GBIF | Global Biodiversity
Information Facility

Draft 2019 Annual Work Programme

Event	25th meeting of the GBIF Governing Board (GB25) 16-17 October 2018, Kilkenny, Ireland
Agenda item	8. Presentation of draft 2019 Work Programme and 2019 Budget
Release date	17-09-2018
Doc. status	Draft ☒ Approved ☐
Action	For approval ☒ For discussion ☐ For information ☐

Documents download: Key meeting documents *and* background documents

This document serves both as the 2019 revision of the GBIF Implementation Plan for the period 2017-2021 and also as the GBIF Annual Work Programme 2019. Both elements are based upon the GBIF Strategic Plan for 2017-2021.

GBIF Implementation Plan 2017-2021 and Annual Work Programme 2019

Overview

This document serves both as the 2019 revision of the GBIF Implementation Plan for the period 2017-2021 and also as the GBIF Annual Work Programme 2019. Both elements are based upon the GBIF Strategic Plan for 2017-2021.
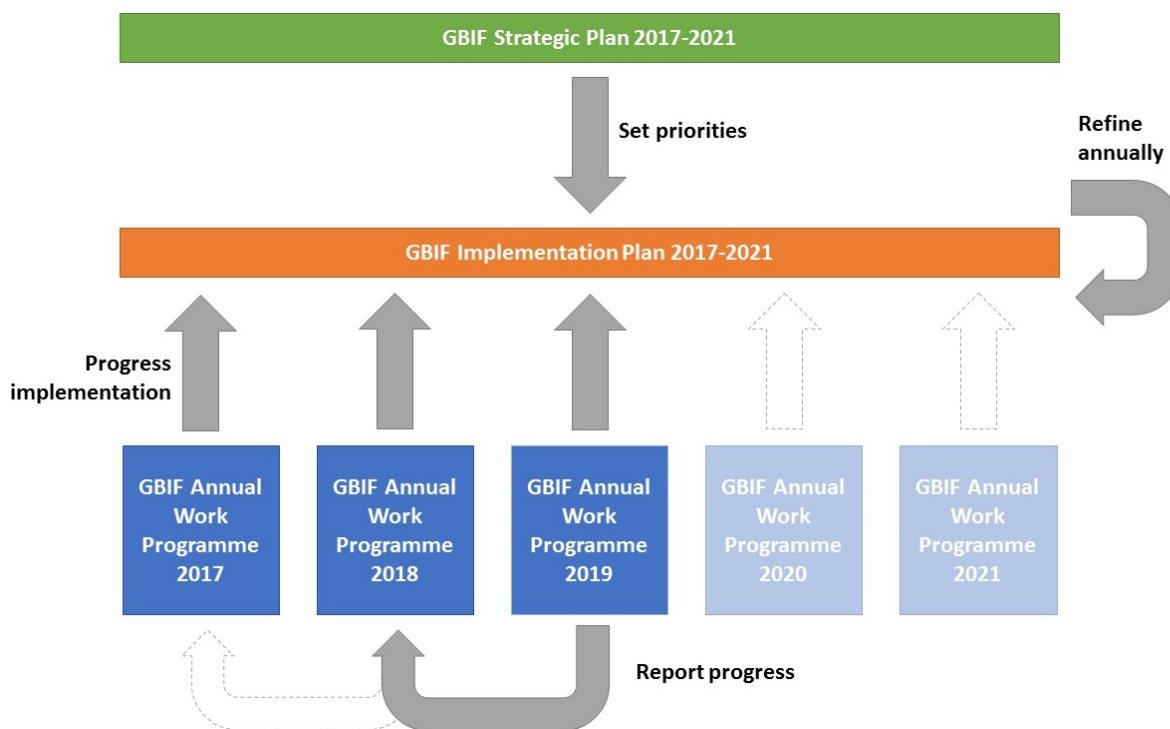
The relationship between these documents is represented by the following graphic.

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

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

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

The Implementation Plan is reviewed and refined each year.

The Implementation Plan is presented here with updates reflecting progress during 2018 and planned activities for 2019 (the 2019 Annual Work Programme).

The Annual Work Programme 2019 is summarized below. The revised Implementation Plan begins on page 9.

GBIF Annual Work Programme 2019

Information on activities planned for 2019 is included under each of the Activity sections included in the Implementation Plan, beginning on page 9.

For convenience, this information has also been collated on page 4 as an overview of the items for which approval is sought from the Governing Board at GB25.

Financial situation

Following the August revision of the GBIF budget, GBIF received additional delayed payments. As of September 2018, an additional €200,000 is available above the level projected in the August revision. As a result, even with the expenditure outlined above, GBIF is expected to close 2018 with net capital of around €1,575,000 (rather than €1,374,825, the figure projected in the August 2018 revision).

The BID programme continues into 2019 and will include around €120,000 in funding for work by GBIF Secretariat staff. Similarly, €20,000 in staff time will be funded from the remaining budget for BIFA. Lastly, the European Commission has funded the Synthesys+ project "to unify European natural science collections, effectively transforming them into an integrated data- driven pan-European research infrastructure". GBIF is funded within this project to advance joint planning for infrastructure components (following the discussions initiated at GBIC2). This will support €50,000 in GBIF staff activity during 2019 (and €50,000 in 2020, €20,000 in 2021 and €20,000 in 2022).

Given the level of capital held by GBIF, the 2019 budget includes a larger allocation for work programme expenditure (i.e. non-salary costs allocated to calls, workshops, contracts, etc.) than has been typical in recent years: €513,000. The planned expenditure for 2019 is €386,294 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.

Secretariat staffing

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

- Participation and Engagement (eight staff)
- Data Products (four staff)
- Informatics (eight staff, with one additional position to be filled)
- Administration (five staff, plus student support)

Planned expenditure

In addition to salary and operating costs, including support for GBIF Governance mechanisms, funding is reserved in the Work Programme 2019 for the following activities:

- **Tools to Manage Documentation on Data Publishing, etc.: €30,000** – See: *Activity 1b - Strengthen Skills*.
- **Global Nodes Meeting: €60,000** – See: *Activity 1c - Equip Nodes*.
- **Nodes Training Event: €20,000** – See: *Activity 1c - Equip Nodes*.
- **Support for Living Atlas Activity: €40,000** – See: *Activity 1c - Equip Nodes*.
- **Arabic Translations for GBIF.org: €8,000** – See: *Activity 1e - Expand National Participation*.
- **GBIC2 Follow-up to Develop Global Alliance: €50,000** – See: *Activity 1f - Plan Implementation*.
- **Capacity Enhancement Support Programme: €80,000** – See: *Activity 1g - Coordinate Resources*.

- **Revision of Data Standards and Modernization of IPT: €50,000** – See: *Activity 2a - Modernize data standards*.
- **Development Work for Catalogue of Life Plus: €80,000** – See: *Activity 2b - Deliver Names Infrastructure*.
- **Engage Human Disease Vector Research Community: €25,000** – See: *Activity 3b - Expand Data Streams*.
- **External Review of GBIF: €70,000** (additional to €30,000 set aside in 2018) – See: *Activity 5d - Assess Impact*.

Areas to expand activity (subject to additional funding)

As explained in the Draft GBIF 2019 Budget, there is always uncertainty around the exact level of funding which GBIF will secure in any given year. This Work Programme is based on a realistic assessment of likely funding levels in 2019.

It is possible however that 2019 funding levels will be higher and GBIF will have the opportunity to carry out additional activity. The following are recommended for such activity, and approval is sought from the Governing Board for the Executive Committee and the GBIF Secretariat to expand activity in these areas if budget levels are considered adequate during 2019:

- **Further Effort to Modernize IPT:** €50,000 is already allocated in the budget for work to revise the GBIF data standards and to begin modernization of the Integrated Publishing Toolkit (IPT) – additional funds (c. €100,000) would accelerate the necessary development work and allow GBIF significantly to raise the flexibility and to lower the technical threshold for new data publishing. See: *Activity 2a - Modernize data standards*.
- **Short-term Appointments from GBIF Community to GBIF Secretariat:** During 2017-2018, GBIF was fortunate to have Maofang Luo from the Chinese Academy of Science based at the GBIF Secretariat, learning more about GBIF operations and contributing to activities related to Asian data mobilization. Ideally, GBIF would run an ongoing programme for representatives from GBIF nodes to spend time based in Copenhagen. We consider this an important approach for GBIF to increase skill transfer and to build capacity around GBIF, particularly in regions with lower resources or lower participation within the network. Further work would be required to develop a model and programme for such appointments, but an allocation of around €100,000 would allow GBIF to trial the approach with some initial appointments. See: *Activity 1g - Coordinate Resources*.

We welcome any efforts from GBIF participants to secure supplementary funds for these purposes. Additionally, we encourage participants to explore opportunities for supplementary funding in the following areas:

- **New Capacity Enhancement Campaigns:** The BID programme comes to a conclusion during 2019. GBIF has demonstrated its capability and suitability as a channel for such capacity enhancement activity. We are keen to work with participants to seek funds for future capacity enhancement campaigns, particularly for remaining regions with under-represented or no national participation, namely: 1) the Former Soviet Union including Central Asia, the Caucasus and the Baltic States; 2) the Balkans and other Central and East European countries; 3) North Africa and the Middle East; and 4) China, South and Southeast Asia. See: *Activity 1e: Expand National Participation*.
- **Support for the Living Atlases Community:** GBIF is providing some core funding for meetings and small contracts in support of the Living Atlases community. However, there is a need for more coordination between the community projects to establish requirements, work with the ALA team, seek funding for implementations, translate text and documentation, and develop open source initiatives. We encourage those participants which

are considering deploying a Living Atlas instance to explore opportunities to fund or co-fund a full-time coordinator on behalf of the community. See: *Activity 1c - Equip Nodes*.

Overview of 2019 Work Items

This following is a summary of all 2019 Work Items proposed in the 2019 update of the *Implementation Plan* below. It is intended to summarize the work items for which approval is sought from the Governing Board at GB25. Accordingly, it does not include the additional commitments made and reported by Participants. These are described in the Activity sections within the body of the document below.

Activity 1a - Focus on people

- Evaluate early versions of expanded user profile function in GBIF.org through monitoring of uptake and a survey of GBIF users and stakeholders; based on this evaluation, consider further expansion and adjustment.
- Monitor usage of the GBIF Community Forum and conduct a survey among stakeholders, to inform a recommendation on whether to maintain the platform beyond a 15-month pilot period.

Activity 1b - Strengthen skills

- Package and promote Data Use for Decision Making workshop materials for wider re-use in GBIF programmes and beyond, e.g. for CBD and IPBES capacity building needs.
- Explore opportunities for additional training workshops on data use for decision making in more regions, attached to existing programmes or using supplementary funds.
- Further develop explainer texts and videos to complete self-instruction package for basic introduction to data mobilization and use in GBIF.
- Establish system for reviewing, prioritizing, updating and creating new documentation to support GBIF data publishers and users, including a lightweight editorial panel and peer review process, with annual funding for commissioning high-priority guidance from subject-matter experts. €30,000 has been allocated in the budget to establish the system, including deployment of software tools.

Activity 1c - Equip Participant nodes

- Prepare and run Global Nodes Meeting in association with GB26. €60,000 has been allocated in the budget to support the meeting.
- Develop training modules on managing a GBIF Participant node, including online version and materials to support on-site workshop.
- Run training workshop on managing a GBIF Participant node in association with the 2019 Global Nodes Meeting. €20,000 has been allocated in the budget to support the training.
- Support further development of the Living Atlases community including coherent branding, establish ongoing workshop and documentation programme outside CESP. €40,000 has been allocated for this purpose.
- Consolidate Biodiversity Open Data Ambassador programme with issuing of digital badges, listing ambassadors on country pages and capturing feedback on promotion of GBIF at events.

Activity 1d - Equip data publishers

- Develop helpdesk services for new data publishers using cloud-hosted IPT installations, managed by volunteer mentors experienced in data publishing.

Activity 1e - Expand national participation

- Add support for Arabic language content in GBIF.org, including technical adjustments to accommodate right-to-left script, and commissioning translations of the user interface as well as key content. €8,000 has been allocated in the budget for translation costs.
- Use BID closing regional meetings in Caribbean, Pacific and sub-Saharan Africa to promote

further national participation in these regions, both for countries with funded projects and for additional countries invited from each region.
● Build on current funded projects, and develop additional funding proposals, to implement an intensive engagement strategy for remaining regions with under-represented or no national participation, namely: 1) the Former Soviet Union including Central Asia, the Caucasus and the Baltic States; 2) the Balkans and other Central and East European countries; 3) North Africa and the Middle East; and 4) China, South and Southeast Asia

Activity 1f - Plan implementation

● Develop standardized, annual information requests from Participant nodes to capture relevant updates and work plans based on the GBIF Strategic Plan and Implementation Plan structure, enabling greater visibility for node activities on GBIF.org and supplementing the country-level metrics provided in the automatically-generated Activity Reports.
● Follow up from GBIC2 workshop by leading in communications, consultations and working groups to develop a global alliance for biodiversity knowledge. The EU-funded *Synthesys+* project will provide some funding in 2019 and subsequent years for GBIF staff time and workshops for this purpose. €50,000 has also been allocated in the GBIF 2019 budget to support additional workshops, meetings or communications materials.

Activity 1g - Coordinate resources

● Subject to approval by GB25 of recommendations made in the five-year review of the Capacity Enhancement Support Programme (CESP), launch a 2019 call for proposals under a renewed CESP, taking account of the feedback received during the review. This call specifically to exclude support for core activities under the Living Atlas programme, to be funded and coordinated separately (see Activity 1c). €80,000 has been allocated in the budget for the call.
● Complete implementation of the current BID programme, including a) closing all remaining projects and delivering necessary documentation to the EU; b) commissioning impact report to analyse the return on investment including mobilized data, capacity enhancement and additional policy support in the target regions; c) publishing guiding examples and best practices developed through BID on mobilization and use of data for priority policy needs; d) organizing closing meetings in each target region to showcase outcomes and develop sustainability plans including through increased GBIF participation (see Activity 1e); and e) organizing a BID closing event, probably in association with GB26, to promote the outcomes to the wider community and propose next steps.
● Based on the outcomes of the community consultation on the BID programme, develop proposals for a second phase of BID in collaboration with the European Union and the BID Steering Committee
● Implement a fourth call for proposals under the Biodiversity Information Fund for Asia (BIFA), reflecting priorities agreed by Asian nodes and the BIFA Steering Committee, and conclude discussions with the Government of Japan to decide the future of this supplementary fund beyond 2019.
● Work with all GBIF regions to identify and target potential funding streams to support additional capacity enhancement for data mobilization and use, building on the BID and BIFA models.
● Roll out selected grant management system to streamline project calls, assessment, selection, implementation, budgeting and reporting, with continuous refinement and evaluation of the system to ensure it is meeting the needs of the community.
● Launch 2019 calls for the Ebbe Nielsen Challenge and Young Researchers Awards, taking into account recommendations from the Science Committee based on the 2018

programmes.

Activity 2a - Modernize data standards

- Work in open consultation with the expert community to document a unified information model that covers the scope of content GBIF accommodates and supports richer publishing of interconnected information from external databases. This model should be presentable in various languages. €50,000 is reserved in the budget to support two associated workshops.
- Provide a set of data exchange profiles for sharing data within GBIF in conformance with the unified information model. This should include both existing and new standards and the necessary controlled vocabularies (e.g. sampling protocols).
- Redesign the GBIF Integrated Publishing Toolkit to support these profiles and to address infrastructural needs (ability to install locally, or use a GBIF hosted solution).
- Provide documentation for the data model and for the associated services offered through GBIF.org.
- Review and redesign GBIF data management to accommodate the unified information model as part of data ingestion, quality control and processing necessary.
- Continue technical discussions with other data aggregators to seek closer alignment in practice and, as far as possible, implementation of aggregation and indexing processes.

Activity 2b - Deliver names infrastructure

- Replace the existing processes for constructing the monthly and annual Catalogue of Life checklists and for constructing the GBIF taxonomic backbone with a single solution that delivers both products (together forming a "provisional checklist"). This work is carried out under the NLBIF-funded Catalogue of Life Plus (CoL+) project. €80,000 has been allocated in the budget to support GBIF development costs, with €20,000 in matching funds from the NLBIF funding to CoL+.
- Plan for existing GBIF checklistbank capabilities (registry integration, images, descriptions, APIs) are correctly handled within the CoL+ implementation.
- Ensure that updates to the provisional checklist are reflected directly in the GBIF data index.
- Manage the CoL+ software as part of ongoing infrastructure supported at GBIF.
- Work with Catalogue of Life and other partners to develop responsive expert communities to support sectors which currently lack adequate support and to improve alignment with other checklist efforts including World Flora Online and WoRMS.

Activity 2c - Catalogue collections

- Deliver a web version of the Global Register of Biodiversity Collections (GRBio)/Global Register of Scientific Collections (GRSciColl) dataset as part of ongoing infrastructure supported at GBIF.
- Work with GRSciColl, Index Herbariorum, GBIF nodes and others to establish sustainable processes for updating the registry.
- As the TDWG Natural Collections Descriptions Interest Group revises the NCD standard, enhance the registry to accept NCD documents and develop mechanisms via IPT or otherwise for collections and nodes to publish and update these data.

Activity 3a - Identify priority gaps

- Prioritize data gaps through a 'hunger map' approach, including systematic documenting of user-communicated data content needs and matching them with available data; and extending the analysis of data searches, including no-search results, to help understand categories of data for which GBIF users do not currently find what they are looking for.
- In collaboration with IPBES and others, capture priority knowledge gaps for which lack of accessible primary biodiversity data is identified as a significant cause of uncertainty.
- Based on the gap analysis approaches mentioned above, in combination with earlier efforts such as the 2017 data mobilization priorities , develop actionable guidance for nodes, publishers and funders on addressing priority gaps through mobilization targets and

strategies based on spatial, temporal, taxonomic and thematic dimensions of biodiversity data.

Activity 3b - Expand data streams

- Extend 2018 work on matching stable Operational Taxonomic Units (OTUs) based on DNA sequencing to the GBIF taxonomic backbone, with targets of establishing dynamic links with the Barcode of Life Data System (BOLD) and incorporating occurrence data derived from environmental DNA sampling (eDNA) through the European Bioinformatics Institute (EBI).
- Run a campaign to mobilize data relating to vectors and hosts of human diseases, establishing an expert group to identify priority needs for biodiversity data supporting disease research, critical gaps in availability of such data in GBIF.org, and potential sources of data to fill these gaps. Campaign will use this analysis to engage directly with known holders of missing data and support publication through GBIF, as well as informing data mobilization priorities for use by nodes, publishers and funders (see Activity 3a). €25,000 has been allocated in the budget.

Activity 3c - Engage data holders

- Work with DiSSCo project team to maximize opportunities for mobilizing collections data from European institutions, including in countries not yet participating in GBIF.
- Based on discussions during Q4 2018, work with UN Environment WCMC and industry bodies (e.g. Proteus Partnership, IPIECA, IAIA) to establish node and helpdesk services for publishing private sector data, with options to use national nodes or hosted repository to support data mobilization.
- Establish framework for regular updates of 2016 evaluation of contribution of citizen science contribution to GBIF data index, based on dataset tagging or other means.
- Work with iNaturalist, iDigBio and nodes community to maximize opportunities for public engagement in GBIF data mobilization provided by the City Nature Challenge and WeDigBio

Activity 3d - Rescue datasets

- Establish workflow for prioritizing and mobilizing potential data sources reported through the 'suggest a dataset' tool, including involvement of nodes, mentors and crowd-source solutions.

Activity 3e - Liaise with journals

- Build on engagement with selected authors and academic publishers in 2018 to develop standard workflows and simple recommendations for depositing supplementary data in formats suitable for GBIF publication, as well as data citation practices enforced through editorial guidelines, with the aim of acceptance by all major publisher groups during 2019.

Activity 4a - Ensure data persistence

- Coordinate effort to achieve CoreTrustSeal data repository certification for relevant portions of the GBIF infrastructure. This will include the data management services within GBIF.org, but will also seek to identify a set of trusted repositories for publishing datasets within the GBIF network.
- Develop documentation for all stages in the GBIF data ingestion process, to enable data provenance to be fully understood by users of GBIF data.

Activity 4b - Assess data quality

- Review, consolidate and update existing documentation for data publishers. In particular, provide clear guidance on minimum requirements for published data.
- Develop metrics to track the completeness of core data elements and the degree to which supplied content is appropriate.
- Supply clear indicator measures for the completeness and usability of data as part of GBIF.org dataset pages, based on the example of the GEOLabel data branding model.

- Extend data quality assessment to include aspects only detectable above the level of individual records.
- Start to assess the patchiness of indexed data (geographical clustering, misleading accuracy or precision of coordinates), including evaluation of the apparent causes of data patchiness (e.g. grid-based relevées, other sampling approaches, country centroids), and include measures of data patchiness in the data index, at both dataset and record level in the data index
- Ensure that users of data are able to identify datasets or records that do not fulfil their criteria for geo-accuracy (including via facets within the portal, via the API, and in downloads).

Activity 4c - Enable data curation

- Continue exploration of the use of the GBIF data index to support stable persistent resolvable identifiers for all specimens and occurrence records.

Activity 5a - Engage academia

- Enhance engagement with academic communities through Open Biodiversity Data Ambassadors (see Activity 1c), and sharing best practices from nodes promoting biodiversity informatics in academic curricula
- Develop materials to guide academic audiences on use of GBIF-mediated data in thematic areas, e.g. invasion biology, epidemiology (see also Activity 1b)

Activity 5b - Document needs

- Review and consolidate responses to recommendations from previous fitness-for-use groups, before commissioning new expert groups other than the group advising the campaign on data for disease vectors and hosts (see Activity 3b)
- Analyse and evaluate the requirements for pre-generated downloads based on the analysis of incoming user requests for custom downloads that cannot be served through the user interface at GBIF.org or API access

Activity 5c - Support biodiversity assessment

- Continue contributing to GEO BON activity to develop Essential Biodiversity Variables, particularly in regard to Species Distribution, Population Abundance and Community Composition. As the global infrastructure for organizing species occurrence data in support of EBVs, work with GEO BON to align with the EBV framework.
- Enhance GBIF.org to ensure that sampling event data can be accessed in forms appropriate for further processing as EBV data.
- Explore options for new data download formats which provide pivoted views of available data in forms appropriate for processing as EBV data. Such pivoted views are likely also to assist other users with simplified access to relevant GBIF data.

Activity 5d - Assess impact

- Perform external review of GBIF operation and functions. At GB24, the Governing Board approved €30,000 in the 2018 budget to be allocated towards an external review of GBIF and its operations. Following explorations within the Executive Committee, it is agreed that the review can best be handled via a contract to CODATA, which led the first review of GBIF, completed in 2005. It has now been 20 years since the OECD Megascience Forum recommended the establishment of a Global Biodiversity Information Facility, so an external review led by CODATA as a representative of the global scientific community will be timely. An additional €70,000 is included in the 2019 budget (for a total of €100,000) to support a broad review of GBIF's operation, services, governance and place within the broader landscape of biodiversity-related organizations.

GBIF Implementation Plan 2017-2021

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

The following notes provide context for the activities outlined here.

Delivery

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

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

Structure of GBIF

GBIF operates on three primary levels:

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

This Implementation Plan seeks to address needs at all three of these levels while aiming to reinforce the value of GBIF for stakeholders at each level. It is highly desirable to increase coordination and feedback between these levels. The 2019 version of the document includes reports from a few GBIF Participants on their 2018 activities and 2019 plans. It would be most beneficial to receive inputs from a wider range of Participants to give a clear picture of everything occurring in the network and to assist the Secretariat and all Participants with better alignment and reuse.

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

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

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

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

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

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

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

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

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

Timeline

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

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

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

Resources

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

These resources combine as follows:

- **Core Funding**
  - Supports GBIF Governance structures (Governing Board, committees)
  - Supports staffing and operation of GBIF Secretariat to coordinate global activity, including:
    - Delivery of global-level IT infrastructure (GBIF.org)
    - Coordination of delivery of software components to support Participants and Data Publishers
    - Coordination of capacity enhancement activity and information materials for Participants, Data Publishers and Users
    - Coordination of supplementary-funded activity and Participant investments in shared GBIF activity
    - Monitoring and reporting usage and benefits
    - Administrative support for GBIF activities
    - Communication, outreach and partnership activities
  - Limited funds to support workshops, capacity enhancement and developments to expand or enhance GBIF

- **Supplementary Funding**
  - Additional funds to support workshops, capacity enhancement and developments to expand or enhance GBIF
  - Funding to support Participants and Data Publishers with content mobilisation or development of national Biodiversity Information Facilities

- **Participant Investments**
  - Operating GBIF nodes
  - Engaging with and supporting data publishers and users
  - Curating data from Data Publishers
  - Additional funds or staffing to support workshops, capacity enhancement and developments to expand or enhance GBIF

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

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

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

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

Implementation Plan structure

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

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

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

Overview of Priorities and Activities

Priority 1 – Empower Global Network

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

Activity 1a - Focus on people

Rationale

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

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

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

Approach

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

Tasks

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

    1.a.ii   Consolidate use of digital badges to record skills and contributions

Progress

Users of GBIF.org can already associate their account with their ORCID, and log in using ORCID as the identity provider. Further implementation of this activity during 2018 has been somewhat delayed as GBIF explores the implications of the European Union's new General Data Protection Regulations (GDPR) on measures that directly target publicly-available information about individuals. Having taken the necessary legal advice, GBIF can proceed with confidence to implement an expanded user profile, enabling users to link relevant information about their own activities and affiliations, for public display on GBIF.org if this option is selected. This will include, for example, information on the user's data publishing record, citation of GBIF-mediated data in research, skills certification and volunteer roles through digital badges (see below in this section). Further information may include individual species occurrence records for which the user is named as recorder or identifier, and even verified species descriptions published in the taxonomic literature. The first test version of this new feature may be implemented within 2018, or scheduled for 2019. A mockup of the potential content of these profiles may be seen here.

A new [GBIF Community Forum](#) was launched in May 2018, using the open-source Discourse platform to provide an open discussion space for the GBIF community. Moderated by the Secretariat communications team, the forum combines open discussion among registered users with closed groups for internal communication between nodes. Following a 15-month trial period, usage of the platform will be evaluated and recommendations on its future presented to the Global Nodes Meeting and Governing Board in Q4 2019. The Community Forum is intended to replicate and enhance those functions of the former GBIF Community Site not available through GBIF.org and other platforms such as GitHub, Open Science Framework, Google Drive and mailing lists. The Community Site itself was retired during 2018.

GBIF continued development of the digital badge scheme to recognize skills and roles among individuals in the GBIF community. During 2018, three badges recognizing mentors and certified trainers within the Biodiversity Information Development (BID) programme, were extended to wider application across GBIF programmes and projects, as GBIF Volunteer Mentor (signifying general availability to act as a mentor and familiarity with the mentoring programme), GBIF Project Mentor (signifying specific experience in mentoring a project or providing on-site mentoring at training workshops) and GBIF Certified Trainer (signifying experience of taking part as a trainer in a GBIF capacity enhancement workshop). Three new badges are in preparation covering the new roles of Node Mentor (see Activity 1c), Biodiversity Open Data Ambassador (see Activities 1c and 5a) and Volunteer Translator (see Activity 1e).

Many participants contribute to the distributed collaborative GBIF help desk, as reflected in the [list of volunteer mentors.](#)

2019 work items

- Evaluate early versions of expanded user profiles through monitoring of uptake and a survey of GBIF users and stakeholders; based on this evaluation, consider further expansion and adjustment
- Monitor usage of the GBIF Community Forum and conduct a survey among stakeholders, to inform a recommendation on whether to maintain the platform beyond a 15-month pilot period

Activity 1b - Strengthen skills

Rationale

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

Approach

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

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

Tasks

1.b.i   Develop collaborative help desk capability – including discussion groups, support materials and FAQs

1.b.ii   Articulate training and capacity enhancement activities around a single, global curriculum for GBIF

1.b.iii   Deliver explainer texts and videos to support self-directed instruction and learning

Progress

During 2018, the Secretariat in collaboration with trainers from the GBIF community has made significant progress in refining and streamlining course materials developed through the BID programme to support data mobilization skills among project teams, for adaptation and re-use in a wide range of programmes in any region. In addition, a new set of training modules for use of open-access biodiversity data in decision making has been developed for three separate regional training workshops within BID, also with the intention of making standard materials available for rolling out in any region and for re-use by nodes.

The consolidation of course materials for data mobilization, tested and adapted in four separate BID capacity enhancement workshops in Africa, the Caribbean and Pacific in 2016-17, enabled replication of the training modules in four additional workshop settings during 2018: 1) in Montenegro in April, within a project on biodiversity information management and reporting in the Western Balkans, funded by Deutsche Gesellschaft für Internationale Zusammenheit (GIZ); 2) in

Moscow in April to support mobilization of data from Russian institutions in a project funded by GBIF Finland; 3) in Beijing in June, to support data mobilization projects funded in the latest call under the Biodiversity Information Fund for Asia (BIFA) using supplementary funds provided by the Ministry of Environment, Japan; and 4) at Lake Baikal, Siberia in September, in a training event on mobilization of data using both GBIF tools and barcoding techniques through the Barcode of Life Data System (BOLD), supported through Norway's ForBio programme.

In collaboration with several partners including UNEP-WCMC, IUCN and SANBI, the Secretariat developed an additional training curriculum to support data use for decision making during 2018 as part of the BID programme. The curriculum is designed to equip participants with the theoretical skills they require to produce selected biodiversity information products and to integrate them into the decision-making process. Specific topics included in the curriculum include data cleaning, species distribution modelling and mapping, assessing species conservation status and mainstreaming biodiversity data by identifying key policy entry points. This curriculum was rolled out at a workshop in Cape Town, South Africa in April for all BID-funded project teams in Africa and repeated for Pacific projects in New Zealand in September, with a further edition planned for November in Suriname for BID Caribbean projects.

The mentoring and collaborative helpdesk approach developed through the BID programme has been significantly widened during 2018, to be applied across the GBIF community. A webinar in July helped to recruit new mentors, and a new web page explains the mentoring approach. By September 2018, 76 volunteer mentors had offered to contribute their time and skills to support capacity development in the network. Volunteer mentors contributed on-site support at all of the training workshops mentioned above.

The set of explainer texts for key GBIF activities expanded in 2018 with an explainer on data hosting and two more are in the pipeline on Creative Commons licensing and API access. A set of explainer texts on the use of GBIF-mediated data is in preparation with a dedicated section planned for this topic in the top-level navigation of GBIF.org.

Participant contributions

Argentina, Australia, Belgium, Benin, Colombia, France, Ghana, Madagascar, Norway, Poland, Portugal, South Africa, Spain, Sweden, United States, Zimbabwe, UNEP-WCMC, VertNet: These participants supplied trainers and mentors for workshops organized by GBIF during the year under BID and BIFA, and in Montenegro and Russia for capacity enhancement in data publishing and use.

Colombia: "GBIF habla Español." GBIF Spain and SiB Colombia have produced an animated promotional video about GBIF and 4 additional short videos that showcase multiple experiences from people that take part in the global network from both countries. The materials produced were shared through a wiki hoping that they may serve to build upon them. The animated video is the most viewed video on the GBIF youtube channel and incorporates subtitles in english, french, chinese and portuguese, following its success after its launch on GB24.

Colombia: "CREAR COMPARTIR TRANSFORMAR: guía de acceso abierto del SiB Colombia." SiB Colombia's open access guide is a simple and attractive material, designed to socialize the new policies for accessing and using the data and information available through SiB Colombia. It is also a tool to clear up doubts about the universe of open access for all people interested in the construction of open knowledge about biodiversity. All the materials are available to anyone who wishes to build on them. More info: (https://goo.gl/oMbXQd)

Colombia: Implementation of a capacity enhancement project based on "Sharing VertNet experiences and tools on biodiversity data quality with the Spanish-speaking community". The CESP project concluded with 64 cleaning datasets and the translation from English to Spanish of the Data Migrator Toolkit documentation. There were also some modifications in the SiB Colombia

quality workflow, based on VertNet experiences and tools. It allowed to improve the data quality in more than one million occurrence records published through SiB Colombia's IPT, in 29 Darwin Core prioritized elements (register, event, taxonomy, and georeferencing elements).

Norway: GBIF Norway organized a regional Nordic training workshop on the use of GBIF mediated data in R during the Nordic Oikos conference in Trondheim 18-19 February 2018 (see also activity 5a).

South Africa: SANBI-GBIF is developing a research publication titled: Towards a curriculum for Biodiversity Informatics, as an outcome of a GBIF Nodes Meeting Task Group, summarising survey results.

South Africa: SANBI-GBIF, through the Africa Biodiversity Challenge (ABC) Project, is engaged in the development of the training modules around data application and use focused on mainstreaming biodiversity into decision-making, in collaboration with UNEP-WCMC and the GBIF secretariat as a resource for further training events both physical and online.

Spain: Continued to provide essential support for GBIF-coordinated training through hosting of the [GBIF e-Learning platform](#).

2019 work items

- Package and promote Data Use for Decision Making workshop materials for wider re-use in GBIF programmes and beyond, e.g. for CBD and IPBES capacity building needs
- Explore opportunities for additional training workshops on data use for decision making in more regions, attached to existing programmes or using supplementary funds
- Further develop explainer texts and videos to complete self-instruction package for basic introduction to data mobilization and use in GBIF
- Establish system for reviewing, prioritizing, updating and creating new documentation to support GBIF data publishers and users, including a lightweight editorial panel and peer review process, with annual funding for commissioning high-priority guidance from subject-matter experts

Participant plans for 2019

- **Australia:** Contribute to maintenance of documentation for the Living Atlas community.
- **Belgium:** Complete set of explainer texts in multiple languages: French and Dutch translations
- **France:** Continue support for capacity enhancement activities and organize workshops at national level.
- **Norway:** GBIF Norway and the Norwegian Research School in Bioinformatics (ForBio) secured together with the GBIFS funding for a three-year (2018-2021) biodiversity digitization and mobilization training program (BioDATA) in the Eurasia region. The training program secured project funding from the Norwegian Centre for cooperation in Education (SIU). (See also activity 1e and 3b).
- **South Africa:** SANBI-GBIF is recruiting a Biodiversity Informatics Support Officer to support training and engagement portfolio of work. This will support national and regional efforts and tie into GBIF helpdesk requirements as needed.

Activity 1c - Equip Participant nodes

Rationale

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

Approach

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

Tasks

1.c.i   Perform ongoing nodes capacity self-assessment

1.c.ii   Build reusable national/regional platform tools centred on ALA software

1.c.iii   Develop online collaboration through GBIF.org helpdesk to assist and mentor node managers

1.c.iv   Explore alternative options for lightweight national portal solutions, e.g. CKAN

Progress

The Secretariat is reviewing the well-used guidance document Establishing and Effective GBIF Participant Node, and will provide an updated version before the end of 2018, including better links with the capacity self-assessment tool, guidance for organizational nodes and on the division of roles between node managers and GBIF Heads of Delegation. In addition, the review has identified the need for more examples and case studies from across the nodes community to complement the generic guidance materials. Work has begun to document guiding examples on thematic aspects of node activity, with early contributions contributed by Belgium, Australia and Norway on collaboration with IPBES, engaging academia and mobilization of data from the private sector, respectively.

The national capacity self-assessment tool is in wide use across the nodes community, and has been especially helpful in tracking progress in capacity enhancement over the course of projects funded through the BID programme. A new, more user-friendly format of the tool is under development, making it easier to track progress over time, with new versions contributed by the community in French, Spanish and Portuguese.

In the final quarter of 2018, GBIF is trialling a new Node Mentoring programme in which dedicated individual volunteer mentors (see Activity 1b) are assigned to newly-established nodes, with South Sudan, Zimbabwe and Vietnam providing the first examples.

The bimonthly GBIF network webinars have continued to generate a good response from nodes, as a way of keeping the community updated on implementation of the Work Programme as well as sharing examples and best practices from around the network. Migration from the Adobe Connect to

the Zoom platform has enabled much more efficient packaging of videos from the webinars, and appears easy to use for participants.

The new [Biodiversity Open Data Ambassador](#) programme was developed and rolled out during 2018. Initially scoped at a joint meeting of the Nodes Steering Group and Science Committee in Copenhagen in February, the programme is designed to enlist willing biodiversity professionals to help advocate open data practices in relevant research and policy communities, using a package of tools and resources provided by GBIF. Close connections will be encouraged with nodes to help strengthen connections with national user groups. The first set of ambassadors was recruited at the GBIC2 meeting in July, and the programme will be further developed in the last quarter of 2018, including development of the information toolkit available to ambassadors.

The Living Atlas activity merits greater investment. The ALA has provided significant support from within its team, and GBIF has supported workshops through the CESP funds. For greater stability, it would be preferable for the community of stakeholders outside Australia to secure funding to support a full-time coordinator to mediate requirements with the work of the ALA team and to assist in planning to support more ALA components through collaborative development and support strategies.

Participant contributions

Australia: The ALA co-coordinated the CESP Madrid workshop with GBIF France and GBIF Spain which provided training for staff from GBIF nodes on the Living Atlas platform. The ALA developed the necessary training materials to support this workshop and these are available for future events. The ALA has implemented a support portal and ticketing service that can be shared and aggregated by all Nodes. It enables efficient authoring of support articles and management of support tickets. This is based on the Australian model for ticket and support provision of Australian Computer Research Infrastructure and can be employed world wide for greater sharing of information, common problems and solutions. The ALA team has provided support for numerous nodes and other countries at different stages of their development, this year including: Guatemala, Phillipines, Germany (iDiv), UK, Brazil (SiBBr) and Suriname, and with continued support for countries already using the software, including Canada, Spain and France.

Colombia: Participation in CESP "Regional capacity enhancement to Latin America by establishing Chile's node" On April 2018 SiB Colombia Node made a publishing and data quality workshop, as support for the enhancement of the Chile's emergent node. The workshop include potential Chilen biodiversity data publishers, with 47 participants from more than 20 organizations.

France: GBIF France has helped to develop a portal based on the ALA software for GBIF Benin and contributes to the Living Atlases community, advising on the technology, leading the organisation of workshops in Madrid (hosted by GBIF Spain) and TDWG. GBIF France launched the web site of the Living atlases through the CESP project. GBIF France has also managed the documentation correlated to the project

Japan: Two workshops for local biodiversity data facility (S-Net) convened. Two additional workshops convened. Collaboration with academic community to communicate GBIF relevant matter in biodiversity data (CC licence, ORCID).

Japan: Science Museum Net (S-Net), a website for biodiversity data for local (domestic) use has been updated, including improvements to support fossil records.

Korea: Hosted and installed new KBIF IPT server from KISTI to NSMK as a KBIF secretariat (updated the coordinates for occurrences in Korea's published datasets)

Korea: Translated the 'IPT User Manual' available in GBIF and distributed it to relevant organizations and people in Korea.

Korea: Organized the official KBIF committee consisted of national governments and universities in

Korea to promote sharing biodiversity data and national strategies (total 15 committee members from Ministry of Environment, Ministry of Maritime Affairs and Fisheries, Ministry of Agriculture, Food and Rural Affairs, Cultural Heritage Administration, National Forestry Administration, Rural Development Administration, Kyungpook National University, Pai Chai University and Mokpo National University)

Norway: The GBIF CESP funded European Bireme developed recommendations for involving the GBIF Nodes in national biodiversity status reporting to the European Commission (see also activity 5c).

2019 Work Items

- Prepare and run Global Nodes Meeting in association with GB26
- Develop training modules on managing a GBIF Participant node, including online version and materials to support on-site workshop
- Run training workshop on managing a GBIF Participant node in association with the 2019 Global Nodes Meeting
- Support further development of the Living Atlases community including coherent branding, establish ongoing workshop and documentation programme. €40,000 has been allocated for this purpose. This recognizes that GBIF needs to provide stable support for this community, which is increasingly important to many participants. In 2019 and in future years, Living Atlas work should be separately funded rather than subsumed under the CESP call.
- Consolidate Biodiversity Open Data Ambassador programme with issuing of digital badges, listing ambassadors on country pages and capturing feedback on promotion of GBIF at events

Participant plans for 2019

- **Australia:** Continue support for the Living Atlases community, and coordinate efforts within CESP projects to support new and existing users of the Atlas platform. ALA may put some time into containerisation technology to make installations easier to run. This coupled with helpdesk and infrastructure support from GBIF could help newer nodes get up and running quickly.
- **Colombia:** "Ventanas departamentales / Regional windows" - Starting from a need for regionalization of SiB Colombia that has been latent for some time, in the course of 2019 two new instances to explore biodiversity at departmental level in the country will be published and put to the test. The first departments will be Boyacá and Santander. These windows are tools for exploring the data and information available about a specific department and sow the seed to develop a very useful tool for decision-making and territorial management at the regional level, through appropriation, awareness and empowerment.
- **Colombia:** Data repatriation: from September 2018, the SiB Colombia team will work in the data repatriation process to the national data portal (datos.biodiversidad.co). There is now almost 3 million occurrences published in GBIF.org for non colombian organizations. We are looking also bring to the national portal the occurrences of national organizations that published data in IPT's from other countries (e.g. CIAT).
- **France:** GBIF France will begin development of a portal based on the ALA software for GBIF Togo, and will continue its actions on the Living Atlases Community.
- **Japan:** Continue revision of S-net system.
- **Korea:** Continue to update the coordinates for occurrences in Korea's published datasets
- **Korea:** Promote mutual cooperation between the KBIF committee members and the affiliated institutions.
- **South Africa:** SANBI-GBIF will consider mechanisms to engage the scientific community around the new and evolving concept of the GBIF Open Science Ambassador Concept. Here the Node will take forward the initial advocacy actions with the marine science community, conducted at the South African Marine Science Symposium. SANBI-GBIF has highlighted research aimed at developing the field of biodiversity informatics in South Africa

through the use of primary data (big data –more than 2 million records) and informatics tools (niche models) to address key biodiversity challenges (food security and climate change). This demonstrates the use and relevance of GBIF mobilized data across the value chain. Showcase examples for Southern Africa will be derived.

Activity 1d - Equip data publishers

Rationale

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

Approach

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

Tasks

1.d.i   Promote and support capacity self-assessment for data holders

1.d.ii   Promote publication of collection metadata

1.d.iii   Simplify data publication pathways (spreadsheet-level publishing)

1.d.iv   Manage IPT feature upgrades

1.d.v   Operate hosted IPT infrastructure

1.d.vi   Consolidate NSG-led endorsement process

1.d.vii   Develop online collaboration through GBIF.org helpdesk to assist and mentor data publishers

1.d.viii Provide clear online reporting of the use of data for data publishers

1.d.ix   Promote data management plans as key tool for data publishers (Added 2018)

Progress

The Integrated Publishing Toolkit (IPT) is being maintained to deal with bug fixes, translations etc., and version 2.3.6 was released during 2018. Review of requirements for a revised IPT is on hold until there is greater clarity on how broad a data model for GBIF should be (see Activity 2a). With a new model in place, GBIF will design appropriate tools to map and publish data.

The cloud-based IPT infrastructure maintained by the Secretariat continues to provide a fallback solution for publishers unable to host their own installation or to find a third-party hosting option. As of September 2018, this option was being used by six publishers associated with the BID programme, sharing a total of 52 datasets. Cloud-hosted IPTs are expected to become more widely used in coming months, especially with the plan to use volunteer mentors to provide helpdesk support for users of this option (see 2019 work items in this section).

The Secretariat in collaboration with the Nodes Steering Group continued in 2018 to identify and attempt to resolve remaining log-jams in the process of endorsing new publisher wishing to share

data through the GBIF network. To address isolated cases in which publishers experience extended delays in response to an endorsement request from a participant node, the NSG agreed to amend the [endorsement guidelines](#) to enable community-level endorsement as a last resort if the node remains unresponsive. Further review of these guidelines is planned for the final quarter of 2018.

Participant contributions

Australia: The Atlas delivered enhancements to its platform for registering datasets in GBIF using GBIF's registry APIs. These enhancements are in use by both UK and Australia. In addition, we continue to support thematic portals within Australia for Herbaria and Zoological collections and other groups aimed at showcasing the work of data publishers.

Colombia: Increase of biodiversity data publications through SiB Colombia node. In the current year the SiB Colombia's network increased with 16 new data publishers (14.5%) estimating an increase of almost 20% to the end of the year. The publications at the current year represent an increasement of ~ 124,000 new occurrence, 52 occurrence datasets, 10 Checklist and 5 Events.

France: GBIF France hosts data and maintains IPT instances for 13 southern countries, and web sites for Madagascar and Benin.

Japan: IPT established to publish previously unpublished data.

Korea: Published ca. 1 million collections data from National Institute of Biological Resources (NIBR) of Korea via GBIF portal.

Korea: Seoul National University received a fund from the BIFA project (Collections Data Mobilization Grant).

Norway: GBIF Norway has developed an online tool for generating GBIF-compatible spreadsheet-templates with source code available on GitHub for further collaborative development.

South Africa: 1. SANBI-GBIF provides helpdesk support for Nodes in Africa to host data, through SANBI-GBIF's Africa IPT. 2. SANBI-GBIF hosts an Africa IPT infrastructure to support dataset publishing by Participants in Africa that may require this service. This service has supported some existing GBIF country Participants publishing data as an outcome of BID, as well as some new countries which are not yet GBIF Participants. GBIF Ghana, Liberia, Zambia

2019 Work Items

- Develop helpdesk services for new data publishers using cloud-hosted IPT installations, managed by volunteer mentors experienced in data publishing

Participant plans for 2019

- **Australia:** Data audits with a focus on trait information, and business analysis to assist the mobilisation of trait information.
- **France:** GBIF France will continue to support data publishing for southern countries.
- **Japan:** New IPT server to be established.
- **Korea:** Have a plan to publish new historical herbarium specimen data (nearly 140,000 specimens) from the Korean peninsula (including North Korea) flora checklist provided by T.B. Lee herbarium in Seoul National University (funded by BIFA collections data mobilization grant). Plan to provide new dataset of relatively recent specimens of woody plants in Korea (ca. 100,000 occurrence records with georeference & data publications)
- **Norway:** GBIF Norway will continue to offer data hosting and publishing services (IPT) also for data publishers located outside of Norway.
- **Norway:** GBIF Norway will continue to offer online services for citizen science digitization of museum specimens (see activity 4c).

Activity 1e - Expand national participation

Rationale

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

Approach

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

Tasks

1.e.i   Ensure GBIF.org interface and key content supported in relevant languages

1.e.ii   Improve coordination and mechanisms for engaging new countries

1.e.iii   Provide Governing Board with annual review and progress assessment for GBIF status and engagement in all countries

Progress

Following comprehensive back-end preparation by the Secretariat informatics team during the first part of 2018, the communications team began a pilot phase with the network's Spanish and Japanese language communities to test workflows, tools and processes for translating GBIF.org.

Translation of the user interface (navigation menus, table headings, field labels etc.) relied on CrowdIn, a tool previously used in translating the IPT interface. By mid-August, the work of a team of three Japanese and ten Spanish volunteers had substantially completed translation of the UI, leading to the release of a language switcher that allows users to view GBIF.org in English (https://www.gbif.org), Japanese (https://www.gbif.org/ja) Spanish (https://www.gbif.org/es).

With the pilot phase well-advanced, volunteers from other language communities are translating the UI and key content following a call for volunteers at the end of August, with some including Portuguese making very rapid progress. Where volunteer effort proves insufficient, commercial translators will be contracted. The intention is to have the UI available in five UN languages (English, French, Russian, Spanish, Simplified Chinese), plus Portuguese and Japanese, by the end of 2018. Because of the additional technical adjustments needed to support right-to-left script, Arabic translation is deferred until 2019.

Modest growth in the number of national participants in GBIF continued in 2018, with South Sudan, Zimbabwe, Cameroon and Viet Nam joining as Associate Participant Countries. Liberia also formally communicated its intent to join as a Voting Participant, and several other African countries are working towards participation as part of activities under the BID programme. Discussions continue with a number of other potential participants in Asia, the Caribbean, Central America, North Africa

and the Former Soviet Union regions. The Secretariat continues to seek appropriate routes to build on successful institutional engagement and data publishing activity both in Russia and China to secure national government participation in each of these countries.

Participant contributions

Australia: ALA has facilitated conferences and workshops with many countries and will continue to support facilitation of Living Atlas infrastructure.

Colombia: "Colombia BIO | Nuevos canales de participación." SiB Colombia has updated two of its existing participation channels: Data Portal and Biodiversity Catalog, and has developed two completely new channels. The first one, "Online Collections", is a significant advance in the visibility and availability of the multimedia material associated with specimens in biological collections, and the second, "List of species", allows access to all the lists of species that are shared through SiB Colombia, including the reference lists of the country. Results: Catálogo de la Biodiversidad: http://catalogo.biodiversidad.co Portal de datos: http://datos.biodiversidad.co Colecciones en línea: http://colecciones.biodiversidad.co Listas de especies: http://listas.biodiversidad.co

Colombia: Internationalizing GBIF.org Colombian node is participating in the GBIF Portal Spanish translation in the current year, using the Crowdin platform arranged by GBIF https://crowdin.com/project/gbif-portal. At his moment there is an 83% advance in the Spanish translation with help of Spanish speakers nodes (Spain & Argentina mainly). The results are displayed in https://www.gbif.org/es and the translation will be complete in the second semester of 2018.

France: Invitation of the project leader of a Balkan project and the LifeWatch CEO to the EU nodes meeting, Participation in other meetings: RDA Berlin, Irkutsk in September.

France: Help for French translation.

Japan: Translation of GBIF.org

Norway: Capacity enhancement and data mobilization training(BioDATA, during 2018-2021) by GBIF Norway, the Norwegian Research School in Biosystematics (ForBio), GBIFS, and funded by the Norwegian Centre for cooperation in Education (SIU), will approach potential new GBIF member countries and promote the benefits of GBIF membership in Russia and former Soviet Union republics. (See also activity 1b).

South Africa: SANBI-GBIF, through the JRS funded ABC project is engaged in expanding the networks through hosting Biodiversity Information Management Forums (BIMFs) in 4 countries (Rwanda, Ghana, Namibia, Malawi) in Africa, replicating the South African BIMF model for developing the Biodiversity Informatics community.

South Africa: GBIF-BID/ABC workshop on data use and application held in February 2018, at Kirstenbosch Gardens, Cape Town, SANBI.

South Africa: SANBI-GBIF will provide support in expanding engagement with new countries in Africa. A service provider will be secured to assist with this work

2019 Work Items

- Add support for Arabic language content in GBIF.org, including technical adjustments to accommodate right-to-left script, and commissioning translations of the user interface as well as key content
- Use BID closing regional meetings in Caribbean, Pacific and sub-Saharan Africa to promote further national participation in these regions, both for countries with funded projects and for additional countries invited from each region.
- Build on current funded projects, and develop additional funding proposals, to implement an

intensive engagement strategy for remaining regions with under-represented or no national participation, namely: 1) the Former Soviet Union including Central Asia, the Caucasus and the Baltic States; 2) the Balkans and other Central and East European countries; 3) North Africa and the Middle East; and 4) China, South and Southeast Asia.

Participant plans for 2019

- **Japan:** Translation to be continued for other materials.
- **South Africa:** The ABC project follows an incentivised approach for data mobilisation. As a final outcome of the project, a prize-giving ceremony will be conducted in February 2019 based on a number of criteria. This approach will demonstrating the use and application of data which has been mobilised, for aspects of societal need and/or policy relevance.
- **South Africa:** Through SANBI-GBIF, new countries in Africa will be identified to further the biodiversity informatics agenda in support of science and sustainable development. National networks will be developed in the region through the implementation of further BIMFs and GBIF advocacy will be conducted.

Activity 1f - Plan implementation

Rationale

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

Approach

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

Tasks

1.f.i   Develop better integration between nodes activities and Governing Board processes

1.f.ii   Expand regional meetings to serve as GBIF community Implementation Planning conferences and opportunities to develop fundable programmes of work

1.f.iii Increase alignment between GBIF work programmes and activities of each GBIF Organisational Participant and Affiliate

1.f.iv   Explore increased alignment and joint organisation of workshops with other networks such as TDWG, SPNHC, GEO BON, RDA, etc.

Progress

Regional meetings are being held in all GBIF regions in 2018, strengthening regional engagement and supporting planning and outreach at the regional level. The NSG identified the key topics for discussion across all regions. The results of the meetings will be incorporated into the Nodes Committee Chair's report to GB25.

The 2nd Global Biodiversity Informatics Conference (GBIC2) was held in Copenhagen during July 2018, with more than 100 attendees from all global regions. The workshop received external sponsorship from Agence française pour la biodiversité, the Atlas of Living Australia, DiSSCo, the Field Museum, iDigBio, the JRS Foundation, MCZ Harvard, NLBIF, Pensoft and UN Environment, and GBIF was granted free access to Copenhagen University meeting areas. GBIC2 followed from the first GBIC in 2012, which delivered the Global Biodiversity Informatics Outlook (GBIO). The GBIO provided a vision and framework for "delivering biodiversity knowledge in the Information Age" through modular delivery of biodiversity information infrastructure. Progress since 2012 has been uneven, primarily because of the complexity and diversity within the landscape of stakeholders in biodiversity knowledge, with overlapping missions and weakly aligned funding. GBIC2 attendees proposed the establishment of a lightweight international alliance to support open planning and collaborative delivery of the tools, services and infrastructure necessary to organise all biodiversity data. The GBIF Secretariat was tasked with leading on development of a paper outlining this proposal, and on seeking to establish such an alliance. Relevant models for the alliance can be seen in the Apache Software Foundation (collaborative open source development), ELIXIR (coordination of bioinformatics data services), and the Global Alliance for Genomics and Health (a closely parallel effort to facilitate community delivery of scalable global solutions).

Participant contributions

Australia: The ALA has been working with GBIF towards having a shared code base for the

processing of occurrence data.

France: As EU representative, organied the EU nodes meeting with Estonia in Tallinn.

South Africa: SANBI-GBIF Node Manager and SANBI executive has met with GBIF Head of Delegation of South Africa, with regards to the Africa Coordinating Mechanism Business Case. Further engagement will continue to look at Biodiversity Informatics regional engagement in Africa, in support of sustainable development.

South Africa: SANBI finalised the National Biodiversity Information System (NBIS) design protocol (Phase 1) in February 2018. The NBIS will provide enhanced access to integrated South African biodiversity information through a centralised web-based infrastructure. It will support and facilitate access to collections data, observations, images, species pages, names data, spatial layers, genetics data.

2019 Work Items

- Develop standardized, annual information requests from Participant nodes to capture relevant updates and work plans based on the GBIF Strategic Plan and Implementation Plan structure, enabling greater visibility for node activities on GBIF.org and supplementing the country-level metrics provided in the automatically-generated Activity Reports.
- Follow up from GBIC2 workshop by leading in communications, consultations and working groups to develop a global alliance for biodiversity knowledge. The EU-funded Synthesys+ project will provide some funding in 2019 and subsequent years for GBIF staff time and workshops for this purpose. €50,000 has also been allocated in the GBIF 2019 budget to support additional workshops, meetings or communications materials.

Participant plans for 2019

- **Australia:** Start implementation and consultation plan. Contribute staff time to help support development activities, hackathons, sprints, etc. that could be made widely available.
- **France:** Organization of the EU nodes meeting and election of the new EU representative
- **Norway:** The 2019 European regional Nodes meeting is preliminarily planned to take place in Poland. GBIF Norway has offered to be available as an alternative host for the regional European Nodes meeting in 2019, or to host the European Nodes meeting a later year.
- **South Africa:** Strategic engagements/meetings between SANBI-GBIF Node Manager and South African Head of Delegation will continue in order to evolve the South African Node planning, and Africa portfolio of work and to elaborate the Science Diplomacy role SANBI-GBIF can play.
- **South Africa:** Phase 2 (of national data platform) commenced in October 2018 (24 months), which looks at the implementation phase of the NBIS.

Activity 1g - Coordinate resources

Rationale

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

Approach

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

Tasks

1.g.i   Secretariat to coordinate participant-led activities aligned with Work Programme (process and models to support participant commitments)

1.g.ii   Development of concept notes/proposals for supplementary funding at regional or global scales

1.g.iii   Operate annual Capacity Enhancement Support Programme (CESP) with core funds, aligned with supplementary investments

1.g.iv   Ensure that BID programme and other supplementary funds align with and reinforce CESP and GBIF community, online resources and infrastructure

1.g.v   Align Ebbe Nielsen Challenge and Young Researcher Awards with GBIF priorities

Progress

The BID programme reached its most intensive year of implementation in 2018 with management of 63 projects in sub-Saharan Africa, Caribbean and the Pacific and capacity enhancement workshops in all three regions focused on data use for decision making (see Activity 1b). A full 2018 progress report on BID, including its contribution to wider GBIF activities in all regions, and a process for consulting on a possible second phase of BID once the current programme completes its implementation in 2019, is provided as a separate document to the Governing Board at its GB25 meeting.

The Biodiversity Information Fund for Asia (BIFA) attracted a strong response in its latest call for proposals for data mobilization projects. After a rigorous selection process, eight projects were selected for funding in Cambodia, Indonesia, Korea, Nepal, Pakistan, Philippines and Vietnam. This year's call replicated the model developed under BID in providing a common training workshop for all funded projects to acquire basic data publishing skills to support project implementation. The

workshop was hosted in Beijing by the Chinese Academy of Sciences and was supported by trainers and mentors from several GBIF nodes.

A fifth call for the Capacity Enhancement Support Programme (CESP) was launched in 2018, resulting in selection of six projects for funding, managed by nodes in Europe, North America, Africa and Latin America. As requested by the Governing Board at its GB24 meeting in 2017, the Secretariat carried out a five-year review of the CESP programme involving a survey of nodes and analysis of outcomes of the funded projects, as well as the procedures for running the programme. The results of the review including a set of recommendations is presented to GB25, with an overriding recommendation that the programme should be continued with some adjustments reflecting the feedback received during the review.

Annual calls in 2018 under the The Young Researchers Award and Ebbe Nielsen Challenge programmes attracted significantly better response than in 2017. The results will be announced during the GB25 meeting.

Participant contributions

Colombia: Contributed to the assessment and evaluation processes for BID funding calls. In 2017 from Colombian node we collaborate in the assessment of the "BID Africa 2" concept notes. We read and evaluated 40 concept notes supporting the advance of the African continent in the mobilization and publication of biodiversity data through GBIF.

France: Contribution to 4 CESP: -International Living Atlases workshop-Francophone mentoring and translation of the open-source ALA platform-Mentoring between GBIF Benin and GBIF France-EuropeanBireme: EU Nodes in biodiversity reporting mechanisms

South Africa: SANBI-GBIF has led the development of the GBIF-Africa regional engagement strategy and business case funding proposal - Towards the establishment of the Africa coordinating mechanism 2018-2023. This has been presented to the Department of Science and Technology, with in principal endorsement for the content and strategy.

South Africa: Funding budgeted through FBIP programme for data mobilisation through large and small grants ± € 645 000;

South Africa: The ABC project, through JRS funding provides 80 000 US$ in prize money for data mobilization.

2019 Work Items

- Subject to approval by GB25 of recommendations made in the five-year review of the Capacity Enhancement Support Programme (CESP), launch a 2019 call for proposals under a renewed CESP, taking account of the feedback received during the review. This call specifically to exclude support for core activities under the Living Atlas programme, to be funded and coordinated separately (see Activity 1c).
- Complete implementation of the current BID programme, including a) closing all remaining projects and delivering necessary documentation to the EU; b) commissioning impact report to analyse the return on investment including mobilized data, capacity enhancement and additional policy support in the target regions; c) publishing guiding examples and best practices developed through BID on mobilization and use of data for priority policy needs; d) organizing closing meetings in each target region to showcase outcomes and develop sustainability plans including through increased GBIF participation (see Activity 1e); and e) organizing a BID closing event, probably in association with GB26, to promote the outcomes to the wider community and propose next steps.
- Based on the outcomes of the community consultation on the BID programme, develop

proposals for a second phase of BID in collaboration with the European Union and the BID Steering Committee

- Implement a fourth call for proposals under the Biodiversity Information Fund for Asia (BIFA), reflecting priorities agreed by Asian nodes and the BIFA Steering Committee, and conclude discussions with the Government of Japan to decide the future of this supplementary fund beyond 2019
- Work with all GBIF regions to identify and target potential funding streams to support additional capacity enhancement for data mobilization and use, building on the BID and BIFA models
- Roll out selected grant management system to streamline project calls, assessment, selection, implementation, budgeting and reporting, with continuous refinement and evaluation of the system to ensure it is meeting the needs of the community.
- Launch 2019 calls for the Ebbe Nielsen Challenge and Young Researchers Awards, taking into account recommendations from the Science Committee based on the 2018 programmes.

Participant contributions

- **Australia:** Contribute staff time to help support development activities, hackathons, sprints, etc. that could be made widely available.
- **South Africa:** Further engagement with regards to funding allocation will continue. The Africa Coordination Mechanism (ACM) 5 year strategy is seen as a catalytic investment. Sustainability will be built into the 1st phase, so there will be less concern about what happens subsequently, as this would have been built into the model. The ACM Business Case and Regional Engagement Strategy is a model that could be adopted in other regions of the world.
- **South Africa:** In the 1st three years, through the ACM, GBIF-Africa, through SANBI-GBIF will explore the need and feasibility of developing a bigger platform that can provide for a strengthening of a community of practice for all African biodiversity informatics initiatives, to enhance and make more efficient use of the data. This will be done through leveraging additional regional and global partnerships and initiatives.
- **South Africa:** Data mobilisation funding through FBIP large and small grants ± € 645 000 will continue in 2019. 4. ABC Prize money to be awarded in 2019

Priority 2 – Enhance Biodiversity Information Infrastructure

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

Activity 2a - Modernize data standards

Rationale

The GBIF network participants are able to reliably exchange data thanks to their adherence to a set of standards. As GBIF looks to grow in capability, enable exchange of richer content and improve the quality of data, the standards must be revised and evolve accordingly.

Current standards adopted by GBIF are not yet adequate to accommodate the needs expressed by many potential and existing data publishers. Weaknesses in the model have led to ambiguous or over-complex data representations and unclear documentation, leading to difficulties in data integration and use. The main issues relate to uncertainties around the use of Darwin Core record types, the basisOfRecord element, and the use of Core and Extension vocabularies. Reviewing and updating the core domain model, tightening up the vocabularies and documentation and adopting more robust exchange standards will result in an easier to use, and a wider reaching GBIF data exchange network.

Approach

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

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

Tasks

2.a.i   Promote development of a shared domain model for sharing and linking all components of biodiversity information

2.a.ii   Lead a review of the Darwin Core vocabulary and associated extensions to ensure consistency and full alignment with a shared domain model

2.a.iii Explore opportunities to increase accessibility of biodiversity data through evolution of Darwin Core Archive formats to W3C CSV on the Web formats

2.a.iv   Explore models to enable GBIF and other biodiversity infrastructures to deliver comprehensive global catalogues of instances of key data classes

2.a.v   Improve management of trait data of relevance to GBIF

Progress

TDWG does not yet offer a set of recommendations that can be implemented as an approved standard for a common domain model, although many of the necessary elements have been developed. GBIF is well positioned to move this work forward by offering a new baseline implementation within GBIF.org.

The GBIF Secretariat has recognised that are many of the most serious challenges facing data users arise from the fragmented approach to data integration and data quality that has arisen as multiple infrastructures aggregate overlapping subsets of the same occurrence data. Each data aggregation infrastructure carries out its own harvesting, interpretation, assignment of unique identifiers, normalisation, quality checks, faceting, etc. and offers its own downloads, API and data citation tools. The results of each of these steps are different in each network/infrastructure.

GBIF, OBIS, ALA (and the Living Atlases), iDigBio, VertNet and GGBN are significant existing networks that seek to address these needs in parallel. The GBIF Secretariat has started discussions with these parties to explore the extent to which they all could combine forces and move towards jointly aggregating data as a shared web infrastructure, producing and maintaining a single shared data product which is replicated globally and supports whatever views and access choices each partner/portal requires. This conceptual model could address the following needs:

- Every occurrence record should have a globally unique resolvable identifier for use wherever it is accessed. These identifiers could serve as a robust, sustainable basis for persistent reference (as with the International Geo Sample Number, IGSN).
- Every aggregated record should include the verbatim record, the standard DwC attributes derived from the verbatim form, and an endlessly extensible set of other attributes, including e.g. VertNet interpretation of measurements, OBIS interpretation of associated marine measurements, ALA national taxonomy and spatially-aligned properties, etc. All of the benefits from all of this processing could be accessible to all users regardless of how and where they access the data.
- Data quality checks and associated processing should be fully standardised and documented.
- Data processing should be extensible to allow additional processing and additional attributes based on taxonomy or geography or e.g. for paleo-records.
- Search and download APIs could be standardised to simplify development of data access tools.
- All infrastructures could adopt a single mechanism for data downloads, DOI generation, and citation tracking. This would make it easier for data publishers to get integrated information on the use of their data through all outlets

Nothing would stop existing and future portals offering specialised value on top of these core services.

Initial discussions with other platforms have been positive and there is agreement, particularly with ALA and iDigBio to hold a technical workshop to develop a shared plan, including immediate steps to converge practices and a longer-term set of steps to explore closer unification.

Participant contributions

Australia: The Atlas has contributed to the work of the TDWG Biodiversity Data Quality Interest Group. The Atlas continues to collaborate with the CSA, ECSA, the Australian Citizen Science Association (ACSA) on data standards for Citizen science projects. The Atlas is providing tool support (publishing and aggregation) for these standards through the BioCollect platform.

BHL: Planning BHL next version. Launched full-text search of BHL. BHL participated in GBIC.

Norway: GBIF Norway described for the 2017-2021 implementation plan contribution to the TDWG process to advance biodiversity data standards.

Norway: GBIF Norway contributed as [review manager](#) for the ratification process of the [Vocabulary Maintenance Standard](#)(VMS) and the [Standards Documentation Standard](#)(SDS).

Norway: GBIF Norway is member of the [Darwin Core hour](#) team (organized by iDigBio) and contributes to TDWG interest groups and task groups on topics including persistent identifiers. GBIF Norway contributes to the development of the Event Core model (see also activity 3b).

Japan: GBIF Japan is in discussion with J-OBIS (the OBIS node for Japan) to avoid duplication in data provision from Japan.

2019 Work Items

- Work in open consultation with the expert community to document a unified information model that covers the scope of content GBIF accommodates and supports richer publishing of interconnected information from external databases. This model should be presentable in various languages. €50,000 is reserved in the budget to support two associated workshops.
- Provide a set of data exchange profiles for sharing data within GBIF in conformance with the unified information model.. This should include both existing and new standards and the necessary controlled vocabularies (e.g. sampling protocols).
- Redesign the GBIF Integrated Publishing Toolkit to support these profiles and to address infrastructural needs (ability to install locally, or use a GBIF hosted solution).
- Provide documentation for the data model and for the associated services offered through GBIF.org.
- Review and redesign GBIF data management to accommodate the unified information model as part of data ingestion, quality control and processing necessary.
- Continue technical discussions with other data aggregators to seek closer alignment in practice and, as far as possible, implementation of aggregation and indexing processes.

Participant plans for 2019:

- **Australia:** Work with GBIF on the reference implementation of the data quality tests. Continued engagement with CSA, ECSA and ACSA. The ALA recommends promotion of ALA BioCollect tool as a method to engage disparate groups.
- **BHL:** Establishing new metadata model for BHL
- **Norway:** GBIF Norway will continue contribution to the TDWG process including topics on Darwin Core documentation, persistent identifiers, collection descriptions, common domain model, data exchange models and biodiversity informatics curriculum. GBIF Norway will participate and contribute to the TDWG 2019 annual conference in Leiden.
- **Norway:** The wider Norwegian GBIF community will continue implementation of the sampling event data model for environmental monitoring and survey-based data with focus on national implementation while contributing to the international standardization process (see also activity 3b).

Activity 2b - Deliver names infrastructure

Rationale

The most significant challenge to improving the quality of aggregated occurrence data is the continuing need for a comprehensive checklist of known species, and even for a comprehensive list of published scientific names. Interpreting and mapping names depends on the quality and completeness of these resources. Even in cases where names in occurrence records are incorrect or misspelled, better names infrastructure can assist by increasing confidence that fuzzy match algorithms or human intervention is required.

Delivering these resources is the focus of a number of GBIF Participants and other stakeholders, including the Catalogue of Life partnership, WoRMS, nomenclators (IPNI, Index Fungorum, ZooBank) and many national, regional or taxonomic databases. A comprehensive resource for scientific names and taxon concepts organized at least as a workable reference classification (but with support for additional classifications as appropriate) would also benefit other infrastructures, including Encyclopedia of Life, Biodiversity Heritage Library, Barcode of Life and GBIF nodes, and improve interoperability between data from these infrastructures. It would also be beneficial to accommodate vernacular names, informal names for undescribed species and other identifiers such as Barcode Index Numbers.

Approach

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

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

Tasks

2.b.i   Partner with other biodiversity informatics initiatives and taxonomic database holders to plan and deliver a comprehensive nomenclatural dataset and working consensus classification for all life

2.b.ii   Promote publication of species checklists through GBIF network

2.b.iii   Explore potential use of checklists to assist with data validation or derive augmented data products

2.b.iv   Explore integration of Linnaean nomenclature of formally described taxa with provisional names and species hypotheses and OTU naming

Progress

The goal of the CoL+ project is to innovate the CoL systems by developing a new information technology infrastructure that includes both the current Catalogue of Life and a provisional Catalogue of Life (replacing the current GBIF backbone taxonomy), separates scientific names and taxonomic concepts with associated unique identifiers, and provides some (infrastructural) support for taxonomic and nomenclatural content authorities to finish their work.

The initial steering committee, consisting of the GBIF Secretariat, Species 2000/Catalogue of Life, and Naturalis Biodiversity Center, has been expanded with representatives from the Barcode of Life data systems, Biodiversity Heritage Library, Encyclopedia of Life, and ITIS. The CoL + project was presented to the GBIF community at the 14th global Nodes meeting and the 24th GBIF Governing Board, at the 42nd and 43rd meeting of the Consortium of European Taxonomic Facilities, and at the 2018 SPNHC/TDWG. The CoL+ project is part of the strategic alignment group of the Distributed Systems of Scientific Collections (DiSSCo).

Conversations have taken place about future cooperation with the following parties:

- Kew / International Plant Names Index & Index Fungorum for connection with nomenclature information
- Species File group - Illinois connection with taxonomic editing tool TaxonWorks
- LifeWatch / WoRMS about taxonomic editing tool Aphia
- World Flora Online about taxonomic plant information and editing tools
- European Environment Agency

These partnerships will be further elaborated.

GBIF has allocated funds for workshops to advance development of a common checklist for all vascular plants meeting the needs of CoL, WFO, The Plant List and others and to develop an international editorial model for updating the Lepidoptera component of CoL.

The Catalogue of Life Plus project has so far delivered the following milestones:

- Dataset store and dataset importer for the clearinghouse infrastructure. This includes a back-end API installation and documentation: http://api.col.plus. A demonstration environment is currently hosted at the GBIF Secretariat. The datastore includes data sources coming from the global species databases (CoL GSDs) in ACEF format. It also includes data sources coming from the GBIF Backbone Taxonomy in DWC-A format.
- API debugger: http://tools.col.plus/. This tool enables search for species names and datasets.
- Names index for handling unique scientific names: described here.
- Work has started on building the tools needed to assemble the current Catalogue of Life.

Participant contributions

Australia: The Atlas has continued to support the development of names infrastructure within Australia. The outputs of this infrastructure will provide a source of scientific names and taxon concepts from Australia into international clearing house initiatives.

Colombia: "Colombia BIO | Listas de especies de referencia para el país." Branch of the project that achieved the consolidation of the first two lists of reference species for Colombia: the list of mammals and the Checklist of the freshwater fishes of Colombia. This unprecedented achievement in the country was possible thanks to a collaborative construction of knowledge among networks of experts, universities and biological collections, among others.

Japan: Check list of Fungi in Japan digitization almost completed and to be released. Vascular plant check list "Green List" http://www.rdplants.org/gl/ released.

Switzerland: Publication of national species checklists for red list groups and important invertebrate groups.

2019 Work Items

- Replace the existing processes for constructing the monthly and annual Catalogue of Life checklists and for constructing the GBIF taxonomic backbone with a single solution that delivers both products (together forming a "provisional checklist"). This work is carried out under the NLBIF-funded Catalogue of Life Plus (CoL+) project. €80,000 has been allocated in the budget to support GBIF development costs, with €20,000 in matching funds from the NLBIF funding to CoL+.
- Plan for existing GBIF checklistbank capabilities (registry integration, images, descriptions, APIs) are correctly handled within the CoL+ implementation.
- Ensure that updates to the provisional checklist are reflected directly in the GBIF data index.
- Manage the CoL+ software as part of ongoing infrastructure supported at GBIF.
- Work with Catalogue of Life and other partners to develop responsive expert communities to support sectors which currently lack adequate support and to improve alignment with other checklist efforts including World Flora Online and WoRMS.

Participant plans for 2019

- **Australia:** Delivery of names and taxon concepts to Catalogue of Life plus.
- **Japan:** Additional data and revision to be continued. Endangered species (National and local in Japan) checking application to be developed.

Activity 2c - Catalogue collections

Rationale

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

Approach

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

Tasks

2.c.i   Partner with other biodiversity informatics initiatives to deliver a single, comprehensive catalogue of the world's natural history collections

2.c.ii   Use collection metadata as a first stage in content mobilisation from natural history collections, including identification or development and adoption of necessary data standards

Progress

During 2018 the website for the Global Register of Biodiversity Collections (GRBio) and Global Register of Scientific Collections (GRSciColl) was hacked and is now offline.

GBIF have committed to enhancing the GBIF Registry during 2018 to support representation of collections and institutions as registered entities (in addition to data publisher entities), and to migrate the content from GrBio and GrSciColl, surfacing it through the GBIF registry API. GBIF will then proceed to develop an interface for editors to curate the content.

Within TDWG, work has continued to review and update the Natural Collections Description (NCD) standard. Once this update is complete, GBIF plans to support the inclusion of NCD documents as rich content to attach to GRBio records, and to work with nodes and collection networks to enable publication and update of NCD records through IPT or other channels.

Participant contributions

Switzerland: Acquisition of collection metadata in order to support the definition of priorities for the scientific revision and data digitization of important collection holdings (national and international level) as a joint initiative of GBIF Switzerland, the Swiss Academy of Sciences, the Swiss Association of Natural History Collections and the Swiss Systematics Society.

2019 Work Items

- Deliver a web version of the GRBio/GRSciColl dataset as part of ongoing infrastructure

supported at GBIF.
- Work with GRSciColl, Index Herbariorum, GBIF nodes and others to establish sustainable processes for updating the registry.
- As the TDWG Natural Collections Descriptions Interest Group revises the NCD standard, enhance the registry to accept NCD documents and will develop mechanisms via IPT or otherwise for collections and nodes to publish and update these data.

Participant plans for 2019:

- **Norway:** GBIF Norway will contribute to the TDWG standards process on collection descriptions (CD) (see also activity 2a) and take a leading role nationally in Norway for collecting and publishing metadata on Norwegian natural history collections in GBIF when the global GBIF infrastructure facilitates this data type.

Priority 3 – Fill Data Gaps

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

Activity 3a - Identify priority gaps

Rationale

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

Approach

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

Tasks

3.a.i   Organize and integrate requirements identified by fitness-for-use groups

3.a.ii   Establish open mechanism for researchers and users to document specific data needs

3.a.iii Develop assessments and visualisations of strengths and weaknesses of GBIF data in key dimensions (taxonomic, geographic, environmental, temporal)

3.a.iv   Maintain well-publicized GBIF priorities for data publishing, incorporating needs of IPBES and other networks

3.a.v   Provide Governing Board with annual review and progress assessment for digitisation of major natural history collections

3.a.vi   Provide GBIF Participants with actionable priorities and targets for content mobilization

Progress

The appointment of a Data Analyst in the second quarter of 2018 has enabled work to commence on establishing ways of measuring and visualizing data gaps within GBIF's services. The focus has been on occurrence count representation per taxonomic grouping over geographic grid cells, with various resolutions to evaluate coverage vs. sampling efforts, and to determine confidence levels for the completeness of a given cell. Work will continue to establish a baseline.

The Secretariat has continued to work with the IPBES Task Force on Data and Knowledge to help prioritize data mobilization based on knowledge gaps encountered through the IPBES assessment process. An ongoing analysis of knowledge gaps reported from the IPBES pollinators assessment is likely to be the first deployment of this approach, with early proposals likely to emerge towards the end of 2018 or in 2019.

Participant contributions

Colombia: Colombia BIO. Since 2016 to 2018 The Administrative Department of Science, Technology and Innovation of Colombia - Colciencias organized the project Colombia BIO to promote knowledge, conservation, management and sustainable use of biodiversity in Colombian territories. With the joint work of several research institutes and organizations this project made a large amount of data available through the SiB Colombia in regions of the country previously difficult to access due to problems of armed conflict, in this way it has been possible to fill data gaps in Colombia through this national initiative. Colombia BIO Datasets: https://www.gbif.org/dataset/search?q=Colombia%20BIO

Japan: As a part of research program in the National Museum of Nature and Science, data gap analysis for the endangered species was adopted. The research will be carried out in three years.

Norway: GBIF Norway contributed during 2017 and 2018 to follow-up activities based on priorities identified by the 2016 Task group on GBIF fitness for use in agrobiodiversity that was established by GBIF and Bioversity International. GBIF Norway have together with the appropriate data owners completed data publication of international datasets to address key gaps identified by the task group (including international and Nordic crop wild relative occurrence and checklist datasets) (see also activity 3a, 4a, and 5b).

2019 Work Items

- Prioritize data gaps through a 'hunger map' approach, including systematic documenting of user-communicated data content needs and matching them with available data; and extending the analysis of data searches, including no-search results, to help understand categories of data for which GBIF users do not currently find what they are looking for.
- In collaboration with IPBES and others, capture priority knowledge gaps for which lack of accessible primary biodiversity data is identified as a significant cause of uncertainty
- Based on the gap analysis approaches mentioned above, in combination with earlier efforts such as the 2017 data mobilization priorities , develop actionable guidance for nodes, publishers and funders on addressing priority gaps through mobilization targets and strategies based on spatial, temporal, taxonomic and thematic dimensions of biodiversity data.

Participant plans for 2019

- **Australia:** Recommend that the community works collaboratively to define targets around coverage and extent of our data products (e.g. temporal/ spatial and taxa). This will help define the level of effort required for data publishers to met these gaps. Also define level of use from Occurrence to Abundance to True Absence.
- **Japan:** Research program on gap analysis for endangered species to be continued.
- **Norway:** GBIF Norway will continue to represent and promote implementation of GBIF infrastructure technology and solutions in the international agrobiodiversity community with focus on Nordic and European projects for conservation and utilization of crop wild relatives with secured Nordic project funding (see also activity 3a, 4a, and 5b).

Activity 3b - Expand data streams

Rationale

GBIF serves as an integration point for any source of evidence of the recorded occurrence of species in time and space. A primary role for the GBIF infrastructure is to serve as a comprehensive single point of access for discovery, access, use and curation of all such evidence. Several classes of data are already well-supported within the GBIF network.

These include collections data, observations from field research, and many categories of citizen science data. However, there are other new and developing streams of data which should be accommodated if GBIF is to serve as the platform for supporting comprehensive data assessment and modelling (e.g. for GEO BON Essential Biodiversity Variables, IPBES assessments, Red List assessments, etc.). These include efforts to mine historical data records from literature, genomics activities and particularly barcode-driven surveys, and potentially species-level data from remote-sensing systems. More work is also still needed to engage with the full spread of research activities delivering sampling event data of various kinds. GBIF needs to ensure that it provides simple, effective and beneficial ways for researchers to share these and other streams of Darwin Core compatible data.

Approach

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

Tasks

3.b.i   Promote use of sampling event data model for ecological and monitoring datasets

3.b.ii   Partner with BHL and others to support integration of species occurrence records based on literature

3.b.iii   Work with bioinformatics initiatives and databases to form robust bidirectional linkages with molecular data

3.b.iv   Explore opportunities to integrate species-level data from remote sensing

Progress

The first steps toward integrating environmental DNA sequencing data directly into the GBIF occurrence index were taken during 2018, in a pilot collaboration with the UNITE community of Northern European researchers. This database has defined more than 73,000 species hypotheses based on clustering of DNA sequences taken from fungi. These are represented as stable operational taxonomic units (OTUs), linked to a Digital Object Identifier (DOI) and matched to the nearest known taxon name. By publishing these OTUs as a checklist in GBIF, they are now incorporated into the taxonomic backbone. A dataset based on DNA sampling of Danish fungi has already used this development successfully to publish georeferenced occurrences in which only the OTU is provided. This paves the way for publication of many more categories of data from sources such as eDNA in which established OTUs are recognisable but lack a formal scientific name.

The process of incorporating more data using the richer sampling-event format also advanced during 2018, as the informatics team enhanced the representation of such data through listing individual sampling events on the landing page for each dataset, and providing a page for each event listing the number of occurrences of each species contained in the sample, as well as a visualization of the taxonomic distribution. This is the first step in improved visualization of sampling-event data which will eventually help users to link datasets using similar protocols and to represent trends over time, making GBIF a more appealing destination for ecological monitoring data.

Participant contributions

Australia: The Atlas ran a national level workshop held in Canberra, which included EOL.The Atlas will also run a workshop focussing on traits at TDWG 2018. The goal of this is to attract a comprehensive range of people with expertise in trait data to discuss how we can form a community around trait data and better interact with each other. The Atlas has been working with genomics initiatives to incorporate Genomics data and eDNA data into Atlas infrastructure

BHL: Adding already transcribed handwritten items to BHL; full text search will enable better finding of occurrence data.

Colombia: AmeriGEOSS Week 2018 In August 2018 the Node Manager of Colombia and Nodes regional representative deputy Latin America LA & the Caribbean has the opportunity to assist to AmeriGEOSS Week: a meeting of organizations and experts in satelital images and remote sensing for the Americas that working, as GBIF, for the ecosystems and biodiversity conservation but at large scale. We found a very good opportunity 1. Work in linking effectively GBIF data with AmeriGEOSS datahub, cause that is not happening right now (https://data.amerigeoss.org/organization/gbif) and 2. Work hand in hand with some experts from the AmeriGEOSS community on the visibility of changes over time in event data published through GBIF.

Norway: GBIF Norway has collaborated with OBIS on exploring new extensions and solutions for supporting mobilization of MeasurementOrFact data with the Event Core model. Proposed solutions have been (partly) documented and presented at community workshops and at the regional European Node meeting in Tallinn in 2018.

South Africa: SANBI-GBIF is in the process of developing an MoU with iNaturalist for publishing citizen science data

2019 Work Items

- Extend 2018 work on matching stable Operational Taxonomic Units (OTUs) based on DNA sequencing to the GBIF taxonomic backbone, with targets of establishing dynamic links with the Barcode of Life Data System (BOLD) and incorporating occurrence data derived from environmental DNA sampling (eDNA) through the European Bioinformatics Institute (EBI).
- Run a campaign to mobilize data relating to vectors and hosts of human diseases, establishing an expert group to identify priority needs for biodiversity data supporting disease research, critical gaps in availability of such data in GBIF.org, and potential sources of data to fill these gaps. Campaign will use this analysis to engage directly with known holders of missing data and support publication through GBIF, as well as informing data mobilization priorities for use by nodes, publishers and and funders (see Activity 3a)

Participant plans for 2019

- **Australia:** The Atlas is in the process of auditing the potential sources of trait information in Australia. We will also work to identify use cases for trait information. Recommend developing a joint program with ELIXIR and BPA (Australia)
- **BHL:** Review options for transcribing in BHL
- **Norway:** Capacity training in Eurasia (BioDATA, organized by GBIFS, GBIF Norway and the Norwegian Node in BOLD - NorBOL, and funded by the Norwegian Centre for cooperation in

Education - SIU) includes training on BOLD data types and mobilization of sequence-based occurrence data in GBIF (see also activity 1b).
- **South Africa:** Citizen Science community and records catalysed and available through iNaturalist

Activity 3c - Engage data holders

Rationale

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

Approach

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

Tasks

3.c.i   Engage with natural history collections not yet publishing to GBIF

3.c.ii   Work with national citizen science groups

3.c.iii Promote national policies and platforms to enable publishing of environmental impact and monitoring data

3.c.iv   Promote importance of data sharing to funding bodies, licensing authorities and industry bodies

3.c.v   Address the tropical data gap

Progress

Initial discussions are being held in the final quarter of 2018 to explore how GBIF's partners at the UN Environment World Conservation Monitoring Centre (UNEP-WCMC) could facilitate the publication of biodiversity data from environmental impact assessments and other private sector data, through the Proteus Partnership of large extractive industry corporations. Several nodes have already demonstrated strong engagement with the private sector to mobilize such data, and these are among the guiding examples provided to the wider nodes community through GBIF.org (see Activity 1c).

Continued engagement through the European Citizen Science Association (ECSA) has helped GBIF to establish closer links with this community, both at Secretariat level and through nodes. This has included addressing the question of open licensing of data from citizen science projects, which remains a challenging issue for many of these groups. A related set of discussions with the Global Mosquito Alert Consortium enabled GBIF Spain to secure the publication of an important citizen science monitoring dataset from Mosquito Alert.

Participant contributions

Australia: Ongoing interactions with state agencies through annual meetings. Further engagement achieved through development of the BioCollect platform which is connecting the Atlas with an increasing number of citizen science groups in Australia.

France: Data mobilisation workshops at national and international level (see also 1b). Connection of new data publishers

Japan: Data providers increased to be 95 organisations.

Norway: A use case on occurrence-based data mobilization from the private sector in Norway (Asplan Viak, Multiconsult, and Sweco Norway) was prepared for online publication from GBIF.org in 2018.

South Africa: SANBI-GBIF and the Systematics Division at SANBI, has supported the implementation of the national Biodiversity Information Management and Foundational Biodiversity Information Management Forum (BIMF-FBIP Forum). This has been ongoing for over ten years and is a national conference, between SANBI and Partner Institutions aimed at harmonising biodiversity information sharing and discussing best practice in data and information systems management. It brings together government agencies, funding bodies, scientific, conservation and informatics practitioners to discuss data mobilisation efforts, funding opportunities and showcasing scientific projects which show the use and application of data for science and policy outcomes. SANBI-GBIF biodiversity geo-referencing training conducted in August 2018, alongside the BIMF-FBIP Forum.

Switzerland: Partnership with collection holding and research institutions active in DNA sequencing (linkage of sequence data, DNA-samples and reference specimens). Collaboration with national biodiversity data centers in order to capture and use species trait data.

2019 Work Items

- Work with DiSSCo project team to maximize opportunities for mobilizing collections data from European institutions, including in countries not yet participating in GBIF
- Based on discussions during Q42018, work with UN Environment WCMC and industry bodies (e.g. Proteus Partnership, IPIECA, IAIA) to establish node and helpdesk services for publishing private sector data, with options to use national nodes or hosted repository to support data mobilization
- Establish framework for regular updates of 2016 evaluation of contribution of citizen science contribution to GBIF data index, based on dataset tagging or other means
- Work with iNaturalist, iDigBio and nodes community to maximize opportunities for public engagement in GBIF data mobilization provided by the City Nature Challenge and WeDigBio.

Participant plans for 2019

- **Australia:** Maintain existing relationships, and further outreach with citizen science groups.
- **France:** New workshops
- **Japan:** Promotion of data publishing to be continued.
- **South Africa:** SANBI-GBIF and the Systematics Division will implement the 2019 National Biodiversity Information Management and Foundational Biodiversity Information Management Forum (BIMF-FBIP Forum).
- **South Africa:** SANBI-GBIF 2019 Training event will be conducted alongside the BIMF-FBIP Forum.
- **South Africa:** SANBI-GBIF will conduct a national training event, which will also include regional representation, as well as trainers which form part of the GBIF mentors network. This will support the development of communities of practice in data management nationally and regionally and aligns with the objective to develop the Centre for Biodiversity Information Management.

Activity 3d - Rescue datasets

Rationale

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

Approach

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

Tasks

3.d.i   Develop tools for reporting potential data sources for integration into GBIF

3.d.ii   Develop support materials (including accreditation) for collaborative data preparation and mapping datasets in GBIF

3.d.iii Develop site and support mechanisms for users to adopt and map datasets

3.d.iv   Review and update definitions of data publisher within GBIF to reflect collaborative data publishing

3.d.v   Develop partnerships with data journals to support data papers for rescued datasets

Progress

Since the establishment of the 'suggest a dataset' tool in 2017, little progress has been made in this activity area. Although many ad hoc efforts are being made at publisher, node and Secretariat levels to engage holders of important datasets and encourage publication through GBIF, clearly a more methodical approach needs to be devised and implemented if this activity of the implementation plan is to advance along the lines suggested.

2019 Work Items

● Establish workflow for prioritizing and addressing potential data sources reported through the 'suggest a dataset' tool, including involvement of nodes, mentors and crowd-source solutions

Activity 3e - Liaise with journals

Rationale

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

Approach

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

Tasks

3.e.i   Develop scalable approach to support research journals and data journals in publishing to GBIF network

3.e.ii Produce relevant support materials to justify benefits and explain processes to publish primary data

3.e.iii Integrate support for data journals into hosted IPT infrastructures and data rescue processes

Progress

During 2018, the Secretariat's science communication coordinator has reached out intensively both to journal publishers and to individual authors in an effort to improve the standard of citations of GBIF-mediated data in research. These communications are often based on specific examples of poor citation practices, and although responses have been mixed there has been positive feedback from several publishers and a commitment to improve citations. Collaboration with Elsevier has helped to improve the visibility of GBIF citations on article landing pages through use of the Scholix platform. Efforts to engage more than 150 individual authors have led to assurances of better citations in future, and also the identification of more than 300 download DOIs which had not been cited in the original articles but can now be linked to the research concerned. In cases where publications were still in a pre-proof stage, the citations were amended to include the DOIs in the citations, thus following the GBIF citation guidelines.

2019 Work Items

● Build on engagement with selected authors and academic publishers in 2018 to develop standard workflows and simple recommendations for depositing supplementary data in formats suitable for GBIF publication, as well as data citation practices enforced through editorial guidelines, with the aim of acceptance by all major publisher groups during 2019.

Participant plans for 2019

● **South Africa:** SANBI-GBIF will integrate data paper publication within Bothalia - African Biodiversity and Conservation Journal.

Priority 4 – Improve Data Quality

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

Activity 4a - Ensure data persistence

Rationale

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

Approach

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

Tasks

4.a.i   Identify and verify datasets within GBIF network without current owners

4.a.ii   Publish reference instances of these datasets within hosted IPTs

4.a.iii Develop processes and mechanisms for adoption of orphaned datasets by suitable agencies or experts

Progress

Datasets identified in the [2017 analysis and campaign](#) to identify and where possible re-activate lost data access points were migrated into a dedicated data hosting environment operated by the GBIF Secretariat. From here, they are available for "adoption" by Participant nodes or re-publication by the original owners. All other conditions, including the representation under the original data publishing institution, remain unchanged.

Participant progress

BHL: Migrating BHL wiki archives and public wiki; gathering BHL Europe data

Colombia: During July 2017, the 20 Colombian datasets identified by the IT team of GBIF as "orphaned" were reviewed and published again correctly by the Colombian Node. All the datasets had been detected as orphaned due a domain change from http://data.biodiversidad.co (deprecated) to http://ipt.biodiversidad.co. Tracking to the issue: https://github.com/gbif/watchdog/issues/15

Norway: GBIF Norway contributed in 2017 and 2018 to the Darwin Core compatibility of specimen-level DOI resolver services for the UN Food and Agriculture Organization in coordination with GBIFS and the GBIF data fitness for use priorities (see also activity 3a, 4b, and 5b).

Spain: Several hundred offline Spanish datasets were extracted from the GBIF.org index and relocated to an IPT server under management of GBIF Spain.

2019 Work Items

- Coordinate effort to achieve [CoreTrustSeal](#) data repository certification for relevant portions of the GBIF infrastructure. This will include the data management services within GBIF.org, but will also seek to identify a set of trusted repositories for publishing datasets within the GBIF network.
- Develop documentation for all stages in the GBIF data ingestion process, to enable data provenance to be fully understood by users of GBIF data.

Participant work plans

- **BHL:** Make progress on adding BHL Europe data to BHL
- **Norway:** GBIF Norway will continue contribution to the best practices documentation, capacity training and implementation of specimen-level DOIsin collaboration with the UN Food and Agriculture Organization in coordination with the GBIFS (see also activity 3a, 4b, and 5b).

Activity 4b - Assess data quality

Rationale

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

Approach

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

Tasks

4.b.i   Develop extensible data validation tools framework in partnership with ALA, TDWG and other networks (e.g. Symbiota, iDigBio, VertNet)

4.b.ii   Integrate consistent data validation tools in GBIF.org, national/regional portals, IPT and elsewhere

4.b.iii   Improve presentation and reporting of data validation results

4.b.iv   Develop regular data set reports for data publishers and nodes

Progress

A meeting was held with the ALA System Architect and a collaborative project was created to develop consistent data processing pipelines (parsing, interpretation and quality flagging) for use in GBIF, the Living Atlases project and beyond, including applications ranging in scale from laptop tools to GBIF volumes. Progress has been slower than expected, due to 1) introduction of new staff at GBIF, 2) the amount of support needed for running the existing live infrastructure and 3) lack of hardware to research the necessary technologies at this scale. Work will continue throughout 2018 and 2019 on this.

GBIF.org dataset pages now include a "metrics" tab which includes a summary of data processing issues flagged for records within the dataset. Similar information is also provided for the data associated with any arbitrary search on GBIF.org. This information will evolve as flagging of quality issues is standardised and as the GBIF and ALA ingestion codebases are aligned.

During 2018, the Data Products team started work to review metadata quality and to suggest a minimum set of fields and required content, particularly in the context of the BID programme.

Participant progress

Australia: The Atlas has contributed to the work of the TDWG Biodiversity Data Quality Interest Group. This has led to a set data quality tests that will guide the implementation of new and

improved data quality routines.

Colombia: "Colombia BIO | Gestión y publicación de datos" / Data management and publication Branch of the project that managed to improve the information of the collections in the geographic and taxonomic areas. For this, a team of specialists worked with data from nine biological collections, selected from an open call. In total there were 103,057 structured, georeferenced and published biological records, in a process that also left 13,846 new localities consolidated and available through SiB Colombia. Results: https://goo.gl/AD1fSx

France: GBIF France helps the new publishers to assess data quality before connecting the data

Japan: Assessment of data duplication and missing data implemented for the past 4.5 M specimen data.

Korea: Shared the coordinates (partially updated) for occurrences in Korea's published datasets via GBIF portal

South Africa: SANBI-GBIF has a data workflow in place to support the national community to deliver and share more accurate data.

2019 Work Items

- Review, consolidate and update existing documentation for data publishers. In particular, provide clear guidance on minimum requirements for published data.
- Develop metrics to track the completeness of core data elements and the degree to which supplied content is appropriate.
- Supply clear indicator measures for the completeness and usability of data as part of GBIF.org dataset pages, based on the example of the GEOLabel data branding model.
- Extend data quality assessment to include aspects only detectable above the level of individual records.
- Start to assess the patchiness of indexed data (geographical clustering, misleading accuracy or precision of coordinates), including evaluation of the apparent causes of data patchiness (e.g. grid-based relevées, other sampling approaches, country centroids), and include measures of data patchiness in the data index, at both dataset and record level in the data index
- Ensure that users of data are able to identify datasets or records that do not fulfil their criteria for geo-accuracy (including via facets within the portal, via the API, and in downloads).

Participant plans for 2019

- **Australia:** Work with GBIF on a combined code base to implement data quality routines based on the outputs of the TDWG Biodiversity Data Quality Interest Group.
- **Colombia:** Open Refine Scripts SiB Colombia is working in a set of scripts in OpenRefine for data quality management of primary biodiversity data. Some of these scripts are being used now by the Node Staff in the process of accompaniment that is done with the publishers, assessing the quality and generating quality reports for each dataset before them will be published through the IPT. The Colombian Node is making improvements on these scripts and developing new scripts to consolidate a toolkit in OpenRefine available for all the GBIF and informatics biodiversity community. An article about this work is also being built to be published in 2018 - 2019. OpenRefine scripts Repository: https://github.com/SIB-Colombia/data-quality-open-refine
- **France:** We will continue to help publishers assess data quality before publication.
- **Japan:** Continued assessment of data duplication and missing data
- **Korea:** Continue to share and update the coordinates for occurrences in Korea's published datasets; Find the reason of data lack of Occurrence ID in KBIF IPT

Activity 4c - Enable data curation

Rationale

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

Approach

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

Tasks

4.c.i   Develop network-wide approach to handling and processing annotations and feedback

4.c.ii   Enhance IPT to offer dataset-level peer review and commentary mechanisms and record-level annotations

4.c.iii   Develop mechanism and tools within GBIF.org for sharing cleaned and annotated datasets based on GBIF downloads ("reference datasets")

4.c.iv   Develop GBIF data workbench tool (within GBIF.org and possibly also as standalone) for cleaning and filtering network data (e.g. in red list assessments)

4.c.v   Develop strategy and support mechanisms for expert communities to curate sections of GBIF data

Progress

The GBIF Secretariat met with the Specify collection management software team and presented ways for annotations, services in support of data quality (e.g. verifying names against the Catalogue of Life etc) and issue flags detected from GBIF to be used to enhance data within the Specify platform.

GBIF participated in discussions at TDWG on possible infrastructure approaches to ensure that all specimens (and other occurrence records) are associated with persistent resolvable unique identifiers, based on lessons learned from the International Geo Sample Number (IGSN) system. The GBIF data index or better still, a consolidated data index supported by multiple data aggregation infrastructures (see *Activity 2a - Modernize data standards*), could fulfil the role played by SESAR within the IGSN system and offer stable handle-based resolution for all specimen and occurrence records. This would provide a solid foundation for data annotation and curation services.

Participant contributions

Colombia: "Colombia BIO | Gira de especialistas" / Specialists tours Two specialist tours were developed: the first, led by the CES University, brings together several of the best qualified entomologists in the country, experts in different groups of insects, to visit nine Colombian collections and carry out taxonomic identification, cataloging and skills strengthening, the second, led by the University of Quindío, brings together several experts in different biological groups to advance processes of taxonomic identification in several biological collections of the country.

Results: UCES (https://goo.gl/YwRhBL) UQuindío (https://goo.gl/xZXbqD)

Norway: GBIF Norway described for the 2017-2021 implementation plan contribution of developer resources to development of data annotation tools. Generic data annotation methodology is still in development (delayed because of reduced staff capacity at GBIF Norway). However, the planned approach is in a simplified form implemented for the citizen science digitization and volunteer transcription tools hosted by GBIF Norway – with source code shared in GitHub for further collaborative developments in the GBIF Nodes developer community.

2019 Work Items

- Continue exploration of the use of the GBIF data index to support stable persistent resolvable identifiers for all specimens and occurrence records.

Participant plans for 2019:

- **Norway:** GBIF Norway will engage with GBIF Sweden on possibilities for collaborative development of the ALA-based BioAtlas platform. Capacity at the Norwegian GBIF Node is dependent on securing additional national project funding.
- **Norway:** Further development of data annotation tools (based on the citizen science digitization tools) will depend on securing additional national funding.

Priority 5 – Deliver Relevant Data

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

Activity 5a - Engage academia

Rationale

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

Approach

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

Tasks

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

Progress

The development of a training curriculum focused on data use in the context of the BID programme (see Activity 1b) will help provide the foundations for better engagement with academic communities on the potential of GBIF as a resource for research. Many individual nodes are implementing excellent outreach programmes to improve links with academia, and these are among the guiding examples being collated to inspire the wider community via GBIF.org (see Activity 1c).

The Biodiversity Open Data Ambassadors programme (see Activity 1c) also provides a means to extend the reach of GBIF into relevant academic communities both globally and at national levels.

Participant contributions

Australia: The ALA has helped mobilise data with the Australian Virtual Herbarium for a number of university held collections including La Trobe University, The University of Melbourne and James Cook University.The ALA is engaged in the new EcoCloud project aimed at providing new data and modelling capability to support ecological research and management. The ALA will be working with nine other organisations to collate, integrate and give access to ecological data then develop and train people in the use of modelling and analysis tools to provide dynamic essential environmental

measures and facilitate new insights into ecosystems.

BHL: Working with academic partners on BHL via direct connections and grant opportunities.

France: GBIF France contributes to modules in Biology and Systematics degrees at MNHN and Sorbonne Université

Japan: To promote local use of biodiversity data through S-Net, we collected use case to showcase the possible way to exploit the data use stored in S-Net.

Korea: Increased engagement with National Institute of Biological Resources (NIBR) and T.B. Lee herbarium in Seoul National University as data providers

Norway: GBIF Norway presented a training workshop on GBIF data analysis in Rfor the Nordic Oikos conference in Trondheim 18-19 February 2018 (see also activity 1b). GBIF Norway presented the use of GBIF in agrobiodiversity as invited key speaker for the 8thInternational Triticeae Symposium. The wider Norwegian community is represented in the GBIF Science Committee and promotes the increased use of GBIF in academia with a focus on research topics related to ecological monitoring (see also activity 2a). GBIF Norway provides help desk support and has organized training workshops for academic use of GBIF with focus on support for Norwegian and Nordic scientists. Activities are in part covered by secured additional project funding.

South Africa: SANBI-GBIF has developed a ToR for Research Chair in Biodiversity Informatics. This work forms part of a Strategy to develop a Centre for Biodiversity Information Management.

South Africa: SANBI-GBIF is working with Institutions of Higher Learning to develop and implement a Biodiversity Information Management teaching module at the Honours level. Here an MoU is being developed between SANBI and Sol Plaatjes University.

2019 Work Items

- Enhance engagement with academic communities through Open Biodiversity Data Ambassadors (see Activity 1c), and sharing best practices from nodes engaged in promoting biodiversity informatics in academic curricula
- Develop materials to guide academic audiences on use of GBIF-mediated data in thematic areas, e.g. invasion biology, epidemiology (see also Activity 1b).

Participant plans for 2019

- **Australia:** Further engagement and integration with the EcoCloud project. http://www.ecocloud.org.au/.
- **France:** We will propose a module for the MNHN masters
- **Korea:** Increase engagement with interested biodiversity groups in Korea
- **Norway:** The wider Norwegian GBIF community will continue to engage academia and promote increased use of GBIF data with focus on ecological monitoring and benefits of the sampling event model (see also activity 2a).
- **Norway:** GBIF Norway and GBIFS will provide academic training in Eurasia and in Norway on biodiversity informatics through the Norwegian Research School on Biosystematics (ForBio) with project funding (BioDATA) from the Norwegian Centre for cooperation in Education (SIU) (see also activity 1b).
- **South Africa:** Further engagement to seek funding and consider the BDI Research Chair appointment, to be conducted with DST/NRF, University Partners and other funders.
- **South Africa:** SANBI-GBIF will looks at mechanisms to operationalise the Centre for Biodiversity Information Management, through added strategic and technical assistance.

Activity 5b - Document needs

Rationale

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

Approach

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

Tasks

5.b.i   Prioritize areas for fitness-for-use assessments

5.b.ii   Support fitness-for-use groups

Progress

While no new fitness-for-use groups were established during 2018, considerable progress was made in implementing the recommendations of the group in alien and invasive species, with the publication of country and island checklists from the Global Register of Introduced and Invasive Species (GRIIS) by IUCN's Invasive Species Specialist Group, and the incorporation of these lists onto species pages on GBIF.org so that information on invasiveness status is readily visible to users of the site. Consultations have continued with members of this task group and the Secretariat of the CBD on recommended next steps for improving the usefulness of GBIF to researchers on invasive species and to policy makers dealing with their impacts.

2019 Work Items

- Review and consolidate responses to recommendations from previous fitness-for-use groups, before commissioning new expert groups other than the group advising the campaign on data for disease vectors and hosts (see Activity 3b)
- Analyse and evaluate the requirements for pre-generated downloads based on the analysis of incoming user requests for custom downloads that cannot be served through the user interface at gbif.org or API access.

Participant plans for 2019

- **Australia:** Recommended increased sharing of communications, use cases and policy initiatives. Recommend that the community works collaboratively to define targets around coverage and extent of our data products (e.g. temporal/ spatial and taxa). This will help define the level of effort required for data publishers to met these gaps. Also define level of use from Occurrence to Abundance to True Absence.
- **Norway:** GBIF Norway will continue to capture and document fitness-for-use needs from the

agrobiodiversity community (with focus on Nordic and European crop wild relative communities) and follow up on recommendations from the 2016 fitness for use task group reports (see also activity 3a).

● **Norway:** The wider Norwegian GBIF community will continue to capture and document GBIF fitness for use in academic and applied research on ecological monitoring using the sampling event data model (see also activity 2a).

Activity 5c - Support biodiversity assessment

Rationale

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

Approach

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

Tasks

5.c.i   Enhance GBIF.org presentation of sampling event data

5.c.ii   Support development of species distribution/population abundance EBVs

5.c.iii Showcase tools and practices for use of GBIF in Red Listing and other species assessment processes

Progress

GBIF has continued to participate in GEO BON processes to develop Essential Biodiversity Variables as a mediating structure between 1) the complex spread of measurements captured in regard to biodiversity and natural systems and 2) the needs of researchers and policymakers focused on large scale perspectives regarding biodiversity change and sustainability. Within GEO BON, divergent opinions remain regarding whether the EBV or EBVs for species distribution, population abundance and community composition correspond directly to a matrix of measurements of species occupancy and abundance for any portion of space and time and for any taxa or whether more derived variables representing e.g. some kind of alpha and beta diversity should be adopted as the EBV(s). In practice, for GBIF, the distinction is relatively unimportant, since GBIF can and should serve regardless as the infrastructure for organising all available evidence on the presence, absence and abundance of any species at any time and place. However the EBV definitions are finally resolved, GBIF will play the same fundamental role. Indeed, by focusing on increased mobilisation and integration of all relevant data, and especially sampling-event data, and by improving the options for users to filter, access and pivot GBIF data based on facets such as latitude, longitude, time and taxon, GBIF can play a critical role in evaluating alternative perspectives on EBVs.

A [Memorandum of Understanding between GBIF and IPBES](#) was signed on the margins of the sixth plenary meeting of the platform in Medellin, Colombia, in March. Focusing on specific ways in which

GBIF and IPBES can collaborate better both at global level and between GBIF nodes and IPBES national focal points, the MoU helped to raise the profile of GBIF's role as a foundational source of data to support IPBES assessments, and was well received by many IPBES delegates as well as by the GBIF participant community.

A Memorandum of Cooperation is being finalized between GBIF and IUCN, with joint work areas identified to increase interoperability between IUCN and GBIF portals, to increase data sharing between the two organizations, and to collaborate more closely in development of shared training materials. GBIF will also be involved in discussions concerning the IUCN Monitoring Specialist Group.

Participant progress

Australia: ALA has contributed the EBV goals and within Australia to the Environmental Assessment Measure Program

Australia: The Atlas has extended support for ingestion and indexing of sampling-event based Darwin Core Archives. The Atlas has also invested in joint programs that increase capability and understanding of biodiversity data, management and modeling. This includes the EcoEd program and adoption of tools that simplify analysis of ALA and GBIF data. Elements of these programmes could be of value more broadly within GBIF training and for engagement with tertiary education.

Norway: GBIF Norway was a partner in the GBIF CESP project European Bireme on assessing the role of European GBIF Nodes in the national EU biodiversity reporting and data flow pathways (see also activity 1c).

South Africa: Efforts are underway through both the Research and Policy Advice divisions at SANBI to support biodiversity mapping priorities (Botswana, Ethiopia and Malawi) with UNEP-WCMC;and Red-list assessments in countries in Africa. SANBI is working with Mozambique to support Red-list assessment work, through its Threatened Species Programme, within a BID funded project.

2019 Work Items

- Continue contributing to GEO BON activity to develop Essential Biodiversity Variables, particularly in regard to Species Distribution, Population Abundance and Community Composition. As the global infrastructure for organizing species occurrence data in support of EBVs, work with GEO BON to align with the EBV framework.
- Enhance GBIF.org to ensure that sampling event data can be accessed in forms appropriate for further processing as EBV data.
- Explore options for new data download formats which provide pivoted views of available data in forms appropriate for processing as EBV data. Such pivoted views are likely also to assist other users with simplified access to relevant GBIF data.

Participant work plans for 2019

- **Belgium:** Explore and prototype a tool, based on occurrences API, that will return a species list from a user-defined polygon.
- **Norway:** The wider Norwegian GBIF community will promote the sampling event model for biodiversity assessment activities with focus on national activities in Norway (see also activity 2a).
- **South Africa:** SANBI-GBIF will engage with the IPBES Chair for data and knowledge, to reinforce a stronger IPBES perspective to the ACM work, as it shows the data needs and the potential.   The IPBES Africa Assessment reinforces the need for investment in biodiversity information and the ACM.

Activity 5d - Assess impact

Rationale

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

Approach

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

Tasks

5.d.i   Develop sustainable approach to literature tracking

5.d.ii   Automate detection and reporting of use of GBIF DOIs

5.d.iii Provide mechanisms to report and track uses of data and grey literature

5.d.iv Develop a valuation for the services delivered by the GBIF network

Progress

While the main sources of literature using GBIF-mediated data–such as Google Scholar, Scopus, etc–are mainly based around peer-reviewed journal publications, they do include, to an increasing extent, academic dissertations, official government reports, environmental impact assessments, etc. have been added to the data use archive and featured on GBIF.org. New sources of literature are being considered and added as relevant. In 2018, the communications team added CrossRef Event Data, providing DOI-based citations from a variety of non-traditional sources such as Wikipedia, Twitter, Hypothes.is and several others.

Participant contributions

Australia: ALA completed an impact assessment during 2018.

Australia: The Atlas has added DOI support for occurrence data downloads, and re-uses DOIs supplied by GBIF for datasets. This will simplify and extend tracking of data usage in publications. In addition, the Atlas (including other Living Atlas installations) now uses the GBIF citations API to highlight reuse of datasets.

South Africa: The efforts being conducted through the ABC projects and the Foundational

Biodiversity Information Programme of SANBI, reinforces that data-science-policy value chain and aims to ensure that data mobilised feeds into policy outcomes.

2019 Work Items

- Commission external review of GBIF operation and functions. At GB24, the Governing Board approved €30,000 in the 2018 budget to be allocated towards an external review of GBIF and its operations. Following explorations within the Executive Committee, it is agreed that the review can best be handled via a contract to CODATA, which led the first review of GBIF, completed in 2005. It has now been 20 years since the OECD Megascience Forum recommended the establishment of a Global Biodiversity Information Facility, so an external review led by CODATA as a representative of the global scientific community will be timely. An additional €70,000 is included in the 2019 budget (for a total of €100,000) to support a broad review of GBIF's operation, services, governance and place within the broader landscape of biodiversity-related organizations.

Participant plans for 2019

- **Australia:** Recommend that GBIF explores offering analytical services on top of their data to countries as a service to provide a greater value proposition.
- **Australia:** Recommend continuing work to assess impact and jointly appointing a resource to lead such work. Recommend incorporating qualitative measures as a way of assessing impact.
- **South Africa:** The efforts to reinforce the data-science-policy value chain through the ABC projects and the Foundational Biodiversity Information Programme of SANBI, will continue into 2019.