

# GBIF Implementation Plan 2017-2021 and Annual Work Programme 2020

## Overview

This document serves both as the 2020 revision of the GBIF Implementation Plan for the period 2017-2021 and also as the GBIF Annual Work Programme 2020. Both elements are based upon the [GBIF Strategic Plan for 2017-2021](#).

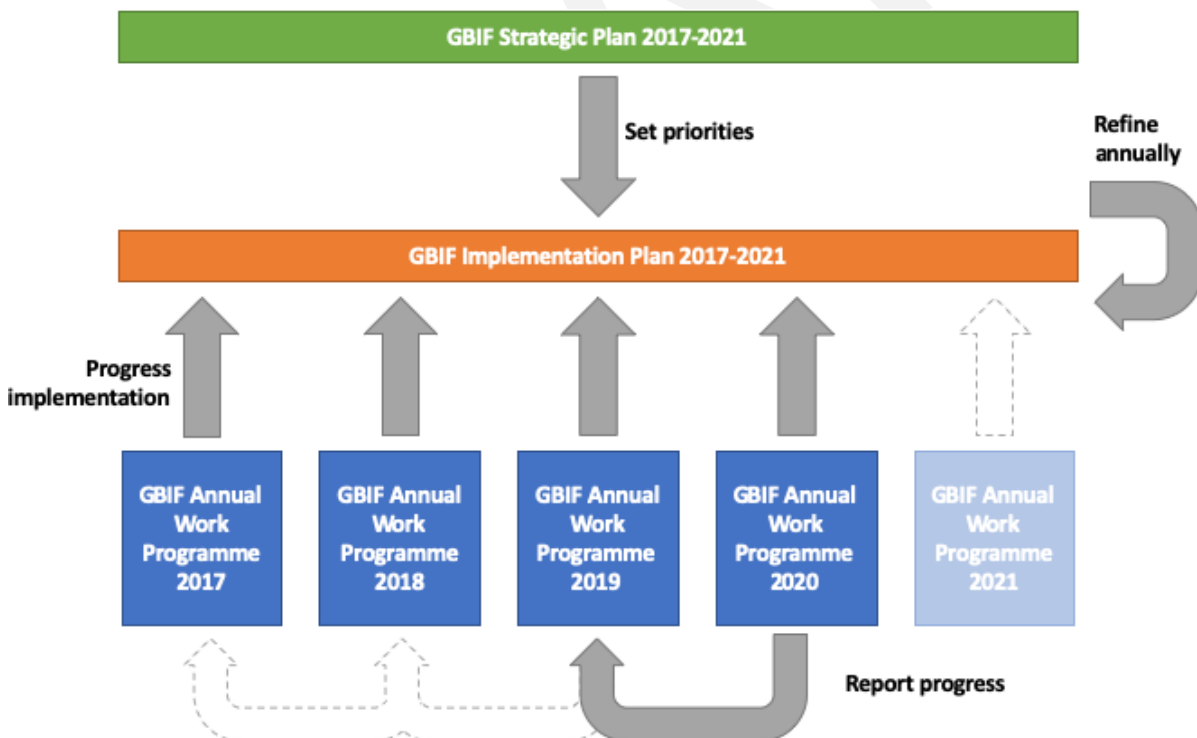
The relationship between these documents is represented by the following graphic.

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 2019 and planned activities for 2020 (the 2020 Annual Work Programme).

The Annual Work Programme 2020 is summarized below. The revised Implementation Plan begins on page 18.

## GBIF Annual Work Programme 2020

Information on activities planned for 2020 is included under each of the Activity sections included in the Implementation Plan, beginning on page 18.

For convenience, this information has also been collated on page 6 as an overview of the items for which approval is sought from the Governing Board at GB26.

---

### Financial situation

As of 24 July 2019, contributions are behind the pace of 2018. Current and imminent contributions are at the 89% of total expected contributions. No payments for former years are expected in 2019. Mid-year 2019 budget changes approved by the Executive Committee increased the overall WP2019 budget. Items added mid-year include:

- Increased funding for Capacity Enhancement Support Programme (CESP) from **€80,000 to €100,000**
- Added funding for Biodiversity Information Fund for Asia (BIFA) capacity enhancement workshop from WP (**€45,000**) to allow more funds for projects in a competitive call with strong proposals for mobilizing data from Asian institutions
- Consultant to explore opportunities for external funding (**€22,000**)
- Consultant contract on the *alliance for biodiversity knowledge* and COL+ (**€30,000**)
- Contract to complete validation and publication of at least 175 national invasive species checklists through the Global Register of Introduced and Invasive Species (GRIIS), managed by the IUCN Invasive Species Specialist Group (**€75,000**)
- Two workshops to implement taxonomy in GBIF/COL+ (**€50,000**)

As a result of the past year's increased staffing and increased Work Programme expenditures requested by the GB, GBIF is expected to close 2019 with net capital of around **€992,558**.

The initial phase of the Biodiversity Information for Development (BID) programme will be completed by the end of 2019; however, the funder—the European Commission Directorate-General for International Cooperation and Development (EU DEVCO)—has expressed interest in continuing the programme with new funding of €1.6m from 2020-2023. If the programme extension is finalized GBIF should again receive around **€120,000** in funding for work by GBIF Secretariat staff in 2020. Because this continued funding has not been finalized, this contribution is not included in the 2020 budget.

A fifth call for proposals under the BIFA programme will be issued in late 2019 or early 2020, using **€131,395** supplementary funds received from the Ministry of Environment, Government of Japan. Subject to agreement by the BIFA Steering Committee, we anticipate that **€20,000** of this supplementary funding is to be allocated to Secretariat staff coordination costs during 2020.

Lastly, the European Commission has funded the SYNTHESYS+ project “to unify European natural science collections, effectively transforming them into an integrated data- driven pan-European research infrastructure”. GBIF is funded for **€50,000** in 2020 (as well as **€20,000** in 2021 and **€20,000** in 2022) to advance joint planning for infrastructure components.

Given the level of capital held by GBIF and the uncertainty of contributions, the 2020 budget includes a smaller allocation (**€413,000**) for contracted activities in the work programme expenditure (i.e. non-salary costs allocated to calls, workshops, contracts, etc.) than proposed for 2019 at GB25, and **€300,000** less than the actual 2019 spend after budget revisions. Staffing levels do not change in the 2020 budget. The planned expenditure for 2020 is **€456,630** greater than the income projected in the budget. This is considered to represent a responsible balance between advancing GBIF's mission and reserving some funds against future risks (**€535,928** net capital forecast for the end of 2020).

This budget narrative also includes a second scenario that considers a brighter financial situation. It outlines priority areas of spending on contracted and workshop activities that would be submitted in subsequent 2020 budget revisions, should conditions allow.

Scenario 2 differs from the submitted budget in the following ways. It assumes an extension of the BID programme and its relevant activities and funds (including approximately **€120,000** in salary subsidy). It also assumes an additional **€200,000** in external funds or increased contributions from voting participants. This budget includes **€660,000** for contracted and workshop activities in the work programme expenditures. The planned expenditure under scenario 2 for 2020 is **€409,940** greater than the income projected in the budget. This is also considered to represent a responsible balance between advancing GBIF's mission and reserving some funds against future risks (**€582,618** net capital forecast for the end of 2020).

Scenario	Baseline (#1)	Optimistic (#2)
Income Assumptions	Participants pay at a similar level	Modifications from Baseline: BID project renewal and additional Voting Participants €200K/year
WP expenditure	<b>€413,000</b>	<b>€660,000</b>
2020 reserve	<b>€535,928</b>	<b>€582,618</b>

## Planned expenditure

In addition to salary and operating costs (including support for GBIF Governance mechanisms), funding is reserved in the Work Programme 2020 for the following activities:

Baseline 2020 Work Programme budget for non-salary items:

### Priority Area 1

- Following the recommendations of the community-based Documentation Editorial Panel, continue to create new documentation supporting participation in GBIF, based on late 2019 review and prioritization framework (**€30,000**).
- Develop joint training resources combining GBIF data mobilization and DNA barcoding, support for pilot workshop with Barcode of Life Data Systems (BOLD), Convention on Biological Diversity (CBD) and GBIF (**€20,000**).
- Consolidate online e-learning platform with training resources, to make the data mobilization, data use for decision making and nodes training modules fully functional as self-instruction modules, using external contracts for some components as necessary; develop a strategy to meet future demand for on-site, moderated and self-instructed training. Work will involve collaboration with key partners: nodes, the pool of trainers and mentors, Nodes Steering Group and numerous international partner projects (**€20,000**).
- Support for 2020 Regional Nodes Meetings (**€60,000**). The Secretariat will investigate possible synergies with BID phase 2 regional meetings, if funding is forthcoming, as well as with a BIFA workshop in Asia.
- Support further development of the Living Atlases community (**€30,000**).
- Support the alliance for biodiversity knowledge by working with steering committee on a governance structure and prioritizing the alliance efforts. Expand communication effort through conferences and the ambassador network (**€10,000**).
- Launch a new 2020 call for proposals under the Capacity Enhancement Support Programme (**€80,000**).

### **Priority Area 2**

- Maintain and update processes for constructing the GBIF taxonomic backbone. **€108,000** has been allocated in the budget to support GBIF contractor costs.

### **Priority Area 3**

- Carry forward the proposed 2019 activity of data mobilization of vectors and hosts of human diseases to 2020. The Secretariat will establish an expert group (**€25,000**) to identify priority needs for biodiversity data supporting disease research, critical gaps in availability of such data in GBIF.org, and potential sources of data to fill these gaps.

### **Priority Area 4**

- Explore bidirectional data linking and synchronization with data management systems and publishers to achieve faster and more accurate mutual updates on data improvements and annotations (**€10,000**).

### **Priority Area 5**

- Contract a study on use of GBIF-mediated data in the academic literature, with the goal of identifying areas of impact, links to research funding sources and networks of researchers. This information will be used to fine-tune GBIF value proposition message and align future Work Programmes to achieve greater impact (**€20,000**).

### **Other Funds**

- Supplementary funds: Implement a fifth call for proposals (**€110,000**) under the Biodiversity Information Fund for Asia (BIFA), reflecting priorities agreed by Asian nodes.
- Governance funds: Launch 2020 calls for the Ebbe Nielsen Challenge and Young Researchers Awards (**€44,000**), considering recommendations from the Science Committee based on the 2019 programmes. The Science Committee is exploring potential external funds for these awards.

If funding allows, the Secretariat will request that the following items be incorporated into the 2020 budget with approval of the GBIF Executive Committee.

#### **• Priority Area 1**

- Increase alliance for biodiversity knowledge activity to €20,000, Living Atlases to **€40,000** and CESP to **€100,000**.
- Create an additional funding priority area alongside the CESP call for GBIF node regional outreach activities as outlined in the Regional Participation Strategy (**€40,000**).
- Continue development and implementation of a strategic plan for targeting external funding streams for capacity enhancement projects (**€22,000**).
- Initiate engagement, data mobilization and capacity enhancement campaign for Arabic-speaking regions (**€20,000**).

#### **• Priority Area 2**

- Redesign the GBIF Integrated Publishing Toolkit (IPT) to support emerging data standards, explore integrations with quality control routines and to address infrastructural needs (ability to install locally, use a GBIF hosted solution or connect to a third-party repository). If funds allow **€50,000** for an external contractor.

#### **• Priority Area 4**

- Increase bidirectional linking activity to **€20,000**.

- **Priority Area 5**
  - Engage with academic based projects that use or could use better GBIF data in their pipelines and protocols. If funding allows hold a workshop to stimulate use (€25,000).
  - Hackathon and workshop to align GBIF with post-2020 Biodiversity Framework (€40,000). The event would develop tools from CESP-CHM project to improve links to data from GBIF.
- **External Funds**
  - Based on a successful negotiation with EU DEVCO, the Secretariat will continue with a BID call for proposals.

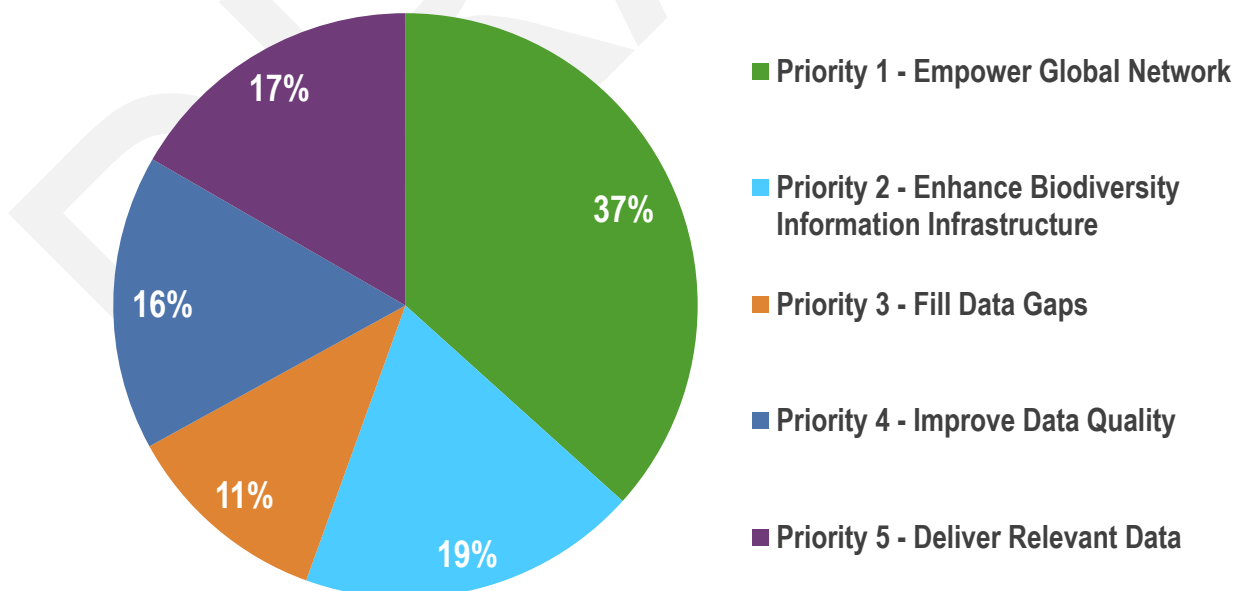
### Secretariat staffing

Staff increases in recent years have enabled the Secretariat to maintain an increasing level of service. We currently have 27 staff (three of whom work part-time) and are in the process of recruiting an additional informatics position. In addition to the Director and Deputy Director, the Secretariat is now structured as four teams, as follows:

- Participation and Engagement (seven staff)
- Data Products (four staff)
- Informatics (seven staff, one contractor, one additional staff position to be filled)
- Administration (six staff, plus student support)

All staff work towards accomplishing the goals of the five Priority Areas in the 2017-2021 Strategic Plan, with most staff working towards multiple areas. The Secretariat does not currently allocate salary to the Work Programme Priority budget areas. The 2020 budget Work Programme categories only represent non-salaried items such as external contracts and workshops. During 2019 our estimate of all staff work the across five Priority Areas is represented below.

### Staff effort across Work Programme



## Overview of 2020 Work Items

<b>Activity 1a</b> - Focus on people.....	7
<b>Activity 1b</b> - Strengthen skills.....	7
<b>Activity 1c</b> - Equip Participant nodes .....	7
<b>Activity 1d</b> - Equip data publishers.....	8
<b>Activity 1e</b> - Expand national participation .....	8
<b>Activity 1f</b> - Plan implementation.....	8
<b>Activity 1g</b> - Coordinate resources.....	9
<b>Activity 2a</b> - Modernize data standards.....	9
<b>Activity 2b</b> - Deliver names infrastructure .....	10
<b>Activity 2c</b> - Catalogue collections .....	10
<b>Activity 3a</b> - Identify priority gaps .....	10
<b>Activity 3b</b> - Expand data streams .....	11
<b>Activity 3c</b> - Engage data holders .....	11
<b>Activity 3d</b> - Rescue datasets .....	11
<b>Activity 3e</b> - Liaise with journals.....	11
<b>Activity 4a</b> - Ensure data persistence.....	11
<b>Activity 4b</b> - Assess data quality .....	11
<b>Activity 4c</b> - Enable data curation .....	12
<b>Activity 5a</b> - Engage academia .....	12
<b>Activity 5b</b> - Document needs.....	12
<b>Activity 5c</b> - Support biodiversity assessment.....	12
<b>Activity 5d</b> - Assess impact.....	13

## Overview of 2020 Work Items

This following is a summary of all 2020 Work Items proposed in the 2020 update of the Implementation Plan below. It is intended to summarize the work items for which approval is sought from the Governing Board at GB26. Accordingly, it does not include the additional commitments made and reported by Participants.

---

### Activity 1a - Focus on people

- Explore possibilities of providing citation tracking from dataset level to individual level. Provide ability to claim one's data and promote use; explore building on the example of 'Bloodhound tracker' and moving it into core GBIF operations.
- Clearly communicate and demonstrate the benefits of ORCID use through citation tracking.
- Pending a decision on continuing the GBIF Community Forum, take steps to improve interactive engagement with the broader GBIF community, considering the views gathered during the 2019 survey.

---

### Activity 1b - Strengthen skills

- Develop joint training resources combining GBIF data mobilization and DNA barcoding, and support pilot workshop with BOLD, CBD and GBIF (**€20,000**).
- Explore opportunities for additional training workshops on data use for decision making.
- The BioDATA project will continue to train students in Armenia and in Ukraine in 2020, and plans for BioDATA II project are in place to further increase the geographic coverage in Northern and Central Asia and Southern Caucasus.
- Following the recommendations of the community-based Documentation Editorial Panel, continue to commission new documentation, based on late 2019 review and prioritization framework (**€30,000**).
- Consolidate online e-learning platform with training resources, to make the modules on data mobilization, data use for decision making and nodes training fully functional as self-instruction modules. Use external contracts for some components as necessary. Develop a strategy to meet future demand for on-site, moderated and self-instructed training. Work in collaboration with key partners, including GBIF nodes, the pool of volunteer trainers and mentors, Nodes Steering Group and numerous international partner projects (**€20,000**).

---

### Activity 1c - Equip Participant nodes

- Support for 2020 Regional Nodes Meetings (**€60,000**). The Secretariat will investigate possible synergies with BID phase 2 regional meetings, as well as with a BIFA workshop in Asia.
- The Secretariat will continue enhancing guidance documentation for nodes. This guidance will incorporate priorities in data mobilization, gap filling, private-sector engagement and products from CESP projects.
- The Secretariat will design and implement user interfaces and services necessary to support simple hosted portals. The initial deployment will be of the BID programme portal followed by exploration of country portals and an exemplar virtual natural history collection. Discussions will determine whether such services should be restricted to Voting Participant countries and/or institutions that contribute a fee to create institutional portals.
- Continue conversations to align codebases and infrastructure with ALA, iDigBio, DiSSCo and other potential partners. The *alliance for biodiversity knowledge* will be used as a platform for some of these interactions.
- Support further development of the Living Atlases community (minimum **€30,000**).

- Consolidate Biodiversity Open Data Ambassador programme with issuing of digital badges, listing ambassadors on country pages and capturing feedback on promotion of GBIF at events.

---

### Activity 1d - Equip data publishers

- Promote wider editing of the GBIF registry and the shared help desk activities; including node staff initiating and diagnosing dataset crawling / ingestion.
- Complete implementation (if not finished in 2019) and develop processes to allow open editing of shared vocabularies used in data interpretations of the GBIF ingestion pipelines (e.g. habitat types, occurrence status, etc...).
- Create system of list management, similar to bulk email, to communicate to a larger section of the GBIF community for compliance and notifications. This includes a twice-yearly mandatory communication with data publishers in compliance with General Data Protection Regulation (GDPR: EU privacy regulations) and explore services for publishers to opt-in to receive push notifications for new citations.
- Provide comprehensive guidance and support services to lower the technical threshold of data-hosting options. Clearly document the benefits and implications of each option including aspects of operational cost, deployment model (local/cloud/GBIF-provided) and expectation of users. Use of volunteer mentors will be promoted to enhance help desk services.

---

### Activity 1e - Expand national participation

- The GBIF Secretariat will continue to identify funding opportunities for regional engagement, including inviting observers to regional node meetings. If budget allows, it will create a specific funding priority area alongside the 2020 CESP call for GBIF regional outreach activities, as outlined in regional participation strategies developed across the network.
- The Secretariat will develop and implement engagement, data-mobilization and capacity-enhancement plans for greater participation in all global regions, with an emphasis on the underrepresented areas. This effort will be aided by capacity enhancement workshops, including through a second phase of BID and other supplementary funding that becomes available.
- The internationalization of GBIF.org user interface, content and guidance materials will continue. Assuming the 2019 commitment for six UN languages is complete, this work will focus on maintenance and updates and will allow the addition of more languages on demand, and translating a wider set of content and materials, funded through both specific projects and volunteer efforts.
- Engagement, data mobilization and capacity enhancement campaign for Arabic-speaking region (€20,000, if funds allow).

---

### Activity 1f - Plan implementation

- Convene SYNTHESYS+ workshops and integrate outcomes with relevant Work Programme activities.
- Continue implementation and refinement of the standardized annual information requests from Participant nodes and display up to date content on revised country pages of GBIF.org.
- Lead some and participate in all *alliance for biodiversity knowledge* virtual workshops that will be defined in key areas to increased alignment of GBIF with other networks and infrastructures.
- Manage and improve the virtual conferences infrastructure used by *the alliance for biodiversity knowledge*. Work with steering committee on a governance structure and prioritizing the *alliance* efforts. Expand communication effort through conferences and the



ambassador network (minimum **€10,000**). Provide technical, communication and administrative support for the *alliance*.

---

### Activity 1g - Coordinate resources

- Launch a 2020 call for proposals under a renewed CESP (minimum **€80,000**). If funding allows, this will be supplemented by a special category for regional outreach activities that target increased data mobilization in non-GBIF countries as recommended by Nodes Steering Group ((see Activity 1e).
- Based on a successful negotiation with EU DEVCO, the European Commission funder of BID, the Secretariat will continue with a BID call for proposals that responds to feedback received from community consultations, three regional closing meetings and the BID phase 1 closing meeting.
- Implement a fifth call for proposals (approximately **€110,000**) under the Biodiversity Information Fund for Asia (BIFA), reflecting priorities agreed by Asian nodes.
- Continue collaboration with the BioDATA (Norway > Eurasia) and Russia support (Finland > Russia) supplementary funding programs.
- Develop and implement strategic plan for targeting external funding streams for capacity enhancement projects. This will be based on 2019 external consultation and using the results of BID Phase 1 closing meetings. If budget allows an additional (**€22,000**) will be allocated to this effort. The strategy will work with all GBIF regions to target potential funding streams that support additional capacity enhancement for data mobilization and use, building on the BID and BIFA models.
- Make refinements to the newly selected grant management system, FLUXX, to streamline project calls, assessment, selection, implementation, budgeting and reporting, to ensure it meets the needs of the community.
- Launch 2020 calls for the Ebbe Nielsen Challenge and Young Researchers Awards (**€44,000**), considering recommendations from the Science Committee based on the 2019 programmes. Explore the possibility of soliciting additional donor/sponsor co-funding.
- Begin Secretariat planning to establish a workflow using digital documentation to develop the 2021 work programme and 2022-2026 strategic plan. This planning will address the recommendations of the 2019 20-year review.

---

### Activity 2a - Modernize data standards

- Modernizing data standards is a continuous Work Programme activity for a global infrastructure like GBIF. During 2020 we will focus on advancing and refining data models for Collections, Taxonomic Treatments, Sampling Events, Organisms, Specimens, Organisms, Citations and the linkages between them.
- Provide a set of data-exchange profiles for sharing data within GBIF that conforms with a unified information model that includes both existing and new standards as well as the necessary controlled vocabularies.
- Redesign the GBIF Integrated Publishing Toolkit (IPT) to support these profiles and to address infrastructure needs, such as the ability to support local installations or GBIF-hosted solution. If funds allow, **€50,000** for an external contractor.
- Provide documentation for the data model and for the associated services offered through GBIF.org.
- Review and redesign GBIF data management system to accommodate the unified information model as part of data ingestion, quality control and processing where necessary.
- Continue technical discussions with other data aggregators to seek closer alignment in practice and, as far as possible, implementation of aggregation and indexing processes.

- Demonstrate improvements of information in GBIF.org and hosted national portals in specimen-level information, links to material citations, and links between specimens and sequence data from sources such as BOLD.
- Explore approaches for adding a phylogenetic/evolutionary dimension to the GBIF taxonomic backbone. Pilot phylogenetic browsing capabilities of occurrence data.
- Open discussion with GB participants to provide project funders with an overview of the resulting value relating to their investment (e.g. data mobilization, publications).
- In collaboration with international partners, explore the desirability and scope of “catalogue services” that are targeted specifically at physical specimen collections. Examples could include displaying duplicate or derived specimens across collections, type information, citations in taxonomic treatments and trait data.
- Explore options for displaying occurrence data from long-term sampling sites, piloting with projects like BIOSCAN 2 and/or Norwegian ecological datasets.

---

### Activity 2b - Deliver names infrastructure

- Maintain and update processes for constructing the GBIF taxonomic backbone, including monitoring the content and helping to prioritize editorial effort. **€108,000** has been allocated in the budget to support GBIF costs. This work is in collaboration with the Catalogue of Life.
- Implement a process enabling key checklists to be used in filtering occurrence data, such as Red Listed species and invasive alien species.
- Consult with relevant regulatory agencies, such as the European Environment Agency (EEA), for guidance on which legislative checklists should be incorporated to increase the relevance of COL+ to governments.
- Explore feasibility of supporting national taxonomies for exploring GBIF occurrence data to better enable national level reporting.
- Develop and pilot a process that allows qualified users to collaborate and edit sectors that contribute to the GBIF backbone taxonomy, aimed at reducing the delays before such edits appear on occurrence records from months to days.

---

### Activity 2c - Catalogue collections

- Based on community consultation, build mechanism to synchronize Global Registry of Scientific Collections (GRSciColl) with other catalogues.
- Improve linkages between collections, institutions and occurrences (or specimen) objects indexed by GBIF.
- Develop the user interfaces and services necessary to support a collection catalogue system.
- Work with the community to ensure the content is fit for use, and promote community editing of the registered content.

---

### Activity 3a - Identify priority gaps

- Continue work on items initiated in late 2019, with an emphasis on developing actionable guidance for data publishers and nodes, integration of user needs into prioritization for data mobilization, and data search analysis.
- Continue to improve visualizations of GBIF-mediated data that identify gaps by engaging in interactive community consultations. This work is a candidate for curated discussion through the *alliance for biodiversity knowledge* prior to implementation on GBIF.org.

---

### Activity 3b - Expand data streams

- Enhance the data exchange standards for sampling-event data, collaborating with partners that generate data to provide sources for filling current gaps. This work aims to establish partnerships with long-term monitoring communities.
- Improve linkages between records originating from museums and BOLD in order to link information that is currently treated as two occurrences.
- Carrying over the proposed 2019 work item, mobilize data on vectors and hosts of human diseases. Establish an expert group (**€25,000**) to identify priority needs for biodiversity data supporting disease research, critical gaps in availability of such data in GBIF.org, and potential sources of data to fill these gaps. The campaign will use this analysis to engage directly with relevant data holders, support data publication through GBIF and inform data mobilization priorities for use by nodes, publishers and funders (see Activity 3a).
- Continue linking and integration of sequence-based data streams.

---

### Activity 3c - Engage data holders

- Continue work with DiSSCo project team to maximize opportunities for mobilizing collections data from European institutions, including in countries not yet participating in GBIF.
- Work with iNaturalist, iDigBio and nodes community to maximize opportunities for public engagement in GBIF data mobilization.
- Develop private-sector data mobilization guidance and training. Promote revised guidance on mobilizing EIA data and run training programme for private sector consultants at IAIA conference in Spain 2020, and incorporate guidance from CESP project in nodes guidance package.

---

### Activity 3d - Rescue datasets

- Continue to implement workflow for prioritizing and drawing upon potential data sources reported through the 'dataset catcher' tool, including involvement of nodes, mentors and crowdsourced solutions.
- Roll out a workflow for 'Suggest a dataset' processing.

---

### Activity 3e - Liaise with journals

- Work with journals to make it easier for article authors to deposit supplementary data in formats suitable for GBIF publication.
- Establish model guidelines for data deposition and citation and develop compliance criteria for distribution to publishers. Build on recent developments of COPDESS and Research Data Alliance with GBIF-specific guidance. Develop additional communication materials to describe the benefits of DOI-citation best practices and work with journals on implementation.

---

### Activity 4a - Ensure data persistence

- Continue revision and documentation of flagging routines used in GBIF data ingestion pipelines.

---

### Activity 4b - Assess data quality

- Review, consolidate and update existing documentation for data publishers. In particular, provide clear guidance on minimum requirements for published data.
- Develop metrics to track the completeness of core data elements and the degree to which supplied content is appropriate.

- Supply clear indicator measures for the completeness and usability of data as part of GBIF.org dataset pages, based on examples such as the GEOLabel data branding model.
- Extend data-quality assessment to include aspects only detectable above the level of individual records.
- Assess the patchiness of indexed data (geographical clustering, misleading accuracy or precision of coordinates), including evaluation of the apparent causes of data patchiness and include measures of data patchiness in the data index, at both dataset and record level in the data index.
- Ensure that users of data are able to identify datasets or records that do not fulfil their criteria for geo-accuracy, whether they are accessing data through facets in the GBIF.org, via the API or in downloads.

---

#### Activity 4c - Enable data curation

- Continue to explore the use of the GBIF data index to support stable persistent resolvable identifiers for all specimens and occurrence records.
- Explore bidirectional data linking and synchronization with data management systems and publishers to achieve faster and more accurate mutual updates on data improvements and annotations (minimum **€10,000**).

---

#### Activity 5a - Engage academia

- Conduct survey on how GBIF informatics is incorporated into relevant graduate and undergraduate curricula to identify gaps and opportunities and key entry points. Use survey findings to prepare a campaign that could roll out relevant resources through GBIF nodes in 2021.
- Engage with academic-based projects that use or could use better GBIF data in their pipelines and protocols. If funding allows, hold a workshop to stimulate use (**€25,000**).

---

#### Activity 5b - Document needs

- Explore the creation of lightweight, customized website landing pages to address thematic interests on GBIF.org. The pages will increase flexibility to include curated, more informative clustered information, ensuring that users have access to both broad search results and prioritized views of data and information.
- Continue to implement recommendations of past expert user groups. In particular, build on the incorporation of GRIIS checklists to support richer, more targeted information on invasive and alien species, e.g. by highlighting documented occurrences with a relevant IAS status by country. Explore best options to identify and alert users of new occurrences of potential invasive species. In addition, review and address the most feasible and valuable recommendations of the agrobiodiversity group, (see 5c below) with support of continued community involvement.

---

#### Activity 5c - Support biodiversity assessment

- Explore [BIOSCAN](#) as a pilot for visualizing sampling-event data. Pending advances in the revised data model, improve the representation of sampling-event data, particularly abundance measures. Identify sources to support a consistent, controlled list of sampling protocols.
- Continue work with GEO BON to establish needs to deliver relevant data in suitable formats to support ongoing community-agreed implementation of Essential Biodiversity Variables (EBV). Support further community scoping as and if appropriate within the *alliance for biodiversity knowledge* framework.

- If funds are available, the Secretariat will engage visiting scientists from the assessment/modelling community for part-time secondments to improve shared understanding of data-delivery needs for the EBV and associated communities.
- Develop and roll out guidance for use of GBIF-mediated data in species risk assessments. Continue work based on the IUCN memorandum of cooperation and proposed training event at the 2020 IUCN World Conservation Congress in Marseille.
- Host hackathon to align GBIF with post-2020 Biodiversity Framework (**€40,000** if available). The event would develop tools from the CESP CHM project aimed at improve CBD links to GBIF-mediated data. As post-2020 framework/goal structure emerges, develop comprehensive guidance on the role of primary data to support targets and indicators.

---

#### **Activity 5d - Assess impact**

- Implement recommendations of the 20-year review and use it as a basis for the development of the next five-year strategic plan for the period 2022-2026.
- Contract out study of use of GBIF-mediated data in the academic literature with a goal to identify areas of impact, links to research funding sources and networks of researchers. This information will be used to fine-tune GBIF value proposition message and align Work Programme to further impact (**€20,000**).

## 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 2020 version of the document includes reports from a few GBIF Participants on their 2019 activities and 2020 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.

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 2020 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 2020 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 are 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 within each section describing the Activity.

DRAFT

## 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**

The Secretariat informatics team has begun investigating citation tracking of individual records, and developing the ability for users to ‘claim’ data they have contributed, to promote use and to demonstrate the value proposition of GBIF. User profile work is scheduled to begin in late 2019 and early discussions have been initiated with GBIF Canada around lessons learned in claiming specimens through <https://bloodhound-tracker.net> and how to integrate that tool into GBIF.

The Secretariat has been carrying out an evaluation of the GBIF Community Forum (<https://discourse.gbif.org/>) during its first year of operation, to inform a decision on whether to continue maintaining the platform. Preliminary metrics suggest that use of the forum by the community is well below the targets set as benchmarks when the platform was established as a successor to the GBIF Community Site. 336 posts were made over a year with 87 new topics started by non-Secretariat staff. On the other hand, the time spent by Secretariat staff on moderation/curation was not significant.

**2020 work items**

- Explore possibilities of providing citation tracking from dataset level to individual level. Provide ability to claim one's data and promote use; explore building on the example of 'Bloodhound tracker' and moving it into core GBIF operations.
- Clearly communicate and demonstrate the benefits of ORCID use through citation tracking.
- Pending a decision on continuing the GBIF Community Forum, take steps to improve interactive engagement with the broader GBIF community, considering the views gathered during the 2019 survey.

DRAFT

---

## 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 strengthen key skills, GBIF needs to develop and maintain a comprehensive set of clear reference information and training materials that support all GBIF audiences. Since 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). 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
- 1.b.iv Establish system for reviewing, prioritizing, updating and creating new documentation to support GBIF data publishers and users

### **Progress**

The Secretariat continues to improve and develop new current training materials as well as standardized arrangements for setting up training workshops based on re-usable curricula in key skills related to GBIF participation. This work is accomplished in close collaboration with trainers and mentors from the GBIF community, who contribute their time and expertise at little or no cost to GBIF's core budget. The capacity enhancement programmes (e.g. BID, BIFA and CESP and GBIF involvement in other externally-led projects) are major motivating factors for the Secretariat to improve materials and support a growing community of practice.

Two ways to measure the value of this investment in generating training materials are, first, the increased number of volunteer mentors and trainers across the network, which now stands at 93 in

total, with 58 involved in the BID programme. Second, these programmes have sparked a sharp increase in both the number of data publishers and the amount of data from countries and regions not previously engaged in biodiversity data sharing, with the BID programme adding 90 new publishers from sub-Saharan Africa, the Caribbean and the Pacific. Together, with organizations in these regions that previously shared data, the new publishers introduced through the BID programme have added nearly 1.2 million occurrence records, filling critical knowledge gaps by increasing in the percentage of data for a country being published within that country and adding records for 1,189 species listed as threatened in the IUCN Red List.

Multiple training events involving the Secretariat and community mentors and trainers occurred or are planned in 2019, including the following:

- ‘Train the mentors’ event to support data mobilization training within the BioDATA project covering Eastern Europe, the Caucasus and Central Asia (Belarus, February)
- Training course on data skills within the BioDATA project (Tajikistan, June)
- Training sessions on establishing a GBIF Participant Node, including a focus on the benefits of GBIF participation, at the BID regional meetings in the Caribbean (Trinidad and Tobago, June), Pacific (Fiji, August) and Africa (Cameroon, September)
- Capacity enhancement workshop on data mobilization for projects selected under the Biodiversity Information Fund for Asia (BIFA), as well as for GBIF nodes in Asia (Viet Nam, July)
- Training on development of GBIF Participant Nodes, offered to all node managers as part of the GB26 series of meetings (Netherlands, October)
- Capacity enhancement workshop on data use for decision making for participants in GIZ-funded project to improve biodiversity information management in southeast Europe (Bosnia and Herzegovina, November/December to be confirmed)
- The BioDATA project focussed on developing skills in biodiversity data management and data publishing for undergraduate and postgraduate students from Tajikistan, Belarus, Ukraine, and Armenia. The project is run by the University of Oslo and GBIF Norway and exceeded expectations, with more than 8 mentors and 40 students trained on the GBIF publishing and data skills in two events that took place in Belarus and in Tajikistan.

As well as these training events directly involving the Secretariat, a large number of replicated workshops based on the same materials have been carried out in association with various funded projects as well as node activities. For example, BID participants have already replicated the data mobilization and data use workshops 66 times, extending the impact of the BID programme to nearly 1,500 new students within the corresponding regions.

The Secretariat has continued to update training materials for data mobilization, decision making, for use in on-site training and in development of online modules. For the data use curriculum, GBIFS has been working with IUCN to refine guidance on the use of species occurrence data in species Red List assessments. This is one of the areas of collaboration included in a Memorandum of Cooperation between IUCN and GBIF Secretariat, finalized and signed in July 2019.

The Secretariat has established a community-based Documentation Editorial Panel, aimed at prioritizing, developing and updating a core set of electronic guidance documents to support participation in the GBIF community. The panel includes membership from four global regions, and it will have held three meetings by August 2019. A consortium led by the VertNet team is producing an initial package of new and updated documentation in 2019, namely:

- Basic Guide to Using OpenRefine (in Spanish)
- Quick Guide to Georeferencing
- Best Practices in Georeferencing
- Georeferencing calculator

- Best Practices for Generalizing Sensitive Species Occurrence Data

### **2020 work items**

- Develop joint training resources combining GBIF data mobilization and DNA barcoding, and support pilot workshop with BOLD, CBD and GBIF (**€20,000**).
- Explore opportunities for additional training workshops on data use for decision making.
- The BioDATA project will continue to train students in Armenia and in Ukraine in 2020, and plans for BioDATA II project are in place to further increase the geographic coverage in Northern and Central Asia and Southern Caucasus.
- Following the recommendations of the community-based Documentation Editorial Panel, continue to commission new documentation, based on late 2019 review and prioritization framework (**€30,000**).
- Consolidate online e-learning platform with training resources, to make the modules on data mobilization, data use for decision making and nodes training fully functional as self-instruction modules. Use external contracts for some components as necessary. Develop a strategy to meet future demand for on-site, moderated and self-instructed training. Work in collaboration with key partners, including GBIF nodes, the pool of volunteer trainers and mentors, Nodes Steering Group and numerous international partner projects (**€20,000**).

DRAFT

## 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 help desk 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 help desk to assist and mentor node managers
- 1.c.iv Explore alternative options for lightweight national portal solutions, e.g. CKAN

### **Progress**

Based on discussions alongside the GB25 meeting in Ireland in 2018, the Kilkenny Accord developed a roadmap towards an open governance and funding model for the Living Atlas community. GBIFS established and has overseen contracts with an Administrative and Technical coordinator to support the community (**€40,000**) in 2019. GBIFS assisted in the ALA Advanced Workshop in Paris to help adopters move towards Spatial Portal capabilities. The Secretariat informatics team held a technical meeting with the ALA at Canberra in March to explore possibilities to collaborate, with the Registry/Collectory and a hosted data indexing process identified as an opportunity to reduce the costs of operation for Australia and show the value proposition of GBIF.

In 2019 the Secretariat started exploring the potential for a lightweight portal solution that has been requested by several participants to lower the technical threshold. A concept of the simple hosted solution (static site generator and occurrence search) has been demonstrated to the Science Committee and Nodes Steering Group, and a visual design is in development for GB26. A showcase portal containing data generated through projects funded by the BID programme in Africa, Caribbean and the Pacific has been chosen as the initial project in 2019.

The informatics team opened the GBIF Registry administration (<https://registry.gbif.org>) and piloted editing by external users (starting with GBIF Norway node staff). The administration console provides editing and registration capabilities of the core entities (institutions, datasets, etc.) and also provides oversight and control of the dataset crawling infrastructure and diagnostics. This is part of

a strategy to progress towards becoming a more open platform for others to use by lowering the technical threshold through enabling infrastructure.

A global nodes meeting and training event is taking place in conjunction with GB26 in Leiden, Netherlands in October (see 1b above). The training is a two-day course focussed on the functional capacities involved in node development: positioning a node in the broader landscape, identifying key stakeholders, developing strategic and implementation plans and running effective training workshops. The online materials are being made public by the end of the 2019.

The GBIF network has initiated a programme creating a new role, the Biodiversity Open Data Ambassador, aimed at supporting a corps of experts who promote free and open access to biodiversity data. During its initial development phase without significant promotion, 42 professionals from 19 countries, representing each of GBIF's six regions, volunteered to become Biodiversity Open Data Ambassadors. The programme is intended to recognize and support the active role that ambassadors have alongside the network's node managers and staff, formal delegates, committee members and Secretariat staff. The GBIF Secretariat continues to develop resources to help ambassadors' efforts, including presentation slides, talking points, posters, and support materials in both digital and print formats.

### **2020 work items**

- Support for 2020 Regional Nodes Meetings (**€60,000**). The Secretariat will investigate possible synergies with BID phase 2 regional meetings, as well as with a BIFA workshop in Asia.
- The Secretariat will continue enhancing guidance documentation for nodes. This guidance will incorporate priorities in data mobilization, gap filling, private-sector engagement and products from CESP projects.
- The Secretariat will design and implement user interfaces and services necessary to support simple hosted portals. The initial deployment will be of the BID programme portal followed by exploration of country portals and an exemplar virtual natural history collection. Discussions will determine whether such services should be restricted to Voting Participant countries and/or institutions that contribute a fee to create institutional portals.
- Continue conversations to align codebases and infrastructure with ALA, iDigBio, DiSSCo and other potential partners. The *alliance for biodiversity knowledge* will be used as a platform for some of these interactions.
- Support further development of the Living Atlases community (minimum **€30,000**).
- Consolidate Biodiversity Open Data Ambassador programme with issuing of digital badges, listing ambassadors on country pages and capturing feedback on promotion of GBIF at events.



---

## 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 help desk 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**

The Integrated Publishing Toolkit (IPT) is being maintained to deal with bug fixes, translations etc., and version 2.4 was released in 2019. Review of requirements for a revised IPT is on hold until there is greater clarity on how broad a data model for GBIF should be (see Activity 2a). With a new model in place, GBIF will design appropriate tools to map and publish data. The cloud-based IPT infrastructure maintained by the Secretariat continues to provide a fall-back solution for publishers unable to host their own installation or to find a third-party hosting option. As of June 2019, this option was being used by nine publishers associated with the BID programme and one publisher associated with the BIFA programme, sharing a total of 100 datasets. Cloud-hosted IPTs are expected to become more widely used in the coming months, especially with the use of volunteer mentors to provide help desk support for users of this option.

**2020 work items**

- Promote wider editing of the GBIF registry and the shared help desk activities; including node staff initiating and diagnosing dataset crawling / ingestion.
- Complete implementation (if not finished in 2019) and develop processes to allow open editing of shared vocabularies used in data interpretations of the GBIF ingestion pipelines (e.g. habitat types, occurrence status, etc...).
- Create system of list management, similar to bulk email, to communicate to a larger section of the GBIF community for compliance and notifications. This includes a twice-yearly mandatory communication with data publishers in compliance with General Data Protection Regulation (GDPR: EU privacy regulations) and explore services for publishers to opt-in to receive push notifications for new citations.
- Provide comprehensive guidance and support services to lower the technical threshold of data-hosting options. Clearly document the benefits and implications of each option including aspects of operational cost, deployment model (local/cloud/GBIF-provided) and expectation of users. Use of volunteer mentors will be promoted to enhance help desk services.

DRAFT

---

## 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**

The internationalization of the GBIF.org user interface and core content elements, in the six UN languages as well as selected additional languages, is expected to be complete by the end of 2019.

Technical adjustments have been made to GBIF.org to accommodate right-to-left script needed for Arabic and other languages. A volunteer Arabic-speaking network is being developed. Translations of GBIF.org into Russian and Chinese (both simplified and traditional) are far progressed and are being launched later in 2019, adding to the existing interfaces and core content in English, French, Spanish, Portuguese and Japanese.

Three 2019 regional meetings within the BID program are providing opportunities to strengthen and expand national participation in the Caribbean, Pacific and sub-Saharan Africa. The Caribbean meeting in Trinidad and Tobago was attended by over 40 participants from across the Latin American and Caribbean region, including a strong delegation from the host country coming from various institution and decision-making bodies. Strong attendance is expected in the other regional meetings later in 2019.

The main outputs of the meetings are regional statements on regional priorities (e.g. from [the Caribbean](#)), and roadmaps for ensuring long term sustainability of biodiversity data sharing. The final day of the event focuses on node management training including the benefits of GBIF participation and being part of the GBIF nodes community.

During 2019, Angola and Belarus joined the community as Associate Participants, and Chile returned to Voting Participant status after a period as an Associate Participant. Brazil communicated its intention to become a Voting Participant. Israel, Indonesia and India dropped to observer status

after passing the maximum period as Associate Participants. Argentina, Ghana, Kenya, Uganda and the United Kingdom moved from Voting to Associate status on 1 January 2019 due to non-payment of their contributions for two consecutive years. A complete and updated progress assessment of the GBIF status and engagement in all countries is being presented at the GB26 meeting in the Netherlands.

**2020 items**

- The GBIF Secretariat will continue to identify funding opportunities for regional engagement, including inviting observers to regional node meetings. If budget allows, it will create a specific funding priority area alongside the 2020 CESP call for GBIF regional outreach activities, as outlined in regional participation strategies developed across the network.
- The Secretariat will develop and implement engagement, data-mobilization and capacity-enhancement plans for greater participation in all global regions, with an emphasis on the underrepresented areas. This effort will be aided by capacity enhancement workshops, including through a second phase of BID and other supplementary funding that becomes available.
- The internationalization of GBIF.org user interface, content and guidance materials will continue. Assuming the 2019 commitment for six UN languages is complete, this work will focus on maintenance and updates and will allow the addition of more languages on demand, and translating a wider set of content and materials, funded through both specific projects and volunteer efforts.
- Engagement, data mobilization and capacity enhancement campaign for Arabic-speaking region (€20,000, if funds allow).

---

## 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**

GBIF leads work package five of the European initiative SYNTHESYS+ project, which includes running three workshops. The focus of these workshops, defined by the global community, is to be specified in Q3 2019, but is anticipated to include activities around collection cataloguing, richer specimen models and tracking of citations of specimens, collections and the individuals, all of which are also priorities for GBIFS (see also activity 1a).

The Secretariat has been involved at many organization levels (Steering Committee, Programming committee) of the Biodiversity\_Next conference held in conjunction with GB26. Secretariat staff are involved in over 20 talks and organizing several sessions that relate directly to the Work Programme. Involvement has begun with a similar meeting scheduled to take place in September 2020 in Virginia, USA.

The Secretariat is developing standardized annual information requests based on the GBIF Strategic Plan and Implementation Plan structure aimed at capturing relevant updates and work plans from Participant nodes. The work is intended to improve the visibility of the GBIF Participants, and especially of the nodes, on GBIF.org and to better reflect the network as a whole, particularly with regard to the active contributions toward work programme implementation.

The *alliance for biodiversity knowledge* was established as an outcome of the 2018 GBIC2 conference hosted by GBIF in Copenhagen. The Secretariat has reconstituted the GBIC2 steering committee as the steering committee for the *alliance*. In conjunction with external stakeholders, the Secretariat is developing a virtual workshop infrastructure and will be hosting the first of a series of virtual workshops under this infrastructure later in 2019. The Secretariat is also driving efforts to develop and disseminate *alliance* communication materials and to build expertise networks for managing virtual workshops and networking models.

### **2020 work items**

- Convene SYNTHESYS+ workshops and integrate outcomes with relevant Work Programme activities.
- Continue implementation and refinement of the standardized annual information requests from Participant nodes and display up to date content on revised country pages of GBIF.org.
- Lead some and participate in all *alliance for biodiversity knowledge* virtual workshops that will be defined in key areas to increased alignment of GBIF with other networks and infrastructures.
- Manage and improve the virtual conferences infrastructure used by *the alliance for biodiversity knowledge*. Work with steering committee on a governance structure and prioritizing the *alliance* efforts. Expand communication effort through conferences and the ambassador network (minimum **€10,000**). Provide technical, communication and administrative support for the *alliance*.

DRAFT

---

## 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**

The initial five-year phase of the BID programme reached its final year with a flurry of activity, closing all of the 63 projects in sub-Saharan Africa, Caribbean and the Pacific supported by the European Union-funded programme. The Secretariat organized closing meetings and training events in all three regions focussed on showcasing the achievements of BID, planning regional coordination of future biodiversity data mobilization and use, highlighting the benefits of GBIF participation and establishing a GBIF Participant Node (see Activity 1b). A review of the impacts of BID in terms of data mobilization and use, capacity enhancement and sustainable outcomes, was commissioned following a competitive call and will be finalized later in 2019. An event to close the first phase of BID is being held in Brussels in late 2019, with the aim of summarizing the programme's results to the funders and other stakeholders, as well as setting priorities for future biodiversity informatics development using the model of BID to strengthen and expand data availability in all global regions.

Based on the perceived success of the first phase of BID, the Development and Cooperation Directorate of the European Commission (EU DEVCO, or EuropeAid) expressed interest in extending the programme with additional funding. Details of this extension are under negotiation, but it is expected that GBIF will have flexibility to expand the geographic scope of the programme beyond the three regions targeted in the first phase. The closing event of the first phase of BID will include engagement with other providers of overseas development assistance to explore the potential to supplement the funding from EU DEVCO for future phases of the programme.

During the April 2019 budget revision, €22,000 was added to hire an external consultant to identify potential funding streams from philanthropic organizations, and to help the Secretariat develop a strategic plan for a funding stream for these programmes. This work commences in late 2019.

The Biodiversity Information Fund for Asia (BIFA), based on supplementary funding provided by the Japanese Ministry of Environment, attracted a strong response in its fourth call for proposals for data mobilization projects. After a thorough two-stage selection process, nine projects were selected for funding. As with the last BIFA call, this year's call replicated the model developed under BID in providing a training workshop, held in Viet Nam, for all funded projects to acquire basic data publishing skills to support project implementation.

A further call under the Capacity Enhancement Support Programme (CESP) was launched in 2019, resulting in selection of five projects for funding, managed by nodes in Europe, North America, Africa and Latin America. Each of these projects contribute to activities in the 2020 work programme items:

- [Using the CBD Clearing-House Mechanism to strengthen biodiversity data acquisition and data sharing](#) aims to establish technical links between GBIF national nodes and the network of CBD national Clearing-House Mechanisms (5c. Support biodiversity assessment)
- [OpenPSD: Engaging the private sector to promote biodiversity data publication and use](#) will establish deeper connections between the participating nodes and private-sector sources of data (3c. Engage data holders)
- [Data Use for Decision Making Workshop: an Iberoamerican community call](#) replicates and applies training curriculum developed under the BID programme outside the ACP regions (1g. Coordinate resources)
- [Strengthening Zimbabwe's GBIF node through mentoring by GBIF Spain](#) follows a mandate to improve mechanisms for engaging new countries (1e. Expand national participation)
- [Tackling the spatial challenge in the Southern Cone through georeferencing training](#) is applying updated documentation on best practices for georeferencing data (1b. Strengthen skills)

The 2019 calls for GBIF's annual awards programmes—the Young Researchers Award (YRA) and the Ebbe Nielsen Challenge—are attracting many high-quality entries. The results of the YRA are being announced in late August or early September, while the winners of the Challenge incentive prize are being announced during the GB26 meeting.

The GBIF Secretariat has invested in Fluxx, a grants management system. The system will help streamline and automate processes in relation to the management of the grant programmes GBIF is involved in. This should save time for all involved and ensure that the processes are made more efficient. The GBIF deployment of Fluxx is currently in the development phase. A small group of Secretariat staff is testing and offering feedback to the Fluxx developers. Once all issues have been addressed, the testing will be opened up to a wider test group before the final sign-off. The process has somewhat longer than expected as the investment of time from the Secretariat side has been much larger than originally anticipated.

The Secretariat continues to offer support to BID data publishers at the technical and data content level, which improves data consistency and quality control of all data submissions, assuring adherence to the data publishing and quality guidelines. This work supports the external BID impact study. The help desk supports BID 'cloud' IPT data hosting options for publishers.



## 2020 work items

- Launch a 2020 call for proposals under a renewed CESP (minimum **€80,000**). If funding allows, this will be supplemented by a special category for regional outreach activities that target increased data mobilization in non-GBIF countries as recommended by Nodes Steering Group ((see Activity 1e).
- Based on a successful negotiation with EU DEVCO, the European Commission funder of BID, the Secretariat will continue with a BID call for proposals that responds to feedback received from community consultations, three regional closing meetings and the BID phase 1 closing meeting.
- Implement a fifth call for proposals (approximately **€110,000**) under the Biodiversity Information Fund for Asia (BIFA), reflecting priorities agreed by Asian nodes.
- Continue collaboration with the BioDATA (Norway > Eurasia) and Russia support (Finland > Russia) supplementary funding programs.
- Develop and implement strategic plan for targeting external funding streams for capacity enhancement projects. This will be based on 2019 external consultation and using the results of BID Phase 1 closing meetings. If budget allows an additional (**€22,000**) will be allocated to this effort. The strategy will work with all GBIF regions to target potential funding streams that support additional capacity enhancement for data mobilization and use, building on the BID and BIFA models.
- Make refinements to the newly selected grant management system, FLUXX, to streamline project calls, assessment, selection, implementation, budgeting and reporting, to ensure it meets the needs of the community.
- Launch 2020 calls for the Ebbe Nielsen Challenge and Young Researchers Awards (**€44,000**), considering recommendations from the Science Committee based on the 2019 programmes. Explore the possibility of soliciting additional donor/sponsor co-funding.
- Begin Secretariat planning to establish a workflow using digital documentation to develop the 2021 work programme and 2022-2026 strategic plan. This planning will address the recommendations of the 2019 20-year review.

## 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

The *alliance for biodiversity knowledge* is beginning to act as a platform to engage the biodiversity informatics community around community standards. GBIF is active in numerous significant existing networks that seek to address these needs in parallel. This work is ongoing into 2020 and beyond.

At the core of this is the Biodiversity Information Standards (TDWG) community. GBIF continues to participate in open TDWG discussions around ABCD/DwC alignment (<https://doi.org/10.3897/biss.3.37491>) and recognize many other complementary activities and discussions on a biodiversity knowledge graph by partners. Notably:

- Through Plazi taxonomic treatment data are becoming connected
- Pensoft have piloted the OpenBioDiv system to better connect information in scholarly publications with specimens (<https://doi.org/10.3390/publications7020038>)
- The Ozymandias biodiversity knowledge graph pilot by Rod Page (<https://ozymandias-demo.herokuapp.com>)

Proposed SYNTHESYS+ workshops are to focus on modernizing standards activities that seek to improve the representation of information such as 1) Citation and Provenance models for collections data, including links between specimens, individuals and literature and their relationships to DOI, ORCIDs and other important open identifier issuing systems; 2) Information model for representing Natural History Collections including the TDWG natural Collections Descriptions standard and integrations with other collections catalogues and 3) Information model for representing Specimens; reviewing collection management systems and developing an information model with data classes representing all the asset types and linked information of importance in building a fully interconnected virtual natural history collection

During 2019, the informatics team has continued the redesign and implementation of GBIF data ingestion pipelines. Data growth required significant changes to the backend to ensure GBIF can 1) continue to grow with data volume 2) accommodate new feature deployments that require full data reprocessing and 3) look to expand data content types.

## 2020 Work Items

- Modernizing data standards is a continuous Work Programme activity for a global infrastructure like GBIF. During 2020 we will focus on advancing and refining data models for Collections, Taxonomic Treatments, Sampling Events, Organisms, Specimens, Organisms, Citations and the linkages between them.
- Provide a set of data-exchange profiles for sharing data within GBIF that conforms with a unified information model that includes both existing and new standards as well as the necessary controlled vocabularies.
- Redesign the GBIF Integrated Publishing Toolkit (IPT) to support these profiles and to address infrastructure needs, such as the ability to support local installations or GBIF-hosted solution. If funds allow, **€50,000** for an external contractor.
- Provide documentation for the data model and for the associated services offered through GBIF.org.
- Review and redesign GBIF data management system to accommodate the unified information model as part of data ingestion, quality control and processing where necessary.
- Continue technical discussions with other data aggregators to seek closer alignment in practice and, as far as possible, implementation of aggregation and indexing processes.
- Demonstrate improvements of information in GBIF.org and hosted national portals in specimen-level information, links to material citations, and links between specimens and sequence data from sources such as BOLD.
- Explore approaches for adding a phylogenetic/evolutionary dimension to the GBIF taxonomic backbone. Pilot phylogenetic browsing capabilities of occurrence data.

- Open discussion with GB participants to provide project funders with an overview of the resulting value relating to their investment (e.g. data mobilization, publications).
- In collaboration with international partners, explore the desirability and scope of “catalogue services” that are targeted specifically at physical specimen collections. Examples could include displaying duplicate or derived specimens across collections, type information, citations in taxonomic treatments and trait data.
- Explore options for displaying occurrence data from long-term sampling sites, piloting with projects like BIOSCAN 2 and/or Norwegian ecological datasets.

DRAFT

---

## 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**

A new IT infrastructure for building and maintaining GBIF backbone taxonomy is near completion. This new infrastructure came from a collaboration with the Catalogue of Life (COL) and includes a clearinghouse for nomenclature and taxonomy. This infrastructure was designed to replace the GBIF Checklistbank. This infrastructure will be used by GBIF and COL and is available service for other initiatives. GBIF is working with COL to provide tools, including a web-based console for editors and technicians. The checklist is being monitored and will replace the GBIF backbone when suitably complete in coverage. The design and deployment of the new public Catalogue of Life website is scheduled for later in 2019.

The adoption of Operational Taxonomic Unit (OTU) identifiers from the UNITE fungal database and the Barcode of Life Database (BOLD) in the Catalogue of Life is scheduled for the second half of 2019.

## **2020 Work Items**

- Maintain and update processes for constructing the GBIF taxonomic backbone, including monitoring the content and helping to prioritize editorial effort. **€108,000** has been allocated in the budget to support GBIF costs. This work is in collaboration with the Catalogue of Life.
- Implement a process enabling key checklists to be used in filtering occurrence data, such as Red Listed species and invasive alien species.
- Consult with relevant regulatory agencies, such as the European Environment Agency (EEA), for guidance on which legislative checklists should be incorporated to increase the relevance of COL+ to governments.
- Explore feasibility of supporting national taxonomies for exploring GBIF occurrence data to better enable national level reporting.
- Develop and pilot a process that allows qualified users to collaborate and edit sectors that contribute to the GBIF backbone taxonomy, aimed at reducing the delays before such edits appear on occurrence records from months to days.

## 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 GRSciCol 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

In 2019 the GRSciCol database was migrated to a public website managed by GBIF (<https://gbif.org/grscicoll>) and editing capabilities provided through the enhanced, and open, [registry administration console](#). The migration and resurrection of GRSciCol identified that there are several other catalogues (CETAF, IndexHerbariorum) that need consideration; GRSciCol needs to evolve to synchronize with authoritative catalogues such as Index Herbarium and collection management systems directly. Therefore, GBIF continues involvement in the TDWG task group on the development of a collections description data standard (TDWG CD). This work continues with community consultation under SYNTHESYS+ and the *alliance for biodiversity knowledge* to plan future work.

Identifiers have been used over the years by the community and are referenced in 42M records. These identifiers now resolve again (e.g. <http://biocol.org/urn:lsid:biocol.org:col:34994>)

### 2020 Work Items

- Based on community consultation, build mechanism to synchronize Global Registry of Scientific Collections (GRSciColl) with other catalogues.
- Improve linkages between collections, institutions and occurrences (or specimen) objects indexed by GBIF.
- Develop the user interfaces and services necessary to support a collection catalogue system.

- Work with the community to ensure the content is fit for use, and promote community editing of the registered content.

DRAFT



## 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**

During 2019 the Secretariat is developing actionable guidance for nodes, publishers and funders on addressing priority gaps through mobilization targets and strategies based on spatial, temporal, taxonomic and thematic dimensions of biodiversity data. This guidance document will indicate best practices on how to prioritize mobilization to develop national priorities and how to integrate those priorities into a mobilization strategy. The guide content undergoes assessments that can be updated upon knowledge gained for the GBIF communities in activities such as the global node events.

The Secretariat performed data analyses building on the input from the 2016 Ebbe-Nielsen Challenge. It further examined possible approaches of 'hunger-mapping' to the prioritization of data gaps, as well as analysing the availability and coverage of regional and national checklists of taxa that could support this approach, and the possibility of making use of ES50 maps as a measure of species richness. (data-blog.gbif.org).

## **2020 Work Items**

- Continue work on items initiated in late 2019, with an emphasis on developing actionable guidance for data publishers and nodes, integration of user needs into prioritization for data mobilization, and data search analysis.
- Continue to improve visualizations of GBIF-mediated data that identify gaps by engaging in interactive community consultations. This work is a candidate for curated discussion through the *alliance for biodiversity knowledge* prior to implementation on GBIF.org.

## 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**

2019 saw the continued growth of occurrence data on GBIF.org. Major new data types available are occurrence records derived from single sequence and metagenomic datasets of fungi and bacteria. GBIF is also ingesting occurrence data arising from barcode sequences deposited in the Barcode of Life Database (BOLD). GBIF presented at the Living Norway symposium targeting ecological datasets and is investigating protocols for sampling event data. GBIF is proactively investigating new data streams through outreach seminars at major events.

### **2020 Work Items**

- Enhance the data exchange standards for sampling-event data, collaborating with partners that generate data to provide sources for filling current gaps. This work aims to establish partnerships with long-term monitoring communities.
- Improve linkages between records originating from museums and BOLD in order to link information that is currently treated as two occurrences.
- Carrying over the proposed 2019 work item, mobilize data on vectors and hosts of human diseases. Establish an expert group (**€25,000**) to identify priority needs for biodiversity data supporting disease research, critical gaps in availability of such data in GBIF.org, and potential sources of data to fill these gaps. The campaign will use this analysis to engage directly with relevant data holders, support data publication through GBIF and inform data mobilization priorities for use by nodes, publishers and funders (see Activity 3a).
- Continue linking and integration of sequence-based data streams.

DRAFT

---

## 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**

The Secretariat has signed a Memorandum of Cooperation (MoC) with IUCN to further collaborations. The purpose of the MoC is to strengthen technical and institutional collaboration between GBIF and IUCN with a view to improving the visibility, timeliness and usefulness of data and information exchanged between the two networks, thus adding value to their respective products and services. The immediate objectives are to build upon past close work on red-listed and invasive species in data sharing, visualization and indicator development.

Contributions of the citizen science community to the [publication](#) of primary occurrence data are analysed through an algorithm-assisted process, tagging datasets for further processing and metrics generation. An update to the citizen science study from 2016 is given in the [GBIF data blog](#).

### **2020 Work Items**

- Continue work with DiSSCo project team to maximize opportunities for mobilizing collections data from European institutions, including in countries not yet participating in GBIF.
- Work with iNaturalist, iDigBio and nodes community to maximize opportunities for public engagement in GBIF data mobilization.
- Develop private-sector data mobilization guidance and training. Promote revised guidance on mobilizing EIA data and run training programme for private sector consultants at IAIA conference in Spain 2020, and incorporate guidance from CESP project in nodes guidance package.

---

## 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**

Explored metadata-only dataset publication of the datasets proposed through <https://www.gbif.org/suggest-dataset>. Currently developing categorization of datasets to improve implementation of the 'Suggest a dataset' tool in GitHub.

### **2020 Work Items**

- Continue to implement workflow for prioritizing and drawing upon potential data sources reported through the 'dataset catcher' tool, including involvement of nodes, mentors and crowdsourced solutions.
- Roll out a workflow for 'Suggest a dataset' processing.

---

## 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**

In 2019, the focus is on the development of standard workflows and simple recommendations that support and eventually mandate the process of depositing supplementary primary biodiversity data, both to aid submitting authors and publishing houses. By the end of 2019, a first version of an information page on GBIF.org will be available for journal publishers to reference when recommending GBIF as a data repository, outlining the data publication process for authors, and pointing at simple data spreadsheet templates. The option of offering hosted IPT installations for journal publishers for this purpose is under evaluation.

### **2020 Work Items**

- Work with journals to make it easier for article authors to deposit supplementary data in formats suitable for GBIF publication.
- Establish model guidelines for data deposition and citation and develop compliance criteria for distribution to publishers. Build on recent developments of COPDESS and Research Data Alliance with GBIF-specific guidance. Develop additional communication materials to describe the benefits of DOI-citation best practices and work with journals on implementation.

## 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
- 4.a.iv Coordinate effort to achieve CoreTrustSeal data repository certification for relevant portions of the GBIF infrastructure (added 2019)
- 4.a.v Develop documentation for all stages in the GBIF data ingestion process, to enable data provenance to be fully understood by users of GBIF data (added 2019)

#### **Progress**

The exploration of necessary steps to achieve CoreTrustSeal data repository certification is starting in Q4/2019. This includes the data management services within GBIF.org, but also seeks to identify a set of trusted repositories for publishing datasets within the GBIF network.

Documentation for all stages in the GBIF data ingestion process to start Q3/Q42019, following deployment of the informatics infrastructure upgrade (pipelines project).

#### **2020 Work Items**

- Continue revision and documentation of flagging routines used in GBIF data ingestion pipelines.



---

## 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**

Performed data analysis on implicit geographic data accuracy/patchiness within datasets (gridded data, centroids), documented in the GBIF data blog, estimating the size of impact on further analyses (EBV, species distribution). Tagging of indexed datasets to support further automated data processing.

Exploring of suitable measures to indicate metadata completeness and appropriate content, focused on the contribution from the BID data mobilization.

### **2020 Work Items**

- Review, consolidate and update existing documentation for data publishers. In particular, provide clear guidance on minimum requirements for published data.
- Develop metrics to track the completeness of core data elements and the degree to which supplied content is appropriate.
- Supply clear indicator measures for the completeness and usability of data as part of GBIF.org dataset pages, based on examples such as the GEOLabel data branding model.
- Extend data-quality assessment to include aspects only detectable above the level of individual records.
- Assess the patchiness of indexed data (geographical clustering, misleading accuracy or precision of coordinates), including evaluation of the apparent causes of data patchiness and

include measures of data patchiness in the data index, at both dataset and record level in the data index.

- Ensure that users of data are able to identify datasets or records that do not fulfil their criteria for geo-accuracy, whether they are accessing data through facets in the GBIF.org, via the API or in downloads.

DRAFT

---

## 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**

A pilot handle-based server was deployed demonstrating GBIF infrastructure is capable to participate in a handle network and issue record identifiers.

### **2020 Work Items**

- Continue to explore the use of the GBIF data index to support stable persistent resolvable identifiers for all specimens and occurrence records.
- Explore bidirectional data linking and synchronization with data management systems and publishers to achieve faster and more accurate mutual updates on data improvements and annotations (minimum **€10,000**).

## 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 help desk 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**

The BioDATA project is a training programme for higher-education students from Tajikistan, Belarus, Ukraine, and Armenia that offers an intensive course of study leading to their certification in biodiversity data management. Launched in 2018 and running through 2021, the Finnish-funded programme has already engaged 23 volunteer mentors and is training half of the programme's planned 64 students on GBIF publishing and data skills by the end of 2019. Altay State University, supported by a GBIFS letter, received funding to integrate GBIF and data education in regular biology training for the BSc and MSc programme, with 70 students expected to take part each year.

#### **2020 Work Items**

- Conduct survey on how GBIF informatics is incorporated into relevant graduate and undergraduate curricula to identify gaps and opportunities and key entry points. Use survey findings to prepare a campaign that could roll out relevant resources through GBIF nodes in 2021.

- Engage with academic-based projects that use or could use better GBIF data in their pipelines and protocols. If funding allows, hold a workshop to stimulate use (**€25,000**).

DRAFT

---

## 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

Continued consideration of recommendations from the Data Fitness for Use reports 2016/2017 (agrobiodiversity, invasive and alien species) in user interface and data processing changes. Some of the recommendations were implemented already, particularly around the incorporation of (to date) 122 GRIIS checklists on introduced and invasive species for countries and territories; other recommendations are still valid as possible future options, some pending structural changes (e.g. the taxonomic backbone offering more flexible ranks).

### 2020 Work Items

- Explore the creation of lightweight, customized website landing pages to address thematic interests on GBIF.org. The pages will increase flexibility to include curated, more informative clustered information, ensuring that users have access to both broad search results and prioritized views of data and information.
- Continue to implement recommendations of past expert user groups. In particular, build on the incorporation of GRIIS checklists to support richer, more targeted information on invasive and alien species, e.g. by highlighting documented occurrences with a relevant IAS status by country. Explore best options to identify and alert users of new occurrences of potential invasive species. In addition, review and address the most feasible and valuable recommendations of the agrobiodiversity group, (see 5c below) with support of continued community involvement.

---

## 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**

Development of SQL based downloads now allows that arbitrary data queries can be accommodated without the need to download all data.

For further discussion of proposals for sampling event data interaction, user interface visuals are being developed in 2019. Meanwhile, a review of available data showed that structured information on sampling protocols is scarce, and will require further work in 2020 to enable visualization and analyses across study sources.

A contract was developed under the IUCN-GBIF Memorandum of Cooperation to allocate **€75,000** from the 2019 budget to the IUCN Invasive Species Specialist Group, to fund completion of work on the Global Register of Introduced and Invasive Species (GRIIS) and its connection to GBIF.org. The work will be completed in early 2020.

Data analysis explored options for species density/species richness representation, documented in the data blog to open for further discussion.

### **2020 Work Items**

- Explore [BIOSCAN](#) as a pilot for visualizing sampling-event data. Pending advances in the revised data model, improve the representation of sampling-event data, particularly

abundance measures. Identify sources to support a consistent, controlled list of sampling protocols.

- Continue work with GEO BON to establish needs to deliver relevant data in suitable formats to support ongoing community-agreed implementation of Essential Biodiversity Variables (EBV). Support further community scoping as and if appropriate within the *alliance for biodiversity knowledge* framework.
- If funds are available, the Secretariat will engage visiting scientists from the assessment/modelling community for part-time secondments to improve shared understanding of data-delivery needs for the EBV and associated communities.
- Develop and roll out guidance for use of GBIF-mediated data in species risk assessments. Continue work based on the IUCN memorandum of cooperation and proposed training event at the 2020 IUCN World Conservation Congress in Marseille.
- Host hackathon to align GBIF with post-2020 Biodiversity Framework (**€40,000** if available). The event would develop tools from the CESP CHM project aimed at improve CBD links to GBIF-mediated data. As post-2020 framework/goal structure emerges, develop comprehensive guidance on the role of primary data to support targets and indicators.

DRAFT



---

## 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**

It has now been 20 years since the OECD Megascience Forum recommended the establishment of a Global Biodiversity Information Facility and a systematic review of GBIF had not been undertaken since 2005. A contractor, CODATA, which led the first review of GBIF, was hired after GB25 to review GBIF's operation, services, governance and place within the broader landscape of biodiversity-related organizations. The review is near completion and will be presented to the GB26 in October.

### **2020 Work Items**

- Implement recommendations of the 20-year review and use it as a basis for the development of the next five-year strategic plan for the period 2022-2026.
- Contract out study of use of GBIF-mediated data in the academic literature with a goal to identify areas of impact, links to research funding sources and networks of researchers. This information will be used to fine-tune GBIF value proposition message and align Work Programme to further impact (€20,000).