be developed should combine with the work of the World Flora Online (WFO)<sup>17</sup> to deliver a single, consistent, community-managed plant list. GBIF and Catalogue of Life aim to explore this in more detail alongside the WFO Council meeting to be held in Paris in November 2017.

In preparation for task 2.b.iv, the GBIF Secretariat has tested presentation of species concepts based on Barcode of Life Data Systems (BOLD) barcode index numbers (BINs). This will lead to recommendations on the use of Darwin Core terms in such situations.

**Encyclopedia of Life** (EOL) has carried out work to incorporate elements of the GBIF backbone taxonomy into its new EOL working hierarchy. This will open opportunities for better integration of EOL trait data with GBIF data.

## 2018 Work Items

- Deliver public prototype and solicit community feedback on comprehensive nomenclator and checklist using clearing-house model.
- **GBIF Netherlands** has provided two years of funding and staff time for the project to deliver a comprehensive nomenclator and checklist.
- **GBIF Mexico** will promote publication of checklist datasets, in particular those used for data validation within CONABIO, and proposes to collaborate with the GBIF community to standardise methods for using checklists to validate specimen records. **GBIF Mexico** is also working on mechanisms for handling data records which are not associated with formally published names, or for species of *incertae sedis*.
- **GBIF Spain** plans to work on a comprehensive species checklist for its territory.
- **GBIF UK** is using its national species lists for automated data validation and will be reviewing the associated rule sets.
- **Encyclopedia of Life** (EOL) proposes to work with GBIF during 2018 on improvements to the management, representation and interoperability of taxonomic checklists.

## See Also

- [1d. Equip data publishers](#) – key tool to support data publishers in improving data
- [1f. Plan implementation](#) – requires joint activity with multiple partners

<sup>16</sup> <http://www.gbif.org/newsroom/news/species-names-in-november-workshop>

<sup>17</sup> <http://www.worldfloraonline.org/>

- [2a. Modernize data standards](#) – part of broader standards revision
- [3a. Identify priority gaps](#) – complete taxonomy would assist with evaluating taxonomic coverage
- [3c. Engage data holders](#) – benefits for data holders in linking to consistent framework
- [4b. Assess data quality](#) – lack of complete checklist a limit on quality processing
- [4c. Enable data curation](#) – curation of taxonomic framework and taxonomic aspects of other data
- [5a. Engage academia](#) – need to engage effectively with taxonomic societies and nomenclatural codes
- [5c. Support biodiversity assessment](#) – taxonomic organisation of data fundamental

## Activity 2c - Catalogue collections

### Rationale

Natural history collections are the largest source of data on biodiversity outside nature itself. Collectively the world's natural history collections number about 3 billion specimens and document over 300 years of active human exploration of biodiversity on earth. In addition the fossil remains provide us with glimpses into the very far past before humans existed. The collections and their ancillary materials (images, collectors' notes, sequences, measurements, etc.), contain colossal amounts of data that should be digitized and shared. Only about 10% of the world's collections have been digitized and only a portion of digitized collections are shared publicly through the internet. A large number of current GBIF publishers comprise natural history museums and herbaria. The Secretariat will work with Participants to deliver the most comprehensive catalogue possible of collections, including metadata to publicize undigitized collections as a first step towards their digitisation and mobilisation.

### Approach

GBIF will collaborate with existing biodiversity collection registries such as GRBio to develop a rich collections catalogue to facilitate the discovery and use of the world's collections. Collections will be guided and facilitated to publish metadata-only datasets by drawing on the recommendations of the GBIF task force on accelerating the discovery of bio-collections data. The generation of a collection catalogue through metadata assessments will help establish roadmaps for further digitization of the collections.

### Tasks

- 2.c.i Partner with other biodiversity informatics initiatives to deliver a single, comprehensive catalogue of the world's natural history collections
- 2.c.ii Use collection metadata as a first stage in content mobilisation from natural history collections, including identification or development and adoption of necessary data standards

### Progress

The GBIF Secretariat has engaged in discussions with GRSciColl<sup>18</sup> regarding the possibility of GBIF hosting and taking responsibility for the registry of scientific collections. This makes excellent sense for GBIF, since good management of metadata relating both to digitized and undigitized collections will improve GBIF's ability to clean and improve existing specimen records, and also can underpin broader effort to seek funding for collection digitization and to build a global virtual natural history collection.

More work is required. This can most readily proceed following the cut-over to the new GBIF.org site.

**GBIF Mexico** has a reviewed catalogue of the main collections and institutions which contribute data to the national information facility, SNIB.

### 2018 Work Items

- Develop joint action plan with GRSciColl for joint delivery and maintenance of the catalogue of collections.
- **GBIF Spain** is using its national collection metadata catalogue to help mobilize new data into GBIF.

### See Also

- [1c. Equip nodes](#) – nodes contribute to and benefit from data on collections

---

<sup>18</sup> <http://grscicoll.org/>

- [1d. Equip data publishers](#) – assisting collections in publicising their holdings is a route towards seeking funding for digitisation
- [1f. Plan implementation](#) – need to coordinate with GRBio and other catalogues
- [2a. Modernize data standards](#) – catalogue should build on common data standards
- [3a. Identify priority gaps](#) – knowledge of collections can assist with identifying digitisation opportunities
- [3c. Engage data holders](#) – catalogue can assist with engagement of collections as publishers
- [5a. Engage academia](#) – catalogue should be an operational tool for researchers
- [5b. Document needs](#) – fitness-for-use groups may identify priority collections for mobilisation
- [5d. Assess impact](#) – catalogue can measure progress in global digitisation

## Priority 3 – 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.”*

### Activity 3a - Identify priority gaps

#### Rationale

GBIF has a range of tools, including fitness-for-use groups, other community consultations, feedback channels, direct communication with authors of scientific studies, and societal demands, to identify and collect data needs. The Ebbe Nielsen Challenge for 2016 focuses specifically on tools and algorithms to identify significant gaps. These gaps may relate to different facets of the data, including geography, taxonomy, time periods, and coverage of particular ecosystems or land units. Addressing these gaps may require focus on gaining additional occurrence records, targeting data areas missing from published records, or getting additional metadata elements. By consolidating and prioritising demands for data content, in the context of already accessible data and knowledge of resources which are not yet available as open data, GBIF will be positioned to inform collection and data holders, funding institutions and political decision makers of the most worthwhile and cost-effective ways to extend the available knowledge base.

#### Approach

The GBIF Secretariat will harmonize and document data mobilization demands from different sources. Simple tools are required to support needs capture, including informative documentation and justification for such demands. Automated assessment and reporting of gaps will be included where this proves possible and valuable. This combined information can support transparent decision making and target setting for gap-filling efforts, allowing all interested actors to step in at appropriate levels. GBIF will coordinate with efforts through the Intergovernmental Platform on Biodiversity and Ecosystem Services (IPBES) to identify and address significant knowledge and data gaps, including outreach and funding strategies for gap-filling. A thorough, regularly updated overview of data coverage in [gbif.org](http://gbif.org) both makes it easier to identify gaps, and to monitor progress and efficiency of mobilization efforts over time. GBIF should offer (e.g. annually) a brief report of significant gaps which need to be addressed. Such a report may be valuable to Participants and funding bodies to stimulate and evaluate digitisation and mobilisation options.

#### Tasks

- 3.a.i Organize and integrate requirements identified by fitness-for-use groups
- 3.a.ii Establish open mechanism for researchers and users to document specific data needs
- 3.a.iii Develop assessments and visualisations of strengths and weaknesses of GBIF data in key dimensions (taxonomic, geographic, environmental, temporal)
- 3.a.iv Maintain well-publicized GBIF priorities for data publishing, incorporating needs of IPBES and other networks
- 3.a.v Provide Governing Board with annual review and progress assessment for digitisation of major natural history collections
- 3.a.vi Provide GBIF Participants with actionable priorities and targets for content mobilization

## Progress

The 2016 Ebbe Nielsen Challenge<sup>19</sup> targeted solutions for exploring and presenting information on gaps in GBIF data. The Secretariat will seek to integrate some of these ideas into its tools as part of the next stage in enhancing the GBIF.org portal. Following the 2016 Challenge, Map of Life has developed a proposed indicator based on GBIF data coverage.<sup>20</sup>

This will provide a foundation for GBIF to maintain a set of annually updated content mobilization priorities and to develop tools to enable Node Managers and others to address known issues and weaknesses in existing data. The Secretariat has started this approach by highlighting three general priority areas for 2017: mobilization of data relating to the countries with the lowest density of existing data; continued mobilization of data from natural history collections (the major source for the widest possible range of species data); and mobilization of sampling event data (the major source for comparative data on species distributions and abundance). In 2018 and onwards, these priorities will be refined and used as measures for progress in content mobilization.

The requirements from the GBIF fitness for use working groups on agrobiodiversity, species distribution modeling and invasive species are being compiled into requirements for implementation and uptake both by GBIFS and other actors. Some of these requirements will be addressed as part of a 2018 focus on improved and more transparent data quality checks during indexing. Publication of national lists by the Global Register of Introduced and Invasive Species (GRIIS)<sup>21</sup> will enable other recommendations from the group on invasive species data to be implemented. A GBIF CESP project, co-funded by the Portuguese Instituto Superior de Agronomia, has progressed the work of the group on agrobiodiversity by designing a training course on the use of such data from GBIF.<sup>22</sup>

## 2018 Work Items

- Synthesize outcomes of recent activities (including Ebbe Nielsen Challenge 2016) to visualize and measure data gaps and to offer data mobilization targets
- Encourage additional research communities to provide feedback on possible improvements to GBIF data management and presentation to meet their needs – this may be delivered following the fitness-for-use group model.

## See Also

- [1d. Equip data publishers](#) – priorities can assist with planning or funding digitisation
- [1f. Plan implementation](#) – implementation should respond to critical gaps
- [2b. Deliver names infrastructure](#) – taxonomic framework important for identifying gaps in taxonomic coverage
- [2c. Catalogue collections](#) – support identification and targeting of undigitized collections
- [3b. Expand data streams](#) – prioritize sources of data to address gaps
- [3c. Engage data holders](#) – highlight value of targeted mobilisation
- [3d. Rescue datasets](#) – additional justification for mobilisation
- [5a. Engage academia](#) – communicate priority needs
- [5b. Document needs](#) – fitness-for-use priorities should be included
- [5c. Support biodiversity assessment](#) – addressing gaps will be critical to providing high-quality data for assessments

---

<sup>19</sup> <http://www.gbif.org/newsroom/news/2016-ebbe-nielsen-winners>

<sup>20</sup> <https://mol.org/indicators/coverage/background>

<sup>21</sup> <http://griis.org/>

<sup>22</sup> <https://demo.gbif.org/project/82833/agrotraining-proofing-gbif-use-on-agrobiodiversity-through-needs-assessment-and-training>

## Activity 3b - Expand data streams

### Rationale

GBIF serves as an integration point for any source of evidence of the recorded occurrence of species in time and space. A primary role for the GBIF infrastructure is to serve as a comprehensive single point of access for discovery, access, use and curation of all such evidence. Several classes of data are already well-supported within the GBIF network. These include collections data, observations from field research, and many categories of citizen science data. However, there are other new and developing streams of data which should be accommodated if GBIF is to serve as the platform for supporting comprehensive data assessment and modelling (e.g. for GEO BON Essential Biodiversity Variables, IPBES assessments, Red List assessments, etc.). These include efforts to mine historical data records from literature, genomics activities and particularly barcode-driven surveys, and potentially species-level data from remote-sensing systems. More work is also still needed to engage with the full spread of research activities delivering sampling event data of various kinds. GBIF needs to ensure that it provides simple, effective and beneficial ways for researchers to share these and other streams of Darwin Core compatible data.

### Approach

Existing GBIF models include support for occurrence records and for sampling-event datasets which organize occurrence records as sets of observations deriving from a single field sample (which make provision for GBIF to accommodate “absence data” from surveys which did not record a particular species despite searching). These approaches are core to all potential streams of data to be added. GBIF therefore needs to ensure that existing tools and documentation are clear and usable for relevant research communities and that GBIF sufficiently understands existing data management by these communities to avoid proposing unnecessary additional work. During 2016, GBIF is coordinating a consultation which builds on past engagements with genomics activities such as the Global Genome Biodiversity Network. Recommendations from this consultation are expected to guide improvements in GBIF tools, documentation and communications to support publishing of molecular data in formats which can be integrated within GBIF. Several projects are working on automated or human mining of data records from literature. GBIF needs to learn from these initiatives and ensure that its tools support integration in a simple way. GBIF should also seek exemplar projects for bringing occurrence records from remote sensing into the network.

### Tasks

- 3.b.i Promote use of sampling event data model for ecological and monitoring datasets
- 3.b.ii Partner with BHL and others to support integration of species occurrence records based on literature
- 3.b.iii Work with bioinformatics initiatives and databases to form robust bidirectional linkages with molecular data
- 3.b.iv Explore opportunities to integrate species-level data from remote sensing

### Progress

GEO BON, the Group on Earth Observations Biodiversity Observation Network<sup>23</sup>, has promoted mobilization of sampling event data as a priority for monitoring and representing biodiversity patterns, particularly within the framework of Essential Biodiversity Variables (EBVs). A GEO BON/iDIV/GBIF symposium on sampling event data was held in Leipzig during June 2017, with 60 attendees. This meeting set goals and a timeline to grow volumes of sampling event data from a current baseline of around 13 million records to 100 million

---

<sup>23</sup> <http://geobon.org/>

sampling event records by 2020, via a campaign targeting the data management of networks, journals and publishing houses, individual data holders and research centres.

GEO BON is also exploring development of EBV data from remote sensing, and is organizing a workshop in Twente in September, in which the GBIF Secretariat will participate to explore opportunities to mobilize occurrence and sampling-event data from remote sensing systems.

BHL and GBIF are seeking external funding to prototype extraction and publishing of occurrence data from historical literature using the existing OCR and name discovery functionality of BHL, enhanced using gazetteers and publication metadata.

Several activities have furthered exploration of solutions to integrate molecular (DNA) data both as evidence of species occurrence and to enrich and improve GBIF's taxonomic backbone, including: 1) enhancements to the representation of occurrence data derived from GenBank sequences; 2) exploration with iBOL of steps necessary to integrate data first from barcode sequences associated with binomial names and subsequently from other sequences assigned to Barcode Index Numbers (BINs)<sup>24</sup>; and 3) sandbox tests of integration and representation of informal species concept names from UNITE and BOLD within the GBIF taxonomic backbone.

The GBIF infrastructure does not currently support the data extensions developed jointly by GGBN and GBIF, but combining this standard with the sampling event model offers a possible route for effective mobilisation of metagenomics / eDNA data, to the joint benefit of both networks.

Overall, there are many opportunities for GBIF to integrate significant additional streams of data, and the Secretariat does not have the resources to engage fully in all of these areas. We welcome interest from Participants in leading such engagements.

### 2018 Work Items

- Work with BHL to incorporate literature-based occurrence data.
- Work with BOLD and others to incorporate sequence-based occurrence data.

### See Also

- [1d. Equip data publishers](#) – support for data publishers is fundamental for this activity
- [1f. Plan implementation](#) – broad GBIF planning and engagement is required to work with all potentially relevant data communities
- [2a. Modernize data standards](#) – improvements to data standards will ensure that the underlying characteristics of the evidence for GBIF data are clear
- [3a. Identify priority gaps](#) – new data streams may address gaps in existing coverage
- [3c. Engage data holders](#) – support and tools are inadequate without engaging communities to prioritize sharing data
- [3d. Rescue datasets](#) – mechanisms to liberate historical data will also benefit from support for a broader range of data sources
- [3e. Liaise with journals](#) – journals are important channels for coordination and promotion of research data publishing
- [5a. Engage academia](#) – support from key research bodies and field studies
- [5b. Document needs](#) – requirements from users may assist with prioritising new data streams

---

24

[http://www.boldsystems.org/index.php/resources/handbook?chapter=2\\_databases.html&section=bins](http://www.boldsystems.org/index.php/resources/handbook?chapter=2_databases.html&section=bins)



- [5c. Support biodiversity assessment](#) – some new streams may have high value for delivering comparable data for monitoring and assessment

## Activity 3c - Engage data holders

### Rationale

GBIF has tools and support mechanisms in place to enable publication of several categories of biodiversity data. The key requirement is for increased engagement with, and support for, the communities of institutions and individuals who hold these data. Such engagement is normally most effective at the national level, although international networks also have an important role to play.

### Approach

The primary need is for national Participants to engage with the broadest possible spread of data holders within their countries, for organisational Participants to share relevant data and for the whole GBIF network to promote the importance and value of sharing data. The GBIF Secretariat will focus on enhancement to documentation and tools and on highlighting priorities for complementing existing data and addressing gaps. Data holders should be encouraged and assisted in sharing data in the richest form appropriate for the data in question (sampling event data where relevant elements are available, occurrence data for other spatially explicit data, checklists otherwise) and with the most open data licences possible. As well as natural history collections, Participants should identify opportunities to build partnerships with citizen science groups and promote the value of open access to data from environmental impact assessments and monitoring. GBIF should argue the case for open data as part of the policy for funding agencies, research councils, industry bodies, licensing authorities, development banks and other stakeholder groups.

### Tasks

- 3.c.i Engage with natural history collections not yet publishing to GBIF
- 3.c.ii Work with national citizen science groups
- 3.c.iii Promote national policies and platforms to enable publishing of environmental impact and monitoring data
- 3.c.iv Promote importance of data sharing to funding bodies, licensing authorities and industry bodies
- 3.c.v Address the tropical data gap

### Progress

Support materials for data publishers have been updated and reorganized on the new web site, including information on sampling events<sup>25</sup> and the general publishing guide, and a new access point for information on citizen science<sup>26</sup>.

Both digitization of natural history collections (3.c.i) and mobilization of sampling event data (including 3.c.ii and 3.c.iii) are identified within the 2017 content mobilization priorities for the GBIF network.

GBIF data sharing goals and procedures were publicized at the 2017 Citizen Science Association (Minnesota, U.S.) and the International Association for Impact Assessment Conference, (Montréal, CA), and via various communications with the European Citizen Science Association.

### 2018 Work Items

- Engagement with data holder communities relies on effective national scale coordination. The GBIF Secretariat will revise guidance for Participant Nodes on

---

<sup>25</sup> <https://demo.gbif.org/sampling-event-data>

<sup>26</sup> <https://demo.gbif.org/citizen-science>

engagement with citizen science groups, private sector environmental impact consultants, etc.

- Material will be added to GBIF.org to summarize guidance and best practice relating to data mobilization from environmental impact assessment. The GBIF Secretariat will continue to engage with the International Association of Impact Assessment, World Bank, International Finance Corporation, World Economic Forum, etc. to promote open data publication through GBIF.

- 

#### See Also

- [1a. Focus on people](#) – GBIF must give visibility to the efforts of data publishers
- [1b. Strengthen skills](#) – documentation and support materials must be as clear and simple as possible
- [1c. Equip nodes](#) – nodes are the main agent for engaging data publishers
- [1d. Equip data publishers](#) – GBIF must focus on delivering tools and processes which are intuitive and effective for data holders to use
- [2b. Deliver names infrastructure](#) – better names management and integration will be a benefit for data publishers
- [2c. Catalogue collections](#) – a collections catalogue will assist with engaging collections as data publishers
- [3a. Identify priority gaps](#) – knowledge of key gaps will assist with prioritising focus on particular data holders and may assist them with securing necessary funds
- [3b. Expand data streams](#) – GBIF needs to support the widest possible range of relevant data sources
- [3d. Rescue datasets](#) – processes to assist data holders to publish data even if they themselves have limited capacity
- [3e. Liaise with journals](#) – work with journals to promote and prioritize data publishing
- [4a. Ensure data persistence](#) – provide robust and stable support for published data
- [4c. Enable data curation](#) – offer benefits to data publishers through community collaboration in improving data

## Activity 3d - Rescue datasets

### Rationale

Many researchers hold potentially valuable data which are not yet in a suitable digital format for integration into GBIF. Historical publications are a similar source of valuable data which remain inaccessible. This offers an opportunity to establish a community platform to capture information on such datasets where the researcher or owner lacks the time or capability to make the data available as a GBIF-compatible dataset, and to enable interested individuals to volunteer time to collaborate with the owner to publish a dataset, potentially in conjunction with a data paper credited to all parties. Such a model may address a key bottleneck in bringing valuable data online.

### Approach

The GBIF Secretariat, or an interested Participant, should develop a test environment to explore this model. The model should support identification of basic information on datasets which may be rescued, including details of ownership, etc. Volunteers may be required to undergo some training or demonstrate some knowledge of GBIF data publishing and the taxa concerned prior to adopting a dataset for mobilisation. Mobilisation should include consultation or partnership with the owner and should deliver quality metadata and a valid mapping of the original information. Opportunities should be explored for publication of resulting datasets as data papers as an incentive to all parties.

### Tasks

- 3.d.i Develop tools for reporting potential data sources for integration into GBIF
- 3.d.ii Develop support materials (including accreditation) for collaborative data preparation and mapping datasets in GBIF
- 3.d.iii Develop site and support mechanisms for users to adopt and map datasets
- 3.d.iv Review and update definitions of data publisher within GBIF to reflect collaborative data publishing
- 3.d.v Develop partnerships with data journals to support data papers for rescued datasets

### Progress

During 2017, a tool has been deployed to allow users to suggest datasets<sup>27</sup> for mobilization into GBIF. This captures information on suitable datasets in GitHub<sup>28</sup> (63 data targets in July 2017) with country and taxonomic tags.

During 2018, the Secretariat will work with interested Participants and repositories to develop support tools and processes to enable such datasets to be preserved in a form which can contribute to GBIF's mission. The required solution should build on existing stable trusted repository solutions, include good support for mapping diverse formats into structured data standards such as Darwin Core, and should align with GBIF endorsement processes, possibly recast as peer review quality checks. Ideally, datasets published using this pipeline should also be ready for description via data papers.

### 2018 Work Items

- Establish a repository and processes for hosting, curating, publishing and endorsing datasets suggested for rescue.

---

<sup>27</sup> <https://demo.gbif.org/suggest-dataset>

<sup>28</sup> <https://github.com/gbif/data-mobilization/issues>

## See Also

- [1d. Equip data publishers](#) – a process for data rescue would serve as an additional tool to support data publishers
- [3a. Identify priority gaps](#) – priority gaps may motivate sharing and rescue of data
- [3b. Expand data streams](#) – data rescue processes will benefit from other support for additional data sources
- [3c. Engage data holders](#) – data rescue will reduce the work for some data holders in making their data accessible
- [3e. Liaise with journals](#) – supplementary materials include datasets which may be rescued
- [4c. Enable data curation](#) – linkages between data rescue and curation communities
- [5a. Engage academia](#) – additional mechanism to support work with researchers

## Activity 3e - Liaise with journals

### Rationale

Journals are the traditional established avenues for scientific communication. They not only disseminate research findings and other scholarly communications but are increasingly helping to disseminate research data. These data may be provided as supplementary materials or deposited in biodiversity data repositories as a precondition for publication of the paper. However, such data publishing data does not necessarily facilitate its integration with other related data or make the data discoverable and reusable. In order to benefit from data disseminated through journals, the Secretariat will lead or coordinate activity to engage directly with publishing houses, journal editors and authors to promote GBIF-compatible approaches to publication and the use of GBIF-operated repositories as accepted or preferred destinations for supporting data.

### Approach

GBIF Secretariat will engage with relevant journals in developing and promoting best practices and publishing workflows that reduce the effort required to publish data to GBIF. The aim will be to publish data once but be able to cite it as supporting data for a journal paper and where appropriate produce a data paper as an additional product. The Secretariat will use its experience with Pensoft Publishing where a workflow and a dedicated data paper publishing tool is established.

### Tasks

- 3.e.i Develop scalable approach to support research journals and data journals in publishing to GBIF network
- 3.e.ii Produce relevant support materials to justify benefits and explain processes to publish primary data
- 3.e.iii Integrate support for data journals into hosted IPT infrastructures and data rescue processes

### Progress

Since GBIF started associating data downloads with digital object identifiers (DOIs), there has been a steady growth in citation of these DOIs within research publications. Every time this occurs, GBIF is able to establish a link between the published research and the individual datasets which contributed records to the data download.

By the time of the GBIF Science Review 2017, more than 100 scientific publications<sup>29</sup> had cited GBIF data downloads. This information is now made visible on the associated GBIF dataset pages and can be shared with other infrastructures which graph the connections between researchers, their projects, publications, funders and datasets (e.g. OpenAIRE<sup>30</sup>). This development will enable GBIF to deliver increased value both for data users (through simplified tools for citing data and increasing reproducibility) and for data publishers (through greater visibility of the uses to which their data contribute).

To accelerate uptake during 2018, the GBIF Secretariat will enhance documentation and communication materials, and will seek support from relevant journals to promote best practices on data archival and data citation.

---

<sup>29</sup> <https://www.gbif.org/resource/search?q=gbifDOI&contentType=literature>

<sup>30</sup> <https://www.openaire.eu/>

## 2018 Work Items

- Develop communications with journals around GBIF DOI-based data management, data downloads and data citation policy. Encourage journals, editors and authors to practise responsible data citation.

## See Also

- [1d. Equip data publishers](#) - hosted IPT infrastructure to support journal datasets
- [3b. Expand data streams](#) – support for additional data sources will increase value of GBIF as a data repository for use by journals
- [3c. Engage data holders](#) – data holders may be motivated by increased visibility in conjunction with published research
- [3d. Rescue datasets](#) - engage community assistance in mapping research data into GBIF
- [5a. Engage academia](#) – bidirectional opportunities to build GBIF visibility and influence
- [5d. Assess impact](#) – increase citation levels for GBIF-hosted data

## Priority 4 – 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.”*

### Activity 4a - Ensure data persistence

#### Rationale

There exists a significant portion of data available through GBIF.org that is not actively curated by a data host. In some cases, there are no resources or desire to make further edits to the datasets. These datasets are effectively orphaned and the GBIF.org version of the dataset is often the last remaining version available on the internet. As GBIF develops mechanisms to provide feedback to data publishers and support curation of datasets, we need to consider that these orphaned datasets will not be updated with corrections or migrated to adhere to modern data standards.

#### Approach

The task is to ensure that all datasets have a primary version available on the internet which acts as the source for GBIF.org to index. Orphaned datasets will be identified, extracted from the GBIF.org index and loaded into the most suitable data repository supporting versioning: either run by a GBIF participant or a central cloud installation of an IPT. As issues are identified anyone will be able to volunteer to correct the source data, upload a new version into the data repository, document the changes applied and follow editor guidelines. Once republished GBIF.org will reflect the updated data, and the provenance of changes will be traceable through the repository versioning system. Policies for editors, including attribution and the settlement process for disputes will be documented. This entire activity could be led and implemented by a GBIF Participant.

#### Tasks

- 4.a.i Identify and verify datasets within GBIF network without current owners
- 4.a.ii Publish reference instances of these datasets within hosted IPTs
- 4.a.iii Develop processes and mechanisms for adoption of orphaned datasets by suitable agencies or experts

#### Progress

The GBIF Secretariat has identified datasets in the GBIF data index for which there is no currently accessible online source and for which there is a story of failed indexing attempts. These datasets have been identified as apparently “orphaned”.

Following this, consultation and collaboration began with GBIF Participants in February 2017, aiming to re-activate as many of those datasets as possible, and to clarify responsibilities for further action on the rest.

Where possible, orphaned datasets are transferred for hosting by a Participant node. Where this option is not available, hosting will temporarily be established on a cloud-based IPT instance managed by GBIFS, allowing for later transfer to another host or re-activation by the original publisher.

Policies for reactivation, adoption and curation are currently being drafted.

#### 2018 Work Items

- Establish ongoing detection for future orphan data sets and processes to ensure persistent hosting and management.



- **GBIF Spain** expects to take over hosting for around 100 Spanish datasets which are not currently in an IPT instance.

#### See Also

- [1d. Equip data publishers](#) – ensure sustainable benefits from data publishing
- [3c. Engage data holders](#) – greater security for data holders that data publication brings long-term benefit
- [4b. Assess data quality](#) – data persistence is a key aspect of data quality
- [4c. Enable data curation](#) – data can only be curated if a stable reference copy exists

## Activity 4b - Assess data quality

### Rationale

Assessing data quality includes applying data validation tools to capture and monitor suspected and confirmed errors and ambiguities in data, highlighting useful areas for additional information (metadata and qualifiers) that would improve usability and enhance processing options, and documenting completeness and standardisation of information both within a dataset and within aggregated data. A number of validation tools exist in the wider community, and should be brought together to mutually profit from investments and to more efficiently plan future distributed development efforts. This will benefit data publication frameworks as well as individual data holders, giving concrete feedback on best gains in data management.

### Approach

Consolidation requires an overview of existing data validation tools, their goals and application areas, building on existing community work to produce an annotated tools catalogue (including work by TDWG and the GEO BON “Bon in a Box”). To make best use of development resources, GBIF will support collaboration between networks to bring those developments together and harmonize efforts, so that further development can more efficiently concentrate on new priority areas. Consistent tests and reports will both inform users of the suitability of data for their use, provide feedback to publishers on their holdings, provide a measure for the overall state of the network, and help to prioritize improvement options. Ideally, the most common reporting measures and formats are agreed and unified to a degree that allows publishers an easy cross-walk between and integration of data quality reports supplied by different services and aggregators.

### Tasks

- 4.b.i Develop extensible data validation tools framework in partnership with ALA, TDWG and other networks (e.g. Symbiota, iDigBio, VertNet)
- 4.b.ii Integrate consistent data validation tools in GBIF.org, national/regional portals, IPT and elsewhere
- 4.b.iii Improve presentation and reporting of data validation results
- 4.b.iv Develop regular data set reports for data publishers and nodes

### Progress

A new data validation service will shortly be released on the GBIF website and is in final development. The service allows for users to upload a dataset and run full data interpretation as if it were being indexed, but without actual registration of data. It will be available through both the API and through web forms for users to upload datasets. The API service will be embedded into the IPT and encouraged for use in other platforms to help enable pre-publication data review and fit into a data peer-review process.

The underlying data validation libraries of this process are the same as used in GBIF.org and in part already used in the ALA data products. An agreement has been reached between GBIF and ALA to further align the data management processes by combining the source code and early discussion around what this entails is underway. This activity will start late 2017 and continue into 2018. Once complete, data will be treated consistently: through the pre-publication review and then subsequent indexing by ALA and GBIF. Capacity to maintain and advance these routines will be increased as everyone will be contributing to a common project.

The TDWG data quality interest group have progressed the thinking towards a standardized set of terms for describing data quality and the necessary statements. A workshop was held in Canberra alongside an ALA symposium to progress the validation rules required. This will

provide useful additions for the GBIF / ALA progress but is currently not sufficiently complete to be fully implementable. GBIF needs to fully engage with that process and lead a reference implementation during 2018 to ensure the implementation and associated standards are aligned.

### 2018 Work Items

- Engage with TDWG Data Quality group to define and document the rules and flagging required to assess data quality.
- Engage with the ALA team specifically, and others where possible, to ensure that systems treat data consistently, and use the same software libraries. The GBIF / ALA ingestion processes should become the reference implementation for a TDWG standard.
- Accumulate and present dataset quality statement metrics for the dataset, publisher and national level on GBIF.org to help direct all parties to address data issues and prioritise effort.
- Evaluate options to incorporate expert review as an element in GBIF data management.
- **GBIF Spain** will use tools from GBIF.org to report to Spanish data publishers on use of their datasets.

### See Also

- [1b. Strengthen skills](#) – documentation and training materials must give clear guidance on quality
- [1d. Equip data publishers](#) – tools and processes must reinforce focus on quality
- [2a. Modernize data standards](#) – data provenance and semantics are key to quality
- [2b. Deliver names infrastructure](#) – gaps in the names infrastructure hinder quality assessment and improvement
- [4a. Ensure data persistence](#) – persistence is a key aspect of quality
- [4c. Enable data curation](#) – curation processes should focus on measurable improvement to quality
- [5a. Engage academia](#) – support researchers in publishing high quality data and collaboration to improve data
- [5b. Document needs](#) – gather requirements which can improve quality management at all stages
- [5d. Assess impact](#) – overall quality metrics for GBIF data

## Activity 4c - Enable data curation

### Rationale

In a global network, curation of the shared data pool is increasingly becoming a joint responsibility of aggregators, publishers, experts and data users. The goal is to integrate corrections, improvements, additional information and analysis results in a timely manner, with better visibility to all network participants and data users. Expanding the existing knowledge base requires improved communication channels and workflows for collaboration between all actors, tools to capture and rapidly display new or improved information, commentary and data, and not least tools, credit systems and support to engage expert activities.

### Approach

The main task is to provide tools and mechanisms that make it easy for users and experts to contribute knowledge to the available pool of data. Building on existing data filtering and data improvement workflows in the community, GBIF tools and mechanisms are to support the identification of relevant data, their cleaning and preparation for specific purposes, and the sharing of the results of such processes with the wider community. Input collected through existing feedback mechanisms (Jira feedback on gbif.org, AnnoSys, ...) is to be raised to a visibility level that supports and drives the usefulness of the published data.

### Tasks

- 4.c.i Develop network-wide approach to handling and processing annotations and feedback
- 4.c.ii Enhance IPT to offer dataset-level peer review and commentary mechanisms and record-level annotations
- 4.c.iii Develop mechanism and tools within GBIF.org for sharing cleaned and annotated datasets based on GBIF downloads (“reference datasets”)
- 4.c.iv Develop GBIF data workbench tool (within GBIF.org and possibly also as standalone) for cleaning and filtering network data (e.g. in red list assessments)
- 4.c.v Develop strategy and support mechanisms for expert communities to curate sections of GBIF data

### Progress

As part of the updates to GBIF.org, modifications have been made to the processes for users to offer feedback. Users are asked first to identify the nature of feedback – content issues, software bugs, an idea or a request for help. Where issues relate to content, a user is encouraged to submit feedback via the channel which is most appropriate for the data publisher in question. Where data comes from an active online community, such as iNaturalist a user is encouraged to use the existing discussion forums on the originating portal. Similarly where records come from taxonomic databases connected through BioCASE, the AnnoSys annotation system is suggested. The approach can be expanded to allow any connected system to self-identify the preferred method of receiving feedback. This will continue through 2018 with more systems being connected.

Similarly, GBIF has identified that a significant number of datasets lack anyone to maintain them and respond to data issues, which accordingly will never be addressed. A community process is underway to identify these datasets and to move them into a centralized IPT installation which will allow anyone with sufficient interest to contribute to maintaining and correcting the data. Such an approach will enable a wider community to actively fix issues rather than simply identifying them.

## 2018 Work Items

- Connect more systems to receive direct annotations from GBIF.org users. Symbiota and Specify communities have expressed interest.

## See Also

- [1a. Focus on people](#) – give credit for contributions to data improvement
- [1b. Strengthen skills](#) – documentation and training materials to guide data curation
- [1c. Equip nodes](#) – nodes will be key in engaging expert communities
- [1d. Equip data publishers](#) – ensure that data curation processes bring benefits to publishers
- [2a. Modernize data standards](#) – a common data model underpins effective curation
- [2b. Deliver names infrastructure](#) – normalising names and taxonomy is a key aspect of curation
- [3c. Engage data holders](#) – data holders need to be able to respond to curation actions
- [3d. Rescue datasets](#) - integration of new but previously uncurated data sources
- [4a. Ensure data persistence](#) – curation depends on persistent datasets
- [4b. Assess data quality](#) – understanding issues with data quality can drive future tools and guidance for curators
- [5a. Engage academia](#) – collaborate with taxonomic societies and other expert communities
- [5b. Document needs](#) – incorporate end user requirements in curation processes

## Priority 5 – 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.”*

### Activity 5a - Engage academia

#### Rationale

The most significant user community for GBIF is academic researchers. Even policy-related uses of GBIF often derive from the work of such individuals. It is accordingly important for GBIF to understand the needs of researchers and academic societies and to communicate clearly regarding the tools and services GBIF can deliver. Communication should include information and support materials for students and early-career researchers, on both publication and use of data, including citation, use tracking and data papers. University faculties and libraries may be important channels for this information. In addition, GBIF needs to engage more closely with taxonomic societies and other academic bodies which could be key collaborators in curating and improving data. Achieving such an outcome depends on understanding how GBIF can become a more central tool for their work, so that work on digital knowledge directly benefits those who contribute.

#### Approach

GBIF Participants may be well-positioned to lead in engaging with taxonomic societies and other interested research groups, including assuming responsibility to serve as ambassadors for GBIF within particular communities. Particular focus should be given to opportunities to explore models for such communities to assist GBIF with data curation. Few university programs include clear guidance and promotion for open data sharing. GBIF should make use of its own documentation materials and of curriculum materials from throughout the GBIF community to engage with biological science faculties and university libraries to share information on GBIF tools and resources and practices.

#### Tasks

- 5.a.i Promote biodiversity informatics curriculum, support training of the users
- 5.a.ii Publicize GBIF as data tool via university libraries and faculties
- 5.a.iii Develop online collaboration through GBIF.org helpdesk to assist and support data users
- 5.a.iv Represent GBIF at the relevant science fora, esp. at the frontier directions such as ecology and molecular research

#### Progress

GBIF has been represented or will be represented, either by Secretariat staff or by relevant representatives from Participants, at a significant number of biodiversity and research workshops and conferences, including:

- Computer Science Meets Ecology seminar (Dagstuhl, Germany, February 2017)
- International conference on “The use of modern information technologies in botanical investigations” (Apatity, Russia, March 2017)
- ForBio Annual Meeting (Trondheim, Norway, April 2017)
- Linking Environmental Data and Samples conference (Canberra, Australia, May 2017)
- Inaugural Digital Data in Biodiversity Research Conference (Ann Arbor, USA, June 2017)
- International Botanical Congress (Shenzhen, China, July 2017)
- TDWG conference (Montreal, Canada, October 2017)

- International Barcode of Life Conference (Kruger NP, South Africa, November 2017)
- BES/GFÖ/NecoV/EEF Ecology Across Borders conference (Ghent, Belgium, December 2017)

At three of the larger of these conferences, GBIF operated as an exhibitor, as well as promoting data publishing and use through symposia and talks.

The Science Committee and Nodes Steering Group have reviewed and responded positively to a proposal for GBIF to promote a role for GBIF science ambassadors, established researchers with an interest in promoting data publishing and use of GBIF data through conference talks, organising local symposia, etc.

The Secretariat has streamlined and improved processes for tracking GBIF-relevant scientific literature in order to deliver the Science Review 2017. As part of this activity, authors were contacted directly to gain more detailed information on the number of records used, and to explain and promote use of DOIs for data downloads. This year, for the first time, GBIF has become able to illustrate direct links between research papers and GBIF datasets on which the research is based.

The Young Researchers Award 2017 will reward researchers for publication of sampling event data.

During 2017, it has become clear that current GBIF brochures are not sufficiently clear and effective for communicating the value of GBIF to researchers. A key publicity goal for 2018 is to deliver more appropriate materials for distribution at scientific conferences and elsewhere on the data products offered by GBIF, the need for researchers to share their data, and the benefits associated with data citation.

**GBIF Spain** reports that information on GBIF is included as part of Biology and Environmental Sciences degrees in at least seven Spanish universities.

### 2018 Work Items

- A BID workshop will be held in Africa during February 2018 to assist projects with engaging users in applied use of mobilized data
- New brochures will be produced to highlight the relevance of GBIF for researchers (multiple languages).
- **GBIF Spain** will work to contact additional universities on use of GBIF within lessons and courses.

### See Also

- [1b. Strengthen skills](#) - broader engagement with academic bodies, universities and libraries will require clear standalone information resources
- [1c. Equip nodes](#) – nodes are key to engaging national research councils and academic societies
- [1e. Expand national participation](#) – engagement with academic societies may support further national engagement
- [2b. Deliver names infrastructure](#) – taxonomic societies should be key partners in delivering names infrastructure
- [2c. Catalogue collections](#) - catalogue should be an operational tool for researchers
- [3a. Identify priority gaps](#) – communicate priority needs with academia
- [3b. Expand data streams](#) – work with research communities to incorporate additional sources of data
- [3d. Rescue datasets](#) - additional mechanism to support work with researchers
- [3e. Liaise with journals](#) - bidirectional opportunities to build GBIF visibility and influence

- [4b. Assess data quality](#) – work with research communities to improve data quality
- [4c. Enable data curation](#) – engage research communities to oversee and improve data
- [5b. Document needs](#) – engage relevant academic groups as fitness-for-use working groups
- [5d. Assess impact](#) – published research is a key measure of GBIF impact



## Activity 5b - Document needs

### Rationale

GBIF data are aggregated from many sources and are consequently heterogeneous, varying in fitness for various uses. During 2015-2016, GBIF established three task groups on data fitness-for-use, in agrobiodiversity research, in distribution modelling and in research on invasive alien species, to document how these communities use GBIF data and to understand their data quality demands. The resulting reports inform data mobilization, data processing and improvements to GBIF.org. During 2017, GBIF will expand this work to incorporate more key areas of use, provisionally focussing on: 1) Phylogenetic and (continuation from 2016 work) DNA evidence for names and occurrences, 2) human health and vectored diseases and 3) marine biodiversity. Depending on resources, these groups will operate through a combination of face-to-face meetings and online or remote collaboration.

### Approach

The DNA work will continue from 2016 through online meetings and e-mails. Human health and vectored disease fitness for use group will be formed and will operate using the basic model with two in-person meetings. Work on fitness for use in marine research will continue depending on the availability of marine data from OBIS. Participants are encouraged to organize working groups in additional areas to review the state of GBIF data relevant to particular domains and themes and to provide recommendations on critical gaps, improvements in recommended metadata or data elements, minimum criteria for usable data, etc. The GBIF Secretariat will compile these recommendations and, where possible, develop query profiles to support rapid access and monitor progress by the network in delivering data suited for the needs of these domains.

### Tasks

- 5.b.i Prioritize areas for fitness-for-use assessments
- 5.b.ii Support fitness-for-use groups

### Progress

As part of the upgrades to the GBIF.org site, the GBIF Secretariat is enhancing processing and data validation for individual data records and for datasets. This will include integration of recommendations from the 2015-2016 data fitness for use groups on agrobiodiversity, distribution modelling, and alien and invasive species.

Further fitness-for-use groups have been deferred until these mechanisms are in place.

Nevertheless, the GBIF Secretariat is keen to receive recommendations from Participants and offers to organise and run additional working groups to identify ways to support particular research communities.

### 2018 Work Items

- Address recommendations from 2015-2016 data fitness for use groups.

### See Also

- [1c. Equip nodes](#) – nodes may coordinate requirements gathering in particular areas
- [1f. Plan implementation](#) – GBIF plans should respond directly to needs
- [2c. Catalogue collections](#) – comprehensive knowledge of collections will serve as a basis for identifying collections particularly important for digitisation
- [3a. Identify priority gaps](#) – needs feed into assessing gaps in data
- [3b. Expand data streams](#) – additional data sources may address needs

- [3c. Engage data holders](#) – work with data holders to improve quality and address gaps
- [4b. Assess data quality](#) – needs may determine quality checks which can be performed and metrics to report
- [4c. Enable data curation](#) – assistance with setting curation priorities
- [5a. Engage academia](#) – work with research communities to define needs
- [5c. Support biodiversity assessment](#) – requirements for assessment are a key aspect of needs

## Activity 5c - Support biodiversity assessment

### Rationale

One of GBIF's key roles is as organizer of global evidence for species distribution, based on point records for species in time and space. Expansion of data publishing to accommodate sampling event data enables this evidence base also to mobilize and organize basic data on species populations and abundance. As a result, GBIF is positioned to serve as a critical resource for supporting biodiversity assessment at all scales. In particular, GBIF should serve as the data foundation for GEO BON to deliver Essential Biodiversity Variables (EBVs) for species distribution and population abundance. These EBVs represent a continuum from modelled representation of species occupancy (presence-only) in defined units of space and time through to richer assessment of species abundance in those units. GEO BON should serve as a forum for addressing the challenges of modelling such variables and interpolating sensibly between existing data points. This includes determination of appropriate scale at which modelled variables are adequately supported by current data. GBIF needs to ensure that it delivers the data foundations required for these activities, thereby supporting the requirements of IPBES, species Red Listing through IUCN and national authorities, the CBD and the Aichi Targets and Sustainable Development Goals.

### Approach

This activity depends on the activities defined for all goals, in order to improve the comprehensiveness, coverage and quality of data accessed through the GBIF network. An additional area for enhancement will be GBIF.org search interfaces and presentation (via maps and other visualisations) of sampling event data. Such interfaces should ensure that researchers can properly evaluate and assess data suitable for biodiversity assessment activities. Additionally GBIF needs to be active within GEO BON to ensure that there is close alignment and evolution between the concepts and implementation of EBVs and the data and services offered by GBIF.

### Tasks

- 5.c.i Enhance GBIF.org presentation of sampling event data
- 5.c.ii Support development of species distribution/population abundance EBVs
- 5.c.iii Showcase tools and practices for use of GBIF in Red Listing and other species assessment processes

### Progress

The GEO BON<sup>31</sup> community is working closely with GBIF as the key global repository for occurrence and sampling event data, and is promoting the importance of sampling event data as a generalised approach to bring together large amounts of information on species abundance and distribution. The joint GEO BON/GBIF/iDiv Symposium on the Mobilization of Structured Biodiversity Data<sup>32</sup> was particularly important as a tool to build a joint vision and to secure the engagement of a number of key data holders.

This partnership positions GBIF as a critical infrastructure in support of two of the proposed Essential Biodiversity Variables<sup>33</sup> for Species Distribution and Population Abundance. Workshops under the EU-funded GLOBIS-B<sup>34</sup> project have made progress in defining the parameters for these inter-linked variables, and GLOBIS-B has funded workshops to explore the use of data from GBIF and the ALA to construct such variables.

---

<sup>31</sup> <http://geobon.org/>

<sup>32</sup> <http://geobon.org/a-roadmap-for-the-mobilization-of-structured-biodiversity-data/>

<sup>33</sup> <http://geobon.org/essential-biodiversity-variables/monitoring/>

<sup>34</sup> <http://www.globis-b.eu/>

As the work on EBV develops, GBIF must support best practice for mobilising and integrating data for constructing the best and most fine-scale representations possible of evidence for species occurrence and abundance for any area or time period. Achieving this will require continued improvements in data quality checks and enhancements both to GBIF data indexes and to access and visualisation tools to support EBV workflows.

### 2018 Work Items

- Explore possible enhancements to GBIF data management to optimize access to data in formats which are appropriate for constructing species distribution and population abundance EBVs.
- Further discussion is required to integrate GBIF data into IUCN Red List processes.

### See Also

- [1e. Expand national participation](#) – assessment is a need and motivation for national agencies
- [1f. Plan implementation](#) – delivering high quality global and regional datasets in support of assessment will require coordinated action
- [2a. Modernize data standards](#) – improve information on data provenance and semantics
- [2b. Deliver names infrastructure](#) – names are key to organising data from multiple sources
- [3a. Identify priority gaps](#) – assessment needs will contribute to prioritisation
- [3b. Expand data streams](#) – ensure that all relevant data sources are included
- [5b. Document needs](#) – requirements for assessment are a key aspect of needs
- [5d. Assess impact](#) – many governments, along with intergovernmental collaborations, expect GBIF to deliver solutions in support of assessment activities

## Activity 5d - Assess impact

### Rationale

GBIF Participants require clear evidence of the benefits arising from investments in national-scale content mobilisation and from GBIF global activity. At present, the main source of evidence presented derives from monitoring of published literature to identify uses of GBIF within research. This activity has been reported through annual GBIF Science Reviews and clearly demonstrates growing use of GBIF in research. Monitoring the literature in this way is time-consuming, and becomes more so as the relevant literature increases. A sustainable approach is required for future monitoring of this kind. GBIF now issues and promotes Digital Object Identifiers for data downloads and expects that these can be used both to simplify discovery and to improve the detail offered to Participants and data publishers on some uses of data. Participants also require more information on non-research uses of GBIF infrastructure, particularly in various kinds of government or industry assessments. A broader review of costs and benefits arising from GBIF investment would be valuable for Participants arguing continued engagement within GBIF and other countries considering Participation.

### Approach

The GBIF Secretariat will lead a consultation with Participants on the level of detail desired from literature tracking and possible approaches to maintain this as a shared task across the network. GBIF.org will be enhanced to present known uses of GBIF DOIs and to ensure that data publishers receive information on such uses. Within the BID programme, GBIF is already seeking reports from funded projects on non-research applications of data. A general model is required to enable Participants easily to contribute examples of such uses and for these to be integrated into GBIF's overall reporting on usage. Several countries have requested information on known costs and benefits from GBIF investment. The Secretariat wishes to identify one or more Participants interested in leading case study investigations to value GBIF benefits at national scale and to assist with developing general recommendations for a suitable way for GBIF to track and report this value.

### Tasks

- 5.d.i Develop sustainable approach to literature tracking
- 5.d.ii Automate detection and reporting of use of GBIF DOIs
- 5.d.iii Provide mechanisms to report and track uses of data and grey literature
- 5.d.iv Develop a valuation for the services delivered by the GBIF network

### Progress

The Secretariat has significantly improved and streamlined the process of tracking research uses of GBIF as a data source, with many more examples of data use being summarised and featured in the annual Science Review and on the new GBIF.org website. Monitoring of the research literature during 2017 has revealed a significant acceleration in data use, with the number of peer-reviewed papers citing use of GBIF averaging more than two publications each day. The new website provides much more agile means of searching and filtering such examples based on thematic uses, countries (both researchers and geographic focus of research) and other facets. The growing use of download DOIs in literature citations has enabled the new website to link published datasets directly with research examples.

While non-research uses of GBIF continue to be more challenging to identify, specific examples have included use in national and global Red List processes and identification of Key Biodiversity Areas.

The Secretariat has scoped approaches for commissioning a value proposition study, and potential consultancies have been informally approached with a view to launching a tender for such a study later in 2017.

GBIF has not contracted an external review of its operations and deliverables since 2010. Such a review is overdue. €30,000 has been set aside in the budget to allow the Governing Board to engage a suitable panel of experts to carry out such a review and provide recommendations, including face-to-face meetings. Ideally such a review should include an audit and certification of GBIF as an infrastructure and network delivering repository services.<sup>35</sup>

### 2018 Work Items

- Extend collection of GBIF use cases to include grey literature and other non-research applications of data
- Commission value proposition study – extra funding will be required.
- Contract external review of GBIF operations and deliverables - €30,000 allocated.

### See Also

- [1c. Equip nodes](#) – information on impacts will assist nodes in building strong national networks and nodes will be the key source of information on uses of data outside peer-reviewed literature
- [1d. Equip data publishers](#) – data publishers need information on the benefits arising from use of their data
- [1e. Expand national participation](#) – information on impacts will increase the case for GBIF participation
- [1g. Coordinate resources](#) - understanding impact will assist with future coordination
- [2c. Catalogue collections](#) – catalogue will serve as benchmark for progress in digitising collections
- [3e. Liaise with journals](#) – use of GBIF data in published research a key measure of impact
- [4b. Assess data quality](#) – improvements in quality should be a metric for impact
- [5a. Engage academia](#) - published research is a key measure of GBIF impact
- [5c. Support biodiversity assessment](#) - many governments, along with intergovernmental collaborations, expect GBIF to deliver solutions in support of assessment activities

---

<sup>35</sup> <https://www.rd-alliance.org/groups/repository-audit-and-certification-dsa%E2%80%93partnership-wg.html>