



GBIF

GBIF Work Programme 2022 (draft)
Annual Update to Implementation Plan 2017-2022

Version 1, 2021-07-23

Table of Contents

Colophon	1
Suggested citation	1
Persistent URI	1
Document control	1
Overview	1
GBIF Annual Work Programme 2022	2
Financial situation	2
Planned expenditure	4
Secretariat staffing	5
Overview of 2022 Work Items	5
2022 Implementation Plan Update	10
Priority 1: Empower Global Network	10
Activity 1a: Focus on people	10
Activity 1b: Strengthen skills	11
Activity 1c: Equip Participant nodes	12
Activity 1d: Equip data publishers	14
Activity 1e: Expand national participation	15
Activity 1f: Plan implementation	16
Activity 1g: Coordinate resources	17
Priority 2: Enhance Biodiversity Information Infrastructure	19
Activity 2a: Modernize data standards	19
Activity 2b: Deliver names infrastructure	21
Activity 2c: Catalogue collections	22
Priority 3: Fill Data Gaps	23
Activity 3a: Identify priority gaps	24
Activity 3b: Expand data streams	25
Activity 3c: Engage data holders	26
Activity 3d: Rescue datasets	27
Activity 3e: Liaise with journals	28
Priority 4: Improve Data Quality	29
Activity 4a: Ensure data persistence	29
Activity 4b: Assess data quality	30
Activity 4c: Enable data curation	30
Priority 5: Deliver Relevant Data	31
Activity 5a: Engage academia	31
Activity 5b: Document needs	32
Activity 5c: Support biodiversity assessment	34
Activity 5d: Assess impact	35

Colophon

Suggested citation

GBIF Work Programme 2022 Annual Update to Implementation Plan 2017–2022: GBIF Secretariat.
<https://docs.gbif.org/2022-work-programme/en/>.

Persistent URI

<https://docs.gbif.org/2022-work-programme/en/>

Document control

Version 1.0, 2021-07-23

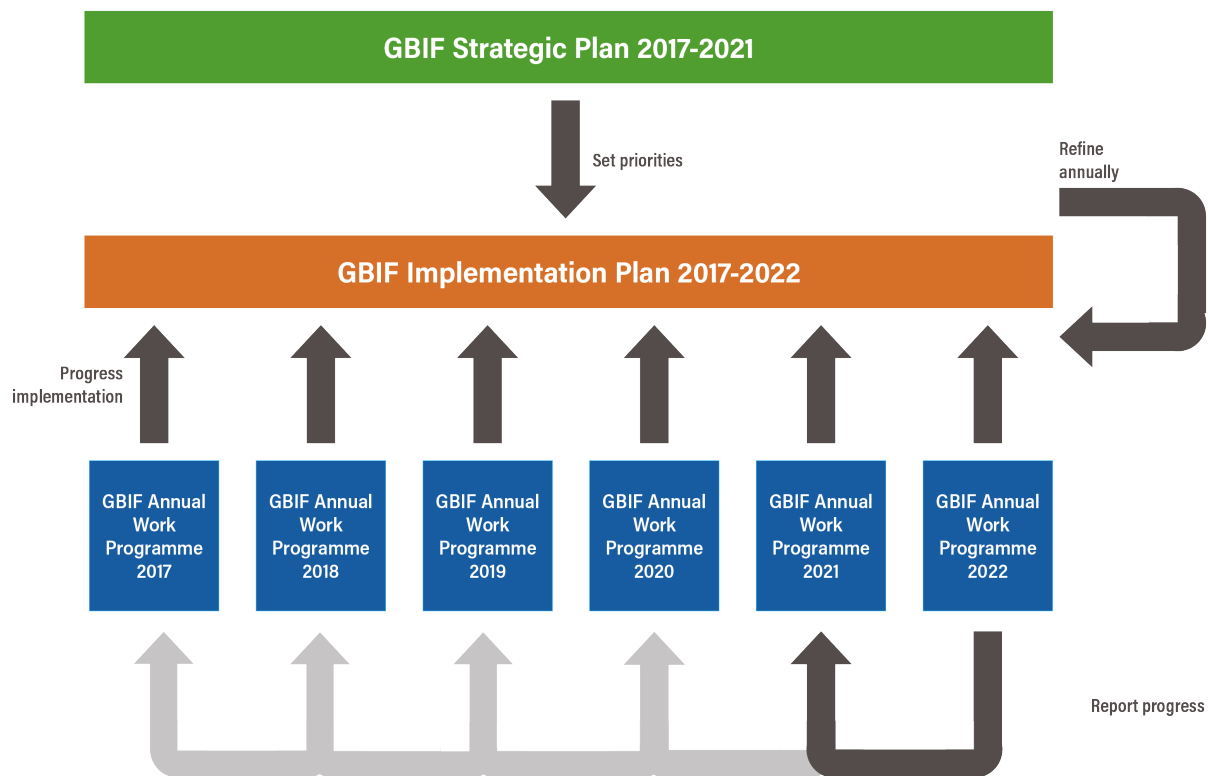
Overview

This document serves both as the 2022 revision of the GBIF Implementation Plan for the period 2017-2022 and also as the GBIF Annual Work Programme 2022. Both elements are based upon the [GBIF Strategic Plan for 2017-2021](#), which was extended for one year due to the challenges created by the global pandemic.

The GBIF Annual Work Programme 2022 represents the final programme based on the current strategic plan. The new strategic planning framework for the period 2023-2027 is due for approval at the 28th GBIF Governing Board meeting in October 2021 (GB28), and this will form the basis of a new implementation plan to be developed during 2022.

The Strategic Plan 2017-2021, agreed at the GB22 meeting in 2015, established five priorities for GBIF's work. The Implementation Plan identified 22 activities which GBIF must undertake to address these priorities during the six-year period. Each year, GBIF has reviewed 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 prioritized set of work items during the year. Annual Work Programmes in the second and subsequent years have also reported on progress against the plan.

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



The Work Programme is presented here with updates reflecting progress during 2021 and planned activities for 2022.

GBIF Annual Work Programme 2022

Information on activities planned for 2022 is included under each of the Activity sections included in the 2022 implementation plan update below. For convenience, this information has also been collated as an [Overview of 2022 Work Items](#) listing the items for which approval is sought from the Governing Board at GB28.

Financial situation

As of July 2021, current and imminent Voting Participant contributions are slightly below expected contributions approved in the GB27 budget. This total, however, takes into consideration expected inability of some countries to pay contributions. Therefore, the total financial contribution provided by countries remains below the goal set forth in the financial model approved at GB22.

The 2021 budget approved by GB27 included some funds for travel and in-person meetings in 2021 (including regional meetings and GB28), but very little of this was used, and these budget items were downscaled in a mid-year revision. Partly due to these savings, €200K budgeted within the 'COVID-19 reserve' was deemed not to be needed, and following mid-year approval by the Executive Committee the Secretariat removed this level of reserve from the budget.

The 2021 Work Programme budget items were adjusted with the overall budget level similar to the budget adopted in GB27. Changes approved by the Budget Committee and Executive Committee for each Strategic Plan Priority are listed below. Information on each new or amended item is included in

the 2021 progress reported for all activities.

Priority 1:

- Regional nodes meetings: decrease by €32K (in-person meetings cancelled, regional breakout sessions held during the virtual Global Nodes Meeting)
- *alliance for biodiversity knowledge*: increase by €10K

Priority 2:

- Consultancy to begin development of improved data models: €30K

Priority 3:

- Human disease vector data mobilization campaign: increase by €25K
- Soil biodiversity data mobilization campaign: €25K

Priority 5:

- Up to €80,000 for a quantitative analysis of the value and impact of the GBIF network

Supplementary income

The following additional expenditure was enabled by supplementary funds received or budgeted during 2021:

- JRS Biodiversity Foundation additional contribution to African Biodiversity Information for Development (BID) projects: €411K
- Biodiversity Community Integrated Knowledge Library (BiCIKL) project (EU Horizon 2020): Developer costs €66K
- Support for Asia regional support contracts through SYNTHESYS+ (European Union): €109K
- Salary subsidies (BID, SYNTHESYS+, DiSSCo Prepare and BiCIKL): increased to €152K

As a result of the past year's staffing and Work Programme changes, GBIF is expected to close 2021 with net capital of around €398,181, with an additional €461,000 in the COVID-19 Reserve.

The 2022 proposed budget includes a smaller allocation (€400K) for contracted activities in the Work Programme expenditure (i.e. non-salary costs allocated to calls, workshops, contracts, etc.) than currently budgeted for 2021. The proposed 2022 budget uses just over half (€261K) of the COVID-19 Reserve. Staffing levels increase by one staff member in the 2022 budget. The planned expenditure for 2022 is €188,155 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 (€471,026, net capital forecast for the end of 2022).

GBIF is a maturing infrastructure and network of people. This maturation has enabled infrastructure improvement, inclusion of new data types and stronger global engagement. It also comes at an increased cost to maintain the infrastructure and engagement at current levels. The current core

budget is based on €3.5 million from Voting Participants, but the amount actually received is typically less than this. Staff growth in past years means that staff salaries and basic operational costs (including insurance) are now similar to or exceed Voting Participants' contributions. This leaves little room for increased feature development and engagement using core funds.

To meet the development needs, the Secretariat has initiated a resource mobilization strategy to meet these needs with supplementary funding. The Secretariat is confident that the value presented by BID and other supplementary programmes, present infrastructure and data quality improvements will be translated into a strong external funding stream.

Planned expenditure

In addition to salary and operating costs (including support for GBIF Governance mechanisms), core funding is reserved in the Work Programme 2022 for the following activities:

Priority Area 1

- Commission risk assessment study for sensitive species data (€15K)
- Launch a new call for proposals under the Capacity Enhancement Support Programme (€80K)
- Support further development of the Living Atlases community (€15K)
- Support for 2022 Regional Nodes Meetings (€32K) - note that some meetings will be supported from supplementary funds or funds reserved from previous budgets
- Support for the *alliance for biodiversity knowledge* (€20K)

Priority Area 2

- Update GBIF data models (€40K)
- Maintain and update processes for constructing the GBIF taxonomic backbone. €108K has been allocated in the budget to support GBIF contractor costs
- Web Developer contract (€30K)

Priority Area 3

- Support for thematic data mobilization (€40K)

Priority Area 5

- Support for phylogenetic integraton with occurrence data (€20K)

Other Funds

- Governance funds: Launch 2022 calls for the Ebbe Nielsen Challenge and Young Researchers Awards (€25K)

Secretariat staffing

Staff increases in recent years have enabled the Secretariat to maintain an increasing level of service. The Secretariat currently has 28 staff (three of whom work part-time) and is currently (July 2021) recruiting for a communications officer. In addition to the Director and Deputy Director, the Secretariat is now structured as four teams, as follows:

- Participation and Engagement (eight staff)
- Data Products (four staff)
- Informatics (eight staff, one contractor)
- Administration (six staff, plus student support)

All staff work towards accomplishing the goals of the five Priority Areas in the 2017-2022 Implementation Plan, with most staff working towards multiple areas. The Secretariat internally reports salary to the Work Programme Priority Areas and Secretariat management based on an estimate of staff time spent on the areas.

Overview of 2022 Work Items

This following is a summary of all 2022 Work Items proposed in the 2022 update of the Implementation Plan below. It is intended to summarize the work items for which approval is sought from the Governing Board at GB28. Accordingly, it does not include the additional commitments made and reported by Participants in the detailed activity sections.

Activity 1a: Focus on people

- Improve representation of individuals on GBIF.org, with more coherent and integrated handling of the roles and profiles of Secretariat staff and network members, including biographies, activities and other related information

Activity 1b: Strengthen skills

- Develop new training module on data quality management, targeted at GBIF nodes
- Monitor impact of GBIF training programmes, based on uptake of existing training resources, skills certification through digital badges, replication of courses throughout the network, and outputs of data and other products linked to participation in training
- Further develop joint guidance with OBIS on publication of marine datasets through both networks

Activity 1c: Equip Participant nodes

- Mature the technical capabilities and community processes to expand the hosted portal service, focussing on the needs of GBIF participant nodes. This activity assumes positive feedback from participants of the pilot program. The expanded programme will determine the level of service that will be available to each node (including number of portals), and possible additional services that would be available through extra financial contributions

- Continue support for the Living Atlases collaboration through extended contract for the technical coordinator (€15K)
- Support in-person regional nodes meetings for 2022, if permitted by the travel situation (€32K from core funds)
- Actively explore supplementary funding for additional regional support contractors in Latin America, the Pacific, and E Europe/Caucasus/Central Asia
- Refresh existing guidance materials for GBIF nodes and associated support, including node-level mentoring, capacity self-assessment and training for national-level project teams
- Develop guidance on best practices for working with indigenous and local knowledge, including application of the CARE principles, following recommendations of the working group established during 2021

Activity 1d: Equip data publishers

- Building on the 2021 work of the sensitive species working group, commission an analysis of the current scale of sensitive species data published through GBIF, evaluation of the evidence that such data presents risks of harm, and recommendations for applying a risk assessment approach (€15K)
- Based on work carried out in 2021, develop plans for the future of the IPT
- Implement new technical documentation website

Activity 1e: Expand national participation

- Review the current status and models of participation in GBIF both at national and organization level, with a view to developing recommendations for the new strategic implementation period at GB29

Activity 1f: Plan implementation

- Organize the 3rd Global Biodiversity Informatics Conference (GBIC3), hosted by the LifeWatch European Research Infrastructure Consortium (ERIC), to address the ongoing operation and governance of the *alliance for biodiversity knowledge*
- Organize and run the 2nd SYNTHESYS+ workshop relating to citation of collections
- Continue to convene activities in the framework of the *alliance for biodiversity knowledge* (€20K)

Activity 1g: Coordinate resources

- Launch a call for proposals under the 2022 Capacity Enhancement Support Programme (€80K)
- Following approval of the resource mobilization strategy at GB28, develop implementation roadmap and begin implementation, with a view to targeting additional supplementary funds as well as consolidating core contributions from Voting Participants
- Subject to an improved travel situation in 2022, organize an in-person engagement meeting in Asia, using supplementary funds from BIFA (€20K) and the SYNTHESYS+ project (€19K), co-located with the 2022 Asia regional nodes meeting

Activity 2a: Modernize data standards

- Continue to participate in the ongoing Open Digital Specimen and Extended Specimen Network standards development and strive towards a common solution
- Together with the Atlas of Living Australia (ALA), support the Living Atlas community, in adoption of the shared data pipelines
- Following open review of the revised data model, prototype new infrastructure that allows the demonstration of data mapping, publishing and indexing into new catalogues, developed in an open manner. This will involve development in the IPT, new data pipelines, APIs and user interfaces to explore content and communities of data publishers willing to shape data for the pilot. It is envisaged one or more developer workshops will be arranged to encourage more open development and collaboration
- Ensure all vocabularies are included in the GBIF Registry and maintained in a collaborative manner

Activity 2b: Deliver names infrastructure

- Complete outstanding tasks to integrate the extended Catalogue of Life in GBIF, partly within the framework of the BiCIKL project
- Maintain and update processes for constructing the GBIF Backbone Taxonomy, including monitoring the content and helping to prioritize editorial effort. €108K has been allocated in the budget to support GBIF costs. This work is in collaboration with Catalogue of Life
- Explore feasibility of supporting national taxonomies for exploring GBIF occurrence data to better enable national-level reporting (carried over from 2020 and likely to be incomplete in 2021)

Activity 2c: Catalogue collections

- Complete outstanding tasks to deploy an enriched GRSciColl providing search and access of collections, specimens and people
- Focus on content of GRSciColl: cleanup of existing entries and registration of new ones by promoting use and giving training and support to editors, and promoting consistent use of codes within data shared
- Seek to identify links between journal articles and collections based on the collection codes, within the framework of the EU-funded BiCIKL project
- Support user interface translations for GRSciColl
- Carried over from 2020: Explore synchronization of content with the Consortium of European Taxonomic Facilities (CETAF) Registry (under development)

Activity 3a: Identify priority gaps

- Publish updated best practices on prioritizing data mobilization based on data gap analysis, for incorporation into guidance for nodes and data publishers
- Explore the use of data gap analysis to prioritize support for data mobilization through funded

programmes coordinated by GBIF, and to document the value added by existing projects

- Work with partners to address the information needs of the post-2020 Global Biodiversity Framework through filling priority data gaps

Activity 3b: Expand data streams

- Building on the framework developed in 2021, establish data mobilization campaigns for additional thematic data streams (possibly freshwater, DNA-derived data and tropical biodiversity) (€40K)
- Seek supplementary funding to support scaling up of support for thematic data mobilization throughout the GBIF nodes network (see also Activity 3c)
- Continue working with partners to increase mobilization and improve visibility of biodiversity monitoring datasets in GBIF

Activity 3c: Engage data holders

- Building on preparatory work and partnerships developed in 2021, scale up mobilization of primary biodiversity data from the private sector, through the Data4Nature initiative and establishment of helpdesk services
- Engage additional sectoral associations, financial and other initiatives to integrate biodiversity data-sharing commitments into best practices for Environmental, Social and Governance (ESG) or regulatory requirements

Activity 3d: Rescue datasets

- Continue to implement workflow for prioritizing and drawing upon potential data sources reported through the 'suggest a dataset' tool, including involvement of nodes, mentors and crowdsourced solutions (carried over from 2021)
- Based on the outcome of the re-evaluation, work with interested parties to roll out a workflow for processing data mobilization requests arising from the 'suggest a dataset' tool

Activity 3e: Liaise with journals

- In consultation with journal publishers, explore solutions to enable academic authors to submit and publish datasets to GBIF alongside research articles
- Further develop the model for sponsored data papers, in connection with mobilization campaigns from thematic data streams
- Continue BiCIKL collaboration to improve integration of data extracted from research literature

Activity 4a: Ensure data persistence

- Support adopters of cloud technologies for accessing GBIF-mediated data, and revise procedures and formats as necessary.
- Explore opportunities to strengthen integration of the GBIF network within research data infrastructures

Activity 4b: Assess data quality

- Implement improvements to data validation reports identified in community consultation
- Promote integration of the data validator in tools connected to GBIF, including the IPT, through use of the validator API

Activity 4c: Enable data curation

- Explore what can be enabled to offer annotation services in GBIF.org following ideas offered during the consultation topic during 2021 (see Activities 1f and 2a).
- Continue to explore the use of the GBIF data index to support stable persistent resolvable identifiers for all specimens and occurrence records (see Activity 2a)

Activity 5a: Engage academia

- Consolidate training resources aimed at academic users and publishers of GBIF-mediated data, for use in funded programmes and guidance for nodes
- Support implementation of the BioDATA Advanced project through partners in Norway, the Russian Federation and South Africa
- Promote GBIF at relevant fora including the 9th International Barcode of Life Conference, scheduled for August 2022, and the International Congress of Entomology, scheduled for July 2022
- Support for phylogenetic integraton with occurrence data (€20K)

Activity 5b: Document needs

- Coordinate work of the expert group on mobilization and use of data on zoonotic diseases based on outputs from the systematic review commissioned in 2021
- Continue programme for mobilization of soil biodiversity data through partnerships and work streams initiated in 2021
- Continue partnership with the IUCN Invasive Species Specialist Group and the Convention on Biological Diversity Secretariat to improve content and curation of data relating to invasive alien species

Activity 5c: Support biodiversity assessment

- Through the *alliance for biodiversity knowledge*, hold a workshop on the outcomes of the analysis of primary data needs for the post-2020 Global Biodiversity Framework, involving partners in GEO BON, the Biodiversity Indicators Partnership, CBD Secretariat and others
- Implement a process enabling key checklists to be used in filtering occurrence data, such as Red Listed species and invasive alien species (carried over from 2020 and likely to be incomplete in 2021)
- Work with the new secretariat of GEO BON to improve connections between Biodiversity Observation Networks (BONs) and nodes at national, regional, thematic and global levels

- Continue collaboration with IUCN to implement the Memorandum of Cooperation, including exploration of publication of reference datasets associated with Red List assessments, and development of training material for use of data in the Key Biodiversity Area (KBA) process

Activity 5d: Assess impact

- Publish quantitative analysis of the value and impact of the GBIF network, and promote in engagement with governments, funders and other partners
- Launch Ebbe Nielsen Challenge and Young Researchers Awards competitions for 2022 (€25K), and generate communications connecting past awardees with subsequent work and impact
- Explore options for a 'Nodes Award Scheme' to recognize progress, performance and innovation across the community
- Produce communication materials including short videos demonstrating the impact of GBIF's funded programmes such as BID and BIFA
- Complete and promote communication materials demonstrating GBIF's impact across thematic communities

2022 Implementation Plan Update

Note that the 'Rationale' section in each activity is presented with the original wording from the Implementation Plan in 2017, to show the original intention of the activity. Participant contributions and Participant plans will be added to the final version of the Work Programme following input from nodes alongside review of the draft Work Programme

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 (for the period 2017-2022)

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.

2021 Progress

Building on previous efforts during the current Implementation Plan, GBIF has continued to make use of the growing cadre of volunteers— **Biodiversity Open Data Ambassadors, mentors and translators**—who contribute their time and energy toward fulfilling the network's global mission. While the opportunities for representing the GBIF community in relevant fora were limited due to the COVID-19 pandemic, the GBIF volunteer community has remained active in numerous virtual events this year, including mentoring and training workshops, e-conferences and symposia.

Apart from maintaining lists of volunteers, progress on this activity since 2017 has mainly focused on establishing the ability to connect individual ORCID IDs with various activities on GBIF.org and developing the **GBIF Community Forum** as an effective space for both general discussions and targeted consultations. More ambitious plans to represent and increase connections between the individual contributions across GBIF.org have not been implemented, due in part to challenges of effectively managing and protecting personal data. The final year of the Implementation Plan presents an opportunity to review and consolidate this work in order to make progress toward the original objectives of this activity.

2022 Work items

- Improve representation of individuals on GBIF.org, with more coherent and integrated handling of the roles and profiles of Secretariat staff and network members, including biographies, activities and other related information

2021 Participant contributions

2022 Participant plans

Activity 1b: Strengthen skills

Rationale (for the period 2017-2022)

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.

2021 Progress

The current implementation period has seen very substantial progress in skills development across the GBIF network, including the consolidation of training resources in multiple languages, mainly developed through BID but re-used across a range of projects and programmes. Increasingly these resources are available both for rollout in Secretariat-led workshops (virtual as well as in-person) and for self-instruction.

A new [training and e-learning page on GBIF.org](#) now provides open access to a core curriculum including courses on a general introduction to GBIF, node management, data mobilization, and data use for decision making. These courses use a flexible format enabling easy updates, self-guided instruction and embedding of videos. They complement the [e-learning platform hosted by GBIF Spain](#) which is still used to organize training workshops for participants in programmes such as BID and BIFA.

Following the release of manuals updating guidance on [georeferencing](#), [data cleaning](#) and [generalizing data on sensitive species](#), GBIF's core documentation programme proceeded with identification of further guidance to be commissioned in 2021. These include updated guidance on digitization and publication of data from camera traps, and new guidelines on species migration and telemetry data, targeted filling of data gaps and data quality.

During 2021, collaboration between nodes and secretariats from GBIF and the [Ocean Biodiversity Information System \(OBIS\)](#) resulted in [a series of videos prepared for the virtual Global Nodes Meeting](#), outlining the overlaps and differences involved in publishing marine data for the two networks. This collaboration is the first step towards aligning guidance and training materials between OBIS and GBIF, with further opportunities provided by the data mobilization workshops in the second phase of BID, especially in the Caribbean where several funded projects have a marine focus.

2022 Work items

- Develop new training module on data quality management, targeted at GBIF nodes
- Monitor impact of GBIF training programmes, based on uptake of existing training resources, skills certification through digital badges, replication of courses throughout the network, and outputs of data and other products linked to participation in training
- Further develop joint guidance with OBIS on publication of marine datasets through both networks

2021 Participant contributions

2022 Participant plans

Activity 1c: Equip Participant nodes

Rationale (for the period 2017-2022)

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.

2021 Progress

Due to the continuing COVID-19 travel restrictions, plans for in-person regional nodes meetings during 2021 were impossible to implement. Instead, and through consultation with the Nodes Steering Group, the Secretariat decided to organize a **virtual Global Nodes Meeting in June/July 2021**, returning to the usual cycle of holding such meetings every second year. As part of the series of sessions held over five days, regional breakout sessions were organized for all six of the GBIF regions. To maximum engagement during the meeting, a series of videos produced by Secretariat staff, nodes and other partners were made available ahead of time, with the live sessions focussed on panel discussion and interventions from participants. The virtual format attracted record attendance for a Global Nodes Meeting, with 168 participants attending at least one live session, including node managers, node staff and observers from all regions.

Following a **call for expressions of interest from nodes to take part in the pilot phase of the GBIF hosted portal programme**, the Secretariat accepted 19 hosted portal projects, covering institutional, national, regional and thematic examples. The first version of the technical framework was complete as of July 2021, and participants were in the process of preparing their content and styling. Terms were drawn up through legal consultation enabling the Secretariat to provide this hosted service through agreements with participating nodes. The first portals were expected to be made public on the internet in August 2021.

In connection with the hosted portal pilot, the Secretariat began development of guidelines to provide more consistent visual branding of products and platforms associated with GBIF nodes, regions and publishers, as well as externally-managed platforms with a dependency on GBIF services.

The Secretariat established a working group involving several nodes to share experiences and make recommendations on working with indigenous and local knowledge (ILK), especially with regard to application of the **CARE Principles for Indigenous Data Governance**.

2022 Work items

- Mature the technical capabilities and community processes to expand the hosted portal service, focussing on the needs of GBIF participant nodes. This activity assumes positive feedback from participants of the pilot program. The expanded programme will determine the level of service that will be available to each node (including number of portals), and possible additional services that would be available through extra financial contributions.
- Continue support for the Living Atlases collaboration through extended contract for the technical coordinator (€15K)

- Support in-person regional nodes meetings for 2022, if permitted by the travel situation. The Asia regional nodes meeting will co-located with the Asia regional engagement meeting and supported by supplementary funds (see Activity 1g); €28,000 is carried over from 2020 for hosting of the Africa regional meeting in Togo; €32,000 will be available from the 2022 Work Programme budget to support additional meetings.
- Actively explore supplementary funding for additional regional support contractors in Latin America, the Pacific, and E Europe/Caucasus/Central Asia
- Refresh existing guidance materials for GBIF nodes and associated support, including node-level mentoring, capacity self-assessment and training for national-level project teams
- Develop guidance on best practices for working with indigenous and local knowledge, including application of the CARE principles, following recommendations of the working group established during 2021

2021 Participant contributions

2022 Participant plans

Activity 1d: Equip data publishers

Rationale (for the period 2017-2022)

Data publishers are an essential component of the GBIF network as they share their content through the common infrastructure. More than 1,700 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.

2021 Progress

Following publication of the [updated guidance on generalizing data on sensitive species](#), the Secretariat set up a working group of nodes to document the various approaches across the community in application of national sensitive species lists. This group will make recommendations on a more consistent approach, and options for helping data publishers to identify records that may present risks if published with precise location details.

At the 2020 TDWG Conference, GBIF and the BioCASE team held two virtual workshops to capture ideas and hold discussion on the future of the IPT. The [final report](#) captures the outcomes which provided clear direction. Following this feedback GBIF has appointed a new lead developer on the IPT product and released the 2.5.0 version of IPT which addresses many of the key issues.

The framework for a technical documentation website has been created, with the [IPT manual](#) being the first component included. Collating information on architecture, data processing, definitions of formats, tutorials, guides and the API is ongoing but requires more dedicated focus during the

remainder of 2021 and 2022.

2022 Work items

- Building on the 2021 work of the sensitive species working group, commission an analysis of the current scale of sensitive species data published through GBIF, evaluation of the evidence that such data presents risks of harm, and recommendations for applying a risk assessment approach (€15K)
- Based on work carried out in 2021, develop plans for the future of the IPT
- Implement new technical documentation website

2021 Participant contributions

2022 Participant plans

Activity 1e: Expand national participation

Rationale (for the period 2017-2022)

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.

2021 Progress

Looking across the whole implementation period to date, 13 new countries have joined GBIF as national Participants. They include Mali (2017); Cameroon, Liberia, South Sudan, Viet Nam and Zimbabwe (2018); Angola, Belarus and Tonga (2019); Armenia, Cambodia and Uzbekistan (2020); and Namibia (2021). Over the same period, seven regional or international organizations have joined as associate participants: the Secretariat of the Pacific Regional Environment Programme (SPREP), iDigBio (Integrated Digitized Biocollections), Lifewatch ERIC, Specify Collections Consortium, Distributed System of Scientific Collections (DiSSCo), the East Asia Biodiversity Conservation Network, and Observation International. Several of the new members joined GBIF as a direct result of engagement through funded programmes led or supported by GBIF, in particular the BID, BIFA and BioDATA programmes.

Opportunities for direct engagement at international conferences and intergovernmental fora continued to be severely limited in 2021 due to the COVID-19 pandemic. However, a series of regional engagement webinars was planned for later in the year, building on the successful outreach achieved by such webinars in 2020. In addition, targeted communication products were developed during 2021 aimed at strengthening GBIF's value proposition for existing and potential national members, as well as supporting the new GBIF resource mobilization strategy (see Activity 1g).

While expansion in formal participation over the period has been modest, engagement through a range of activities has significantly increased GBIF's global footprint both in terms of the sharing and use of data through the network. By July 2021, institutions from 124 countries were publishing data through GBIF, compared with 61 countries that participate through signature of the GBIF Memorandum of Understanding (MoU). Two countries outside the formal network (India and Italy) were within the top ten locations of visitors to GBIF.org, while China ranked second in the world both for data download requests and peer-reviewed studies citing use of data accessed through GBIF. Perhaps the most dramatic increase in activity outside the formal national membership has been in the case of the Russian Federation, where **data publishing** stood at virtually zero in 2017, and by 2021 had reached more than 3.8 million species occurrence records, being shared through GBIF in nearly 500 datasets published by 117 Russian institutions.

Pro-active engagement in the delayed process for agreeing the post-2020 Global Biodiversity Framework through the Convention on Biological Diversity (CBD) has maintained GBIF's profile in supporting national implementation of global goals and targets to address the biodiversity crisis. Acknowledgement of GBIF in decisions and background documents still under negotiation ahead of the COP15 meeting in China will provide important additional support for expansion of GBIF's membership in the coming decade, especially in relation to GBIF's role to support indicators, taxonomy, knowledge management and capacity development (see also Activity 5c).

2022 Work items

- Review the current status and models of participation in GBIF both at national and organization level, with a view to developing recommendations for the new strategic implementation period at GB29

2021 Participant contributions

2022 Participant plans

Activity 1f: Plan implementation

Rationale (for the period 2017-2022)

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.

2021 Progress

As part of the system of reporting on the annual updates of the global GBIF Implementation Plan, the Secretariat has raised the visibility of activities across nodes by including Participant contributions and plans alongside the summaries for each activity within the strategic priority areas. To reinforce this approach, the Secretariat is preparing to capture information from nodes about budgets and in-kind resources dedicated to GBIF activity, so that the total investment in GBIF can be better reflected in financial reporting.

Following the 2nd Global Biodiversity Informatics Conference (GBIC2) in 2018, GBIF has taken a leadership role in convening a number of consultations across the broader biodiversity informatics community through the *alliance for biodiversity knowledge*.

In 2021, as part of the EU-funded **SYNTHEsys+ project** and within the framework of the *alliance*, GBIF convened a workshop focused on Converging Digital Specimens and Extended Specimens - Towards a global specification for data integration. This workshop had been anticipated to be the third in a series of but was moved forward. The consultation covered 11 subtopics and was held in two phases, **Phase 1 in February** and **Phase 2 in June**.

Due to the COVID-19 restrictions, the 3rd Global Biodiversity Informatics Conference (GBIC3) was postponed until 2022. The event will be aligned with the EU-funded **BiCIKL (Biodiversity Integrated Knowledge Library) grant**.

2022 Work items

- Organize the 3rd Global Biodiversity Informatics Conference (GBIC3), hosted by the **LifeWatch European Research Infrastructure Consortium (ERIC)**, to address the ongoing operation and governance of the *alliance for biodiversity knowledge*
- Organize and run the 2nd SYNTHEsys+ workshop relating to citation of collections
- Continue to convene activities in the framework of the alliance for biodiversity knowledge (€10K)

2021 Participant contributions

2022 Participant plans

Activity 1g: Coordinate resources

Rationale (for the period 2017-2022)

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.

2021 Progress

During 2021, the Secretariat selected a total of 42 projects for funding under the second phase of the Biodiversity Information for Development (BID) programme following the calls for the proposals launched during 2020. These included **18 projects in sub-Saharan Africa**, 18 in the Caribbean and 7 in the Pacific. In addition to the top-up funding of €1.6M from the **European Commission's Directorate-General for International Partnerships (EU-INTPA)** to support the new phase of BID,

GBIF secured a funding agreement with the [JRS Biodiversity Foundation](#) to contribute €411K towards BID projects in Africa, enabling approval of five additional projects on biodiversity data mobilization and use.

The continuing COVID-19 travel restrictions inevitably disrupted the original plans for implementing BID activities during 2021. Apart from requiring funded projects to avoid international travel in their plans for this year, the main impact was the need to replace planned in-person training workshops organized by the Secretariat with virtual alternatives for the three target regions. The Secretariat Participation and Engagement team organized virtual data mobilization training for BID grantees in Africa (May), the Caribbean (July/August) and Pacific (August), using the curriculum developed during the first phase of BID and combining a preparatory phase with 'live' sessions over a week and subsequent submission of exercises to demonstrate acquisition of the necessary skills. Additional virtual sessions were conducted for the project teams implementing the data use grants, introduced as an innovation in the second phase of BID.

The BID Steering Committee approved a proposal from the Secretariat to redirect funds originally budgeted for in-person workshops towards regional project support contracts, aimed at providing ongoing technical support for the funded projects within the target regions, and helping to grow the community of practice developed through the BID programme. Following open calls in [Africa](#) and the [Caribbean](#), three individuals in Africa and one in the Caribbean were selected to provide these support services under 2-year contracts starting in August 2021. As well as supporting the projects funded through BID, these contractors will work with regional node representatives to help coordinate and direct technical support for GBIF nodes in the respective regions.

Following a [call for proposals in December 2020](#), the Secretariat selected 13 data mobilization projects in Asia, under the sixth funding round of the Biodiversity Information Fund for Asia (BIFA), using supplementary funds contributed by the Ministry of Environment, Government of Japan. As this was expected to be the last BIFA call in its present form, the BIFA Steering Committee agreed to 'double up' two annual contributions into a single call with €220,000 available for grants up to €20,000 each, with projects implemented over 18 months. Particular priority in the call was given to projects combining DNA-based techniques such as barcoding with mobilization of data through GBIF. A virtual data mobilization training workshop for the BIFA projects was scheduled for November 2021, with funds reserved for an in-person engagement meeting in Asia in 2022 if the travel situation permits.

Based on the model developed for the BID programme, the Secretariat prepared a call for proposals for regional project support contracts for Asia, using funds from the EU-funded SYNTHESYS+ programme under an agreement reached with the programme partners. These roles will be recruited by November 2021 from Asian applicants and will support project teams funded through BIFA as well as nodes and data publishers in the region.

The 2021 call for proposals under the GBIF Capacity Enhancement Support Programme resulted in selection of five projects involving a total of 14 nodes, led from Argentina, Armenia, Brazil, Cameroon and South Sudan. The themes of the selected projects included mentoring to support node development, extending knowledge of data quality, and the role of zoological collections in filling taxonomic and geographic data gaps.

A draft GBIF resource mobilization strategy, aimed at broadening the funding base for GBIF's global activities, was developed by the Secretariat's Strategic Partnership Officer and endorsed by the

Executive Committee, and will be presented for approval by the GB28 meeting alongside the draft Strategic Plan for the period 2023-2027.

2022 Work items

- Launch a call for proposals under the 2022 Capacity Enhancement Support Programme (€80k)
- Following approval of the resource mobilization strategy at GB28, develop implementation roadmap and begin implementation, with a view to targeting additional supplementary funds as well as consolidating core contributions from Voting Participants
- Subject to an improved travel situation in 2022, organize an in-person engagement meeting in Asia, using supplementary funds from BIFA (€20,000) and the SYNTHESYS+ project (€19,000)

2021 Participant contributions

2022 Participant plans

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 (for the period 2017-2022)

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.

2021 Progress

The Secretariat initiated work during 2021 that will result in a new information model for GBIF. This model will allow GBIF to better accommodate wider data types in its data infrastructure, potentially including tracking data (organism-oriented data), monitoring data (site-oriented data), automated

detection (e.g. camera traps), preserved physical material (e.g. tissue or environment specimens), DNA-derived data (e.g. metabarcoding projects) and literature-based data (e.g. material citations). Following approval by the Science Committee, Budget Committee and Executive Committee, the Secretariat is commissioning a contracted expert in the second half of 2021 to compile known use cases, and define a data model and initial formats for data exchange beyond the Darwin Core Archive (DwC-A).

During 2021, the Secretariat and Atlas of Living Australia (ALA) completed the merge of data processing software, and deployed the resulting pipelines in both infrastructures. GBIF and ALA now actively collaborate on defining the roadmap and enhancements to this shared codebase. The teams will look to integrate this into the deployments by the Living Atlas community, to further align infrastructure and increase our collective capacity and collaboration. It is anticipated this collaboration will expand to develop the prototype catalogue services.

The development work to support the pilot hosted portals (see Activity 1c) has resulted in a new GraphQL-based API, and a collection of reusable ReactJS-based 'widgets'.

The open curation of vocabularies has been included in the **GBIF registry** with the LifeStage vocabulary being the first pilot used by GBIF and the ALA. Work is underway with the **TDWG Vocabulary Task Group** to migrate **all existing** vocabularies into the registry. Once complete, communities can get involved in the maintenance of the vocabularies which dictate how data is treated in GBIF.org.

Collaborations with academic projects were initiated to display occurrence data within a phylogenetic framework. These projects are expected to be continued and expanded in 2022 (see Activity 5a).

2022 Work items

- Continue to participate in the ongoing Open Digital Specimen and Extended Specimen Network standards development and strive towards a common solution
- Together with the Atlas of Living Australia (ALA), support the Living Atlas community, in adoption of the shared data pipelines
- Following open review of the revised data model, prototype new infrastructure that allows the demonstration of data mapping, publishing and indexing into new catalogues, developed in an open manner. This will involve development in the IPT, new data pipelines, APIs and user interfaces to explore content and communities of data publishers willing to shape data for the pilot. It is envisaged one or more developer workshops will be arranged to encourage more open development and collaboration (€40K)
- Ensure all vocabularies are included in the GBIF Registry and maintained in a collaborative manner

2021 Participant contributions

2022 Participant plans

Activity 2b: Deliver names infrastructure

Rationale (for the period 2017-2022)

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.

2021 Progress

The new Catalogue of Life infrastructure, powered by and hosted at GBIF, was launched in 2021. This includes a new [public website](#), [COL ChecklistBank](#), and a new [API](#).

The outstanding development tasks were completed for the Catalogue of Life to resume monthly releases. This included the introduction of stable identifiers for the taxon names exposed in the Catalogue of Life.

Citation practice based on Digital Object Identifiers (DOIs) has been introduced to the datasets that constitute the Catalogue of Life. The COL Checklist 2021 will be issued with associated DOIs for the checklist and underlying species data resources.

Development work is still required around the management of user communities within the Catalogue of Life. Once complete, COL and GBIF data management groups can develop the necessary processes around tasks.

The March 2021 update of the [GBIF Backbone Taxonomy](#) included the work of the [Genome Taxonomy Database r95](#) to improve the Bacteria and Archaea sections.

The focus of the later part of 2021 is on building and integrating an extended Catalogue of Life in GBIF as a replacement for the GBIF Backbone Taxonomy, partly within the framework of the the EU-funded BiCIKL project. This will include assessing and reporting on gaps for organizing GBIF occurrence records, and broadening the community of contributors.

2022 Work items

- Complete outstanding tasks to integrate the extended Catalogue of Life in GBIF, partly within the

framework of the BiCIKL project

- Maintain and update processes for constructing the GBIF Backbone Taxonomy, 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 Catalogue of Life.
- Explore feasibility of supporting national taxonomies for exploring GBIF occurrence data to better enable national-level reporting (carried over from 2020 and likely to be incomplete in 2021)

2021 Participant contributions

2022 Participant plans

Activity 2c: Catalogue collections

Rationale (for the period 2017-2022)

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 digitization and mobilization.

2021 Progress

The Secretariat created a [roadmap](#) in 2021 for the [GBIF Registry of Scientific Collections \(GRSciColl\)](#), focussing on the development necessary to allow wider external contribution, and to mature the processes around editing. Editors may now be given scoped responsibility at institutional or national level. Induction webinars have been held with several nodes, resulting in a global team of 45 editors and 12 mediators actively curating content in the registry as of July 2021.

The [iDigBio](#) collection catalogue is now powered by GRSciColl, through its open APIs. iDigBio data managers edit directly through the online [editing interface](#).

Documentation for editors is in progress and can be expected in Q3 2021. All GRSciColl fields are associated with an English-language description available in the online forms. Capabilities for anyone to suggest a change were deployed in May. Proposed data changes are reviewed by the pool of editors and mediators before being applied.

Capability for user interface translations is set up to support multilingual content. Editors and external communities in support of the catalogue are invited to propose translations to support their work. This has already been taken up by the Society for the Preservation of Natural History

Collections (SPNHC) Biodiversity Crisis Response Committee.

A service has been deployed allowing the linking of collections in GRSciColl to specimen records in GBIF. This has resulted in 134 million records being linked to GRSciColl entries. A basic data dashboard is now available for institutions and collections such as [this example](#).

Options for a richer user interface for GRSciColl are being considered within the hosted portal framework (see Activity 1c). There is work remaining on the data model, and how to make best use of the data clustering to link related data before this can progress. It is anticipated the collections catalogue user interface will continue into 2022.

Adoption of persistent identifiers, such as Research Organisation Registry (ROR) identifiers is actively being discussed within European nodes (e.g. through DiSSCo) with some nodes piloting use of ROR IDs on their entries.

Integration with the CETAF registry remains an objective, but has not yet started due to the effort required to enable external editors and focus on content issues. It is expected to extend into 2022. Piloting a profile of the TDWG Collection Descriptions to capture collection-level metadata has also been delayed and will be considered for 2023.

2022 Work items

- Complete outstanding tasks to deploy an enriched GRSciColl providing search and access of collections, specimens and people
- Focus on content of GRSciColl: cleanup of existing entries and registration of new ones by promoting use and giving training and support to editors, and promoting consistent use of codes within data shared
- Seek to identify links between journal articles and collections based on the collection codes, within the framework of the EU-funded BiCIKL project
- Support user interface translations for GRSciColl
- Carried over from 2020: Explore synchronization of content with the [Consortium of European Taxonomic Facilities \(CETAF\)](#) Registry (under development)

2021 Participant contributions

2022 Participant plans

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 (for the period 2017-2022)

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. Addressing data gaps may require focus on gaining additional occurrence records, targeting data areas missing from published records, or getting additional metadata elements. By consolidating and prioritizing 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.

2021 Progress

Analytics work in 2021 focused on the exploration of taxonomic gaps, checking for areas in GBIF-mediated occurrences that are insufficiently mapped to the Catalogue of Life (CoL). No major areas were discovered, most gaps being detected in the area of Operational Taxonomic Units (OTUs) rather than Linnean taxonomy. Minor results in that area were fed back to CoL's issue tracking system.

Updated guidance on prioritizing data mobilization based on analysis of data gaps was being commissioned in 2021 as part of the GBIF documentation programme (see Activity 1b). This guidance will take account of previous work carried out by the Secretariat during the current implementation period to analyse gaps and biases in data available through GBIF.org, such as those reported to the 2020 regional meetings and [documented in the GBIF data blog](#).

The Secretariat continued to work with the [Knowledge and Data Task Force of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services \(IPBES\)](#) to catalyse investment in priority research and data mobilization to address knowledge gaps identified in the IPBES assessment process. This will include participation in a series of regional dialogues with research funders and programmers, to identify potential funding sources to address data gaps.

In the context of discussions on the post-2020 Global Biodiversity Framework carried out through the Convention on Biological Diversity (CBD), the Secretariat has begun discussions with partners including the Group on Earth Observations Biodiversity Observation Network (GEO BON) and Map of Life (MoL) on how indicators of data completeness may be used to prioritize data mobilization, supporting information targets under the new framework (see Activity 5c).

2022 work items

- Publish updated best practices on prioritizing data mobilization based on data gap analysis, for incorporation into guidance for nodes and data publishers
- Explore the use of data gap analysis to prioritize support for data mobilization through funded programmes coordinated by GBIF, and to document the value added by existing projects
- Work with partners to address the information needs of the post-2020 Global Biodiversity Framework through filling priority data gaps

2021 Participant contributions

2022 Participant plans

Activity 3b: Expand data streams

Rationale (for the period 2017-2022)

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.

2021 Progress

The guide to [Publishing DNA-derived data](#) has been released in its final version with the accompanying [Darwin Core extension](#). During the remainder of 2021 GBIF will work with early adopters in preparing data following the guidelines. Mobilization of DNA-derived data was a priority in selection of the projects funded through the latest call under the Biodiversity Information For Asia (BIFA) - see Activity 1g. New training resources combining skills needed for DNA barcoding and GBIF data mobilization were developed and tested during 2021, for rollout in workshops planned under the BioDATA programme later in 2021 and 2022.

Efforts in 2021 to expand the mobilization of ecological datasets through GBIF included discussions with the German Centre for Integrative Biodiversity Research (iDiv) on the establishment of a repository for sampling-event datasets from individual researchers that would be published through GBIF. These discussions are connected with GBIF's involvement in the [Humboldt Core task group in TDWG](#) exploring improved standards for exchanging data on species inventories. The Secretariat also engaged with the [EuropaBON](#) and [Tropical Andes Observatory](#) projects to improve connections between GBIF and biodiversity monitoring activities in Europe and South America, respectively. The Secretariat was also invited to join the [IUCN Species Survival Commission Species Monitoring Specialist Group](#).

During 2021, the Secretariat developed a framework for running data mobilization campaigns based on thematic data streams, including engagement of expert groups, identification of potential mobilization targets, and helpdesk services to support data formatting and publication in collaboration with nodes. Specific mechanisms are being trialled with regard to data on disease

vectors and hosts, as well as soil biodiversity data (see Activity 5b), and data mobilized through the private sector (see Activity 3c). ===== 2022 Work items

- Building on the framework developed in 2021, establish data mobilization campaigns for additional thematic data streams (possibly freshwater, DNA-derived data and tropical biodiversity) - €40k.
- Seek supplementary funding to support scaling up of support for thematic data mobilization throughout the GBIF nodes network (see also Activity 3c)
- Continue working with partners to increase mobilization and improve visibility of biodiversity monitoring datasets in GBIF

2021 Participant contributions

2022 Participant plans

Activity 3c: Engage data holders

Rationale (for the period 2017-2022)

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.

2021 Progress

The Secretariat has continued to engage with nodes and external partners to support sharing of data from the **private sector**, in particular through monitoring activities connected with environmental and social impact assessments. An updated **guide on publishing primary biodiversity data through impact assessments** was published in partnership with the International Association for Impact Assessment (IAIA).

The **Data4Nature initiative**, designed in partnership with **Agence Française de Développement (AFD)** to engage development finance actors in sharing of biodiversity data, achieved visibility during the virtual meeting of the CBD Subsidiary Body on Implementation (SBI3) with support from the European Union's delegation to the CBD, and further buy-in for this initiative was planned for the IUCN World Conservation Congress taking place in Marseille in September 2021. Interest in supporting this initiative and other activities focussed on mobilizing private sector data through GBIF has come from the UN Environment World Conservation Monitoring Centre (UNEP-WCMC) through the Proteus Partnership, bringing together corporations in the oil, gas and mining sectors.

2022 Work items

- Building on preparatory work and partnerships developed in 2021, scale up mobilization of primary biodiversity data from the private sector, through the Data4Nature initiative and establishment of helpdesk services
- Engage additional sectoral associations, financial and other initiatives to integrate biodiversity

data-sharing commitments into best practices for Environmental, Social and Governance (ESG) or regulatory requirements

2021 Participant contributions

2022 Participant plans

Activity 3d: Rescue datasets

Rationale (for the period 2017-2022)

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.

2021 Progress

Converting the growing collection of suggested datasets, submitted through the **user interface**, is work in progress, with most of the 228 entries (as of July 2021) categorized for geographic coverage, country of origin, and broader subject area. All entries are available in the adjoined **GitHub repository**, and relate to data from and about 40 countries, including datasets with a wider geographic scope.

Support for the uptake and data mobilization by the wider GBIF community requires re-evaluation of this work program item in the second half of 2021. With wide variations in the level of detail supplied with submissions, prioritization remains a major challenge, and requires more in-depth evaluation of individual suggestions. A more targeted approach, working with nodes and thematic communities interested in the uptake of suggestions (see Activity 3b), is required to make more efficient use of the resources available.

2022 Work items

- Continue to implement workflow for prioritizing and drawing upon potential data sources reported through the 'suggest a dataset' tool, including involvement of nodes, mentors and crowdsourced solutions (carried over from 2021)
- Based on the outcome of the re-evaluation, work with interested parties to roll out a workflow for processing data mobilization requests arising from the 'suggest a dataset' tool

2021 Participant contributions

2022 Participant plans

Activity 3e: Liaise with journals

Rationale (for the period 2017-2022)

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.

2021 Progress

Following the successful funding of the **Biodiversity Community Integrated Knowledge Library (BiCIKL)** grant through the European Union Horizon 2020 programme, the Secretariat began collaboration with partners **Pensoft Publishers** and **Plazi** to improve integration of data extracted from research literature (see also Activity 2a).

As use of GBIF-mediated data tracked by the Secretariat in research articles continued to increase, the quality of citations passed a significant milestone in 2021, with 50% of articles now correctly citing the Digital Object Identifiers (DOIs) provided with data downloads. The use of DOIs in citations is expected to improve further with the introduction of **a new tool to generate DOIs from derived datasets** cited by researchers following further filtering or cleaning of data downloaded from GBIF.org, or accessed via public cloud services (see Activity 4a).

Following the successful call in 2020 for **sponsored data papers mobilizing data from European Russia**, using supplementary funds provided by the Finnish Biodiversity Information Facility (FinBIF), the model was repeated with a **second call** for Russian-based papers in 2021, and planned as a means of mobilizing datasets relating to vectors, hosts and reservoirs of human diseases (See Activity 5b).

2022 Work items

- In consultation with journal publishers, explore solutions to enable academic authors to submit and publish datasets to GBIF alongside research articles
- Further develop the model for sponsored data papers, in connection with mobilization campaigns from thematic data streams
- Continue BiCIKL collaboration to improve integration of data extracted from research literature

2021 Participant contributions

2022 Participant plans

Priority 4: Improve Data Quality

Activity 4a: Ensure data persistence

Rationale (for the period 2017-2022)

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.

2021 Progress

Work earlier in the current implementation period addressed the issue of **orphaned datasets**, with a major effort to ensure that datasets no longer connected to their original host were transferred to a 'live' version that could be successfully crawled by the Secretariat, with many nodes 'adopting' these datasets for the purposes of hosting and curation.

Following **open discussion on the Community Forum**, GBIF now provides monthly exports of GBIF occurrence records and makes them available on Microsoft Azure in the **Planetary Computer** and on Amazon AWS as **Public Datasets**. Tutorials and guides have been developed, and support will be given to early adopters. The Secretariat anticipates that GBIF-mediated data will be available in a similar manner on the Google Cloud, and made available in Google BigQuery during Q3 2020.

When implementing the new service for exports to cloud environments, the Secretariat was very conscious of the risk of losing the connection between users of GBIF-mediated data and the data publishers whose records are included in applications using these services. To address this, a new citation tool was developed in support of the cloud data services, known as the **Derived Dataset** service. This ensures users of cloud environments, or after further filtering of any GBIF download, can still provide a DOI-based citation for the content used (see Activity 3e).

GBIF has been accepted as a **Data Provider within the European Open Science Cloud**. The Secretariat will now explore opportunities to strengthen integration of the GBIF network within the wider European research data infrastructures.

2022 Work items

- Support adopters of cloud technologies for accessing GBIF-mediated data, and revise procedures and formats as necessary.
- Explore opportunities to strengthen integration of the GBIF network within research data infrastructures

2021 Participant contributions

2022 Participant plans

Activity 4b: Assess data quality

Rationale (for the period 2017-2022)

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 standardization 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.

2021 Progress

Documentation to support better and more targeted interpretation of issue labels has been continued from 2020. A popular post in the [GBIF data blog](#) from October 2020 on "[Issues and Flags](#)", supporting the interpretation and proposed action on diagnosed data issues, is at the centre of communications and further planning. The post has been maintained through additions in 2021, and will be part of the technical documentation under development (see Activity 1d). Other components are development of training resources and materials targeted to alert data users to [post-processing](#) requirements when using GBIF-mediated data.

GBIF participates in the EU-funded [DiSSCo Prepare](#) project, supporting the group with data exports to assess existing content. This has included the creation of a [DiSSCo network entity](#) to easily download the related content, and support for using Google BigQuery to assess the current [Minimum Information about a Digital Specimen\(MIDS\)](#) level of content.

Ongoing work in 2021 on aligning the GBIF [data validator tool](#) with the GBIF.org data processing pipeline, will be supplemented with a community consultation on data validator functionality from Q3 2021.

2022 Work items

- Implement improvements to data validation reports identified in community consultation
- Promote integration of the data validator in tools connected to GBIF, including the IPT, through use of the validator API

2021 Participant contributions

2022 Participant plans

Activity 4c: Enable data curation

Rationale (for the period 2017-2022)

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.

2021 Progress

Feedback received during the 2021 consultation on digital extended specimens through the *alliance for biodiversity knowledge* (see Activity 1f) supported the need for GBIF to accommodate a more expressive data model for specimens, and to consider how communities can assist in annotating, linking and cleaning data (see also Activity 2a). This will help to inform GBIF's future work on enabling data curation.

Further work on the data validator tool in 2021 (see Activity 4b) will flag issues with stable persistent resolvable identifiers, supporting data curation.

The **data clustering algorithm** developed in 2020 has gone through iterative improvements, and by July 2021 was detecting candidate links for 23 million records, including 12 million specimens. A hackathon is being scheduled for Q3 2021 within the European BiCIKL project that will invite additional developers to contribute improvements to this algorithm. The Secretariat has presented the outputs at various conferences, and there is interest in brokering these annotations back to source systems.

2022 Work items

- Explore what can be enabled to offer annotation services in GBIF.org following ideas offered during the **consultation topic** during 2021 (see Activities 1f and 2a).
- Continue to explore the use of the GBIF data index to support stable persistent resolvable identifiers for all specimens and occurrence records (see Activity 2a)

2021 Participant contributions

2020 Participant plans

Priority 5: Deliver Relevant Data

Activity 5a: Engage academia

Rationale (for the period 2017-2022)

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

2021 Progress

The Secretariat developed new training resources in 2021 to support academic uses of GBIF-mediated data, trialled in a virtual workshop for a graduate programme in the United Kingdom coordinated by the [Natural Environment Research Council](#). These resources focus on guidance for finding and downloading data through GBIF.org, data cleaning and best practices for citation. Once consolidated into a standard training package, this material will be incorporated into the GBIF training curriculum for use in GBIF-coordinated workshops and self-instruction (see Activity 1b).

GBIF continues its engagement with the BioDATA programme, funded by the Norwegian Agency for International Cooperation and Quality Enhancement in Higher Education (DIKU), focussed on developing biodiversity data management skills for students in Eastern Europe, South Caucasus and Central Asia. During 2021, the Secretariat worked with GBIF Norway to secure follow-up funding for a four-year 'BioDATA Advanced' project which aims to accelerate biodiversity research through DNA barcodes, collection and observation data with academic partners in Norway, the Russian Federation and South Africa.

2022 Work items

- Consolidate training resources aimed at academic users and publishers of GBIF-mediated data, for use in funded programmes and guidance for nodes
- Support implementation of the BioDATA Advanced project through partners in Norway, the Russian Federation and South Africa
- Promote GBIF at relevant fora including the 9th International Barcode of Life Conference, scheduled for August 2022, and the International Congress of Entomology, scheduled for July 2022
- Support for phylogenetic integration with occurrence data (€20K)

2021 Participant contributions

2022 Participant plans

Activity 5b: Document needs

Rationale (for the period 2017-2022)

GBIF-mediated 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. Further expert groups are considered during the current Strategic Plan period. Depending on resources, these groups will operate through a combination of face-to-face meetings and online or remote collaboration.

2021 Progress

The **GBIF task group on mobilization and use of biodiversity data for research and policy on human diseases** was formally established in November 2020, and during 2021 expanded its membership to include experts from Africa and Asia as well as Europe, North America and South America. Among the early activities of the group was to approve commission of a **systematic review of literature using GBIF-mediated data relating to human health**. Following an open call, Dr Francisca Astorga of Universidad de Chile was selected to carry out this review, which will help to inform GBIF's future work in meeting the biodiversity data needs of the biomedical research community.

Additional work requested by the task group is an initiative to promote mobilization of biodiversity datasets relevant to human health, through support for data papers to be published in a special journal issue following a call later in 2021. This will be enabled through a financial contribution from the **Tropical Disease Research centre (TDR)** of the World Health Organization, with GBIF providing funded helpdesk support to ensure that datasets associated with the data papers are correctly formatted and published through GBIF.org (see also Activity 3e).

With the completion and improved bibliographical content of checklists available in GBIF.org through the **Global Register of Introduced and Invasive Species (GRIIS)**, the Secretariat continued discussions with the IUCN Invasive Species Specialist Group on how data on invasive alien species can be better organized and curated, including through an improved research database managed by LaTrobe University in Australia. An updated research article describing GRIIS is in preparation, and the discussions will focus on how the collaboration can best support information and indicators addressing invasive species within the CBD post-2020 Global Biodiversity Framework (see Activity 5c).

With additional funding approved by the Executive Committee for the 2021 Work Programme, the Secretariat is scoping activities for the second half of the year to engage experts to promote mobilization of soil biodiversity data, recognized by the Food and Agriculture Organization (FAO) as a significant information gap to support links between biodiversity and food security. This work involves collaboration with the **Global Soil Biodiversity Initiative (GSBI)** and the **Soil Biodiversity Observation Network (SoilBON)** - see also Activity 3b.

2022 Work items

- Coordinate work of the expert group on mobilization and use of data on zoonotic diseases based on outputs from the systematic review commissioned in 2021
- Continue programme for mobilization of soil biodiversity data through partnerships and work streams initiated in 2021
- Continue partnership with the IUCN Invasive Species Specialist Group and the Convention on Biological Diversity Secretariat to improve content and curation of data relating to invasive alien

species

2021 Participant contributions

2022 Participant plans

Activity 5c: Support biodiversity assessment

Rationale (for the period 2017-2022)

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.

2021 Progress

Following the contract awarded in 2020 to the VertNet consortium in response to the [call for proposals to analyse the biodiversity data needs in the post-2020 Global Biodiversity Framework](#), the Secretariat worked with the authors to finalize the document which will be published later in 2021. The analysis presents a number of conclusions and recommendations for more consistent and transparent use of species occurrence data in indicators monitoring the goals and targets of the framework. Plans are being developed for a workshop to be held early in 2022 bringing together key partners to help implement the recommendations.

The Secretariat has engaged actively in the delayed preparations for the COP15 UN Biodiversity Conference, through virtual sessions of the Convention's subsidiary bodies in May and June, as well as the final meeting of the Open Ended Working Group in August/September 2021 addressing recommendations for the detail of the framework. Through collaboration with GEO BON and a number of national governments, these meetings have provided an opportunity to highlight the importance of primary data mobilization in successful implementation of the framework, and this is reflected in wording of several of the draft recommendations arising from these meetings.

In April ahead of the subsidiary body meetings, the Secretariat organized and chaired a [virtual round table on biodiversity monitoring and data](#) as part of the [Fifth Science-Policy Forum for Biodiversity and Eighth International Conference on Sustainability Science](#). Including speakers from GEO BON, the Chinese Academy of Sciences, UNEP-WCMC, NatureServe, iDiv and the Instituto de Investigación de Recursos Biológicos Alexander von Humboldt, the session provided an opportunity to reach a

broad science/policy audience regarding the need for increased investment in biodiversity data and monitoring. The key messages from the round table were included in an [information document](#) provided for the virtual SBSTTA24 meeting.

Work continued in 2021 to implement several of the collaborations included in the [Memorandum of Cooperation between GBIF and IUCN](#). This included a survey of experts involved in the IUCN Species Survival Commission to understand better the way that GBIF-mediated data is used in the Red Listing process, and issues identified by IUCN users. The survey will help to inform planned engagement activities including webinars to improve understanding of GBIF's services and functionality, as well as emphasising best practices for data citation.

The Secretariat implemented a new connection with the IUCN Red List web services to ensure that the latest version of the extinction risk category is available on the species pages of GBIF.org. Development has begun to implement filtering of occurrence data by Red List extinction risk category, but issues related to the handling of synonyms are being addressed before this functionality can go live on GBIF.org.

2022 Work items

- Through the *alliance for biodiversity knowledge*, hold a workshop on the outcomes of the analysis of primary data needs for the post-2020 Global Biodiversity Framework, involving partners in GEO BON, the Biodiversity Indicators Partnership, CBD Secretariat and others
- Implement a process enabling key checklists to be used in filtering occurrence data, such as Red Listed species and invasive alien species (carried over from 2020 and likely to be incomplete in 2021)
- Work with the new secretariat of GEO BON to improve connections between Biodiversity Observation Networks (BONs) and nodes at national, regional, thematic and global levels
- Continue collaboration with IUCN to implement the Memorandum of Cooperation, including exploration of publication of reference datasets associated with Red List assessments, and development of training material for use of data in the Key Biodiversity Area (KBA) process

2021 Participant contributions

2022 Participant plans

Activity 5d: Assess impact

Rationale (for the years 2017-2022)

GBIF Participants require clear evidence of the benefits arising from investments in national- scale content mobilization 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.

2021 Progress

The outcomes of the systematic review of research enabled by access to GBIF-mediated data, commissioned by the Secretariat in 2020 and carried out by Dr Mason Heberling were published in the [PNAS journal](#), providing a valuable reference on the range of science across many disciplines that is made possible by the GBIF collaboration and investment.

The total number of [research papers citing use of GBIF as a data source](#) passed the 6,000 milestone in July 2021, and the [2020 Science Review](#), published in February 2021, summarized 67 articles reflecting the range of research and policy applications supported by GBIF. Based on the citation trends observed by July 2021, the number of research papers citing GBIF use was anticipated to exceed 1,200 during the whole year, compared with 987 in 2020.

To support broader demonstration of GBIF's impact, the Secretariat secured approval from the Executive Committee during 2021 to commission a robust quantitative analysis of the value and impact of the GBIF network. Terms of reference for this study will include examining GBIF's investment and use values, contingent value, efficiency impacts, innovation value and return on investment. Responding to a recommendation of the GBIF 20-year review, this work is expected to begin in the last quarter of 2021 and be completed by the middle of 2022.

2022 Work items

- Publish quantitative analysis of the value and impact of the GBIF network, and promote in engagement with governments, funders and other partners
- Launch Ebbe Nielsen Challenge and Young Researchers Awards competitions for 2022 (€25k), and generate communications connecting past awardees with subsequent work and impact
- Explore options for a 'Nodes Award Scheme' to recognize progress, performance and innovation across the community
- Produce communication materials including short videos demonstrating the impact of GBIF's funded programmes such as BID and BIFA
- Complete and promote communication materials demonstrating GBIF's impact across thematic communities

2021 Participant contributions

2022 Participant plans