GLOBAL NODES MEETING 2019
Report

Naturalis Biodiversity Center, Leiden
19 October 2019

Contents

Executive summary by Nodes Chair, André Heughebaert

As far as I can remember, Global Nodes Meetings (GNMs) have always provided a great opportunity for the Nodes community to learn from each other, develop best practices, share their concerns, define recommendations and set priorities. The 15th Global Nodes Meeting in Leiden gave all that to the participants: close to 50 Node Managers plus 20 Nodes Staff/observers with a strong presence of the Secretariat. As usual the heterogeneity of the participants was both a challenge and an opportunity. Naturalis was a superb venue and their professional organisation was faultless.

The one day agenda, crafted by the Nodes Steering Committee, offered presentations from the Nodes and the Secretariat with thematic and regional discussions groups. This formula offered a good balance between one-to-one, one-to-many and many-to-many interactions.

For the Secretariat, Tim Robertson gave an inspiring view of future directions for GBIF.org and the infrastructure services that could lower the technical threshold for participants, publishers and users.

Nodes stories and data use cases were an excellent opportunity to showcase where we are, what are our threats and opportunities, and where we should go next. At no surprise,

sustainability of the nodes' teams and funding is the biggest concern for most of us, even for well experienced participants.
During the thematic breakout groups, participants could exchange on five subjects. Regional breakout groups gave the opportunity to discuss regional priorities. All the regions were not equally represented, but overall discussions were quite open and fruitful.
The conclusion session was shortened and did not allow us to draft a Global Nodes Strategy, but the NSG caught up on that and delivered a clear and concise document based on the themes raised during the global nodes meeting that will guide the Nodes global efforts for the two coming years.
It was very pleasant to chair this meeting and it confirms that the Nodes Committee has a key role to play in the establishment of the GBIF Strategy for 2022-2026.

Meeting preparation and objectives

The meeting was prepared and chaired by the Nodes Steering Group (NSG), with support from the GBIF Secretariat. The NSG invited all nodes to contribute to the meeting by sharing their experiences as lightning talks.

The aims of the meeting were to:

- Share information on the progress made by nodes and regions
- Establish collaboration mechanisms
- Set common priorities for the Nodes Committee for the upcoming period

The expected outcomes were to:

- Foster new collaborations
- Create a better understanding of the global aspects
- Agree on recommendations
- Identify actions for the two coming years
- Inspire the Nodes Committee Chair presentation to the governing board at GB26

Participants and Secretariat support

The meeting was attended by 70 people, of which 46 were Node Managers.

Jean Paul Kubwimana	Node manager	Albertine Rift Conservation Society (ARCOS Network)
Benjamin Komac	Node manager	GBIF Andorra
Esperança Maria Eduardo Francisco da Costa	Node manager	GBIF Angola
Anabela Plos	Node manager	GBIF Argentina
Christian Elloran	Node manager	ASEAN Centre for Biodiversity
David Martin	Node manager	Atlas of Living Australia

Oleg Borodin	Node manager	GBIF Republic of Belarus
André Heughebaert	Node manager	Belgian Biodiversity Platform
Jean Cossi Ganglo	Node manager	GBIF Benin
Martin Kalfatovic	Node manager	Biodiversity Heritage Library
Luiz Henrique Mourao do Canto Pereira	Node manager	GBIF Brazil
Jean François Moussa	Node manager	GBIF Cameroon
James Macklin	Node manager	Canada Biodiversity Information Facility
Carole Sinou	Node manager	Canadensys
Maofang Luo	Node manager	Chinese Academy of Sciences
Chihjen Ko	Node manager	Taiwan Biodiversity Information Facility
Francisco Pando	Node manager	Ciencia y Tecnología para el Desarrollo
Genuar Román Núñez Vega	Node manager	Costa Rica Biodiversity Facility
Veljo Runnel	Node manager	GBIF Estonia
Eija-Leena Laiho	Node manager	GBIF Finland
Anne-Sophie Archambeau	Node manager	GBIF France
Walter Berendsohn	Node manager	GBIF Germany
David Jennings	Node manager	iDigBio
Gerald Guala	Node manager	Integrated Taxonomic Information System
Liam Lysaght	Node manager	National Biodiversity Data Centre
Tsuyoshi Hosoya	Node manager	GBIF Japan
Lawrence Monda	Node manager	GBIF Kenya
Bendictus Freeman	Node manager	GBIF Liberia
Tania Walisch	Node manager	GBIF Luxembourg
Wouter Addink	Node manager	Naturalis Biodiversity Center
Niels Raes	Node manager	Netherlands Biodiversity

		Information Facility
Omokafe Alaba Ugbogu	Node manager	GBIF Nigeria
Dag Endresen	Node manager	GBIF Norway
Donat Agosti	Node manager	Plazi
Piotr Tykarski	Node manager	Polish Biodiversity Information Network
Rui Figueira	Node manager	GBIF Portugal
Fatima Parker-Allie	Node manager	South African Biodiversity Information Facility
Cristina Villaverde	Node manager	GBIF Spain
Geoff Ower	Node manager	Species 2000
Anders Telenius	Node manager	GBIF Sweden
Pascal Tschudin	Node manager	GBIF Switzerland
Pierre Raoufou Radji	Node manager	GBIF Togo
Jo Judge	Node manager	National Biodiversity Network (UK)
Abby Benson	Node manager	U.S. Geological Survey
David Bloom	Node manager	VertNet
Sinh Nguyễn Văn	Node manager	GBIF Viet Nam
Dimitri Brosens	Observer	Belgium
Maxime Coupremanne	Observer	Belgium
Nils Valland	Observer	Norway
Knut Anders Hovstad	Observer	Norway
Natalya Ivanova	Observer	Russia
Maxim Shashkov	Observer	Russia
Olaf Banki	Observer	Species 2000
Sylvie Fanta	Observer	Camercoon
Michéle Marcotte	Observer	Canada
Daphne Duin	Observer	Naturalis Biodiversity Center
Maarten Schermer	Observer	Naturalis Biodiversity Center

Andrea Hahn	GBIF Secretariat staff	
Andrew Rodrigues	GBIF Secretariat staff	
Anne Mette Nielsen	GBIF Secretariat staff	
Joe Miller	GBIF Secretariat staff	
Kyle Copas	GBIF Secretariat staff	
Laura Anne Russell	GBIF Secretariat staff	
Maheva Bagard Laursen	GBIF Secretariat staff	
Marie Grosjean	GBIF Secretariat staff	
Marlene Dalsgaard Nielsen	GBIF Secretariat staff	
Mélianie Raymond	GBIF Secretariat staff	
Morten Høfft	GBIF Secretariat staff	
Tim Hirsch	GBIF Secretariat staff	
Tim Robertson	GBIF Secretariat staff	

Agenda, presentations and discussion summaries

Session 0: Setting up the scene (09:00-09:20)

Chaired by André Heughebaert

The meeting was introduced by the Chair, explaining the meeting structure, goals and practicalities.

- Presentation

Session 1: Common understanding of progress (09:20-10:20)

Chaired by André Heughebaert

Tim Robertson, Head of Informatics at the GBIF Secretariat, presented the work to lower the technical threshold for participating in GBIF.

- Abstract
- Presentation

Session 2: Nodes stories (10:20-11:00)

Chaired by Anders Telenius

Nodes were invited to present lightning talks on national/thematic/sub-regional/regional collaborations.

- Invitation to nodes to prepare lightning talks
- Abstracts
- Presentations:
    - NLBIF activities as DiSSCo national task force lead (Niels Raes)

- International data sharing and data integration through RDA (Wouter Addink)
- Challenges and future direction from recent activities of Japan Node of GBIF (Tsuyoshi Hosoya)
- GBIF Norway after 2019? (Dag Endresen)
- The achievements of GBIF Nigeria - NgBIF (Omokafe A Ugbogu)
- SBDI: New opportunities for data driven research (Anders Telenius)

---Coffee Break 20'---

Session 3: Thematic groups

Session 3a: Parallel thematic discussions (11:20-12:30)

Participants split into five thematic groups:

1. **BID and Beyond**
   - Abstract
   - Discussion notes
2. **Beyond one billion**
   - Abstract
   - Discussion notes
3. **Data repositories**
   - Abstract
   - Discussion notes
4. **Hosted portals**
   - Abstract
   - Discussion notes
5. **Data in GBIF based on generic sequences**
   - Abstract
   - Discussion notes

---Lunch Break 60'---

Session 3b: Thematic discussions continued (13:30-14:00)

Chaired by André Heughebaert

Participants reconvened in plenary for verbal presentations by representatives from the thematic groups.

Session 4: Data Use Cases (14:00-14:50)

Chaired by Anders Telenius

Nodes were invited to present lightning talks on data use stories : Data curation, fitness-for-use, policy driven data mobilization…See final list of Data Use Cases.

- Invitation to nodes to prepare lightning talks
- Abstracts
- Presentations:
  - SBDI: New opportunities for data driven research (Anders Telenius)
  - Regional Master program in Biodiversity Informatics: some research achievements (Jean C. Ganglo)
  - Expanding the workflow from scholarly published data to GBIF (Donat Agosti)

- Improving data availability with sampling event and extended measurement or facts: Examples from OBIS-USA (Abby Benson)
- Status of GBIF data usage in China (Maofang Luo)
- Data use cases: Frictionless DarwinCore (André Heughebaert)

Session 5: Regional groups

Preliminary discussion on the regional nodes meetings in 2020 (Tim Hirsch)

Session 5a: Parallel Regional discussions (14:50-15:40)

Participants split into five regional groups to discuss regional priorities on engagement, capacity strategy, and lowering the technical threshold to participation.

- Session guidelines
- Introductory presentation

1. **Africa**
   - Session notes
2. **Asia**
   - Session notes
3. **Europe**
   - Session notes
4. **Latin America & Caribbean**
   - Session notes
5. **North America** and **Oceania**
   - Session notes

---Coffee Break 20'---

Session 5b: Regional discussions-continued (16:00- 16:30)

Chaired by André Heughebaert

Participants reconvened in plenary for presentations by representatives of the regional groups.

Session 6: Conclusions (16:30-17:00)

Chaired by André Heughebaert

This last session aimed to wrap up discussions and come up with:

- **Draft Nodes Strategy for 2020-2021**
  Based on GBIF Work programme 2020, what are the priorities as Nodes Committee? How could we better contribute to it? What do we want to see included in WP2021. This document will be further elaborated by the NSG after the meeting.
- **NC message to the Secretariat** (to be delivered at the next NSG meeting)
- **NC message to the HoDs** (to be delivered by NC Chair at GB26)

- Presentation introducing the session
- Draft Nodes Strategy 2020-2021

Closing of the Nodes Committee meeting by the Chair and group photo.

Next steps

The next Global Nodes Meeting will take place in 2021 alongside GB27. Until then, the Nodes Committee and NSG will have to:

- Finalize our Nodes Strategy 2020-2021 and report on progress toward the objectives
- Capacitate our network through BID, BIFA, CESP
- Prepare and organize the Regional Nodes meetings in 2020
- Participate in the brainstorm for coming GBIF Strategy 2022-2026
- Bring Nodes perspectives on how to deal with the recommendations of the 20-year Review

ANNEXES

Session 1: Common understanding of progress

Common understanding of progress

Title

Lowering the technical threshold to participate

Tim Robertson, 1hr

Summary

As data publishing activities grow across GBIF it is recognized that the simplicity of the data model supported by GBIF hinders progress. At the same time the GBIF Secretariat receive increasing calls from both new and well-established nodes and groups to provide infrastructure for repositories and for discovery and access services. In some cases this is driven by the cost of both developing and operating this infrastructure nationally and in other cases it is seen as simply too difficult. This presentation will introduce recent work, ideas and opportunities for discussion to improve services and simplify participation in GBIF including:

1. Enabling Nodes to manage content in GBIF (done / underway)
    a. A shared open registry now enabling Nodes to curate content in GBIF
    b. Administration console for managing and debugging data ingestion processes and contributing to the controlled vocabularies used to interpret data
2. Broadening the data model (idea phase)
    a. Why we have limitations
    b. Example visuals illustrating detail pages for new content types including Organism views (such as a tracked individual, catch&release programme), Specimen views (sequences, physical location etc), long term monitoring sites (site/species/time matrices)
    c. Example visuals illustrating how portals (e.g. GBIF.org) may evolve to provide discovery services for these catalogues of data
3. Using GBIF infrastructure to power a portal (moving from idea into early prototypes)
    d. Fully hosted portals on GBIF mediated data with example visuals for
        i. Program specific views (BID)
        ii. National views (Canada)
        iii. Thematic views (e.g. a global virtual NHM collection catalogue and specimen search)
    e. Addressing cross cutting concerns of citation, registration licensing etc
    f. Opportunity to allow nodes to use the infrastructure and focus on data related and training activities
4. Addressing fragility of data repositories through well managed open science repositories
    a. Co-located IPTs
    b. Partnering and using open repositories such as Zenodo

Session 2: Nodes stories

GNM15 Lightning talk sessions

Introduction

Two plenary sessions are open to lightning talks:

- Session2: Node stories
  national/thematic/sub-regional/regional collaborations...
- Session4: Data Use cases
  Data curation, fitness-for-use, policy driven data mobilization...

With these talks, nodes managers have the opportunity to present their activities or projects to their peers. The aims of these talks are to:

- Update the Nodes Committee on the diversity of Nodes activities worldwide
- Exchange best practices
- Foster future collaboration and avoid duplication of effort
- Seed ideas for session 3 (thematic groups) and session 5 (regional groups)

Format

Presentations will take the form of lightning talks, not more than 5 minutes long.

> *"The goal of lightning talks is to articulate a topic in a quick, insightful, and clear manner. These concise and efficient talks are intended to grab the attention of the audience, convey key information, and allow for several presenters to share their ideas in a brief period of time."*[9]

Node managers are therefore kindly invited to stick the time limit, to focus on the essential and to avoid details. If time allows, questions from the assembly will be answered in plenary.

See also general guidelines

Procedure

Step 1

If you are interested to present your work, send the **title** and **abstract** of your talk and your **preferred session** before October 6th to André Heughebaert . Be aware that the number of slots will be limited by the duration of the session. You will get a confirmation if your talk is on the session agenda.

Step 2

The presentation slides, if any, will be made available to the organizers at least three days before the meeting. (Please send your presentation slides to gb26@gbif.org).

Nodes stories

Lightning talks proposals

1. NLBIF activities as DiSSCo national task force lead
2. International data sharing and data integration through RDA
3. Challenges and future direction from recent activities of Japan Node of GBIF
4. Project funding periods for the Norwegian GBIF Node
5. The Effective Node: Doing the greatest good with minimal funding
6. The achievements of The Nigeria Biodiversity Information Facility
7. Progress on the national Swedish Biodiversity Data Infrastructure (SBDI)

---

1. NLBIF activities as DiSSCo national task force lead

Niels Raes (NLBIF)

NLBIF, as partner in DiSSCo has been offered the lead position of the DiSSCo-NL national task force (ntf) and informs and updates the 13 DiSSCo-NL partners on DiSSCo developments. Important for DiSSCo at this stage is to arrive at an accurate estimate of the total European Natural History Collection. This requires a collection description scheme that is adopted by all DiSSCo partners so that summary statistics can be calculated and the level of digitisation can be estimated. The latter is of great importance to facilitate industrial scale digitisation, digitisation-on-demand requests, and data sharing with GBIF. Together with key stakeholders a preliminary classification scheme was designed that is contributing to the development of the TDWG Collection Description (CD) scheme. The implementation of this preliminary scheme was piloted with the DiSSCo-NL partners and the results are visualised in a data dashboard. I will shortly highlight the advantages for NLBIF being a DiSSCo partner, show the preliminary collection description scheme, and the resulting data dashboard.

2. International data sharing and data integration through RDA

Wouter Addink(DISSCo)

The Research Data Alliance provides a neutral space where its 8600 members from 137 countries together develop and adopt infrastructure recommendations to promote data-sharing practices and data-driven research. RDA provides recommendations on e.g.

FAIRSharing, persistent identifier kernel information, attribution metadata and dynamic data citation. Such recommendations underpin the development of a cross-domain 'data fabric', an architecture and set of data services that provide consistent capabilities across scientific domains and technical solutions. Therefore, when applied to GBIF data, these recommendations will enable better integration in the future with other domains, e.g. the medical domain or chemistry. It will allow researchers and innovators to openly share GBIF data across technologies, disciplines, and countries to address the grand challenges of society.

RDA Europe has provided me a RDA ambassador grant to identify the current needs of the international biodiversity and geodiversity data community, including GBIF, for the RDA discipline-specific Interest Group: Biodiversity Data Integration IG . With the outcomes of this activity the group can be adapted to better represent the needs from GBIF and other community members. When combined with a GO FAIR Implementation Network (IN) to define and create specific materials and tools this group can provide a powerful instrument to better integrate GBIF data with research environments and with data from other domains for multidisciplinary research.

Looking at collection specimens as example, DiSSCo is basing its technical architecture design on the concept of Digital Objects (DO) as developed in RDA. It is using RDA recommendations developed in e.g. the RDA PID Kernel Information WG, the RDA/TDWG Attribution Metadata WG and other groups.

3. Challenges and future direction from recent activities of Japan Node of GBIF

Tsuyoshi Hosoya, Japan Node of GBIF (JBIF)

Japan Node was established to facilitate mainly aggregating occurrence data mainly based on natural history specimens and field observation records. National Museum of Nature and Science (NMNS), Tokyo University and National Institute of Genetics (NIG) aggregate the data provided from various museums, universities and institutes throughout Japan, and these three institutions provide the data to GBIF from two IPT servers maintained at NMNS and NIG. To popularize the biodiversity data, the data from specimens are also shared in Science Museum Net (S-Net http://science-net.kahaku.go.jp/) in Japanese.

While these databases are used for scientific researches and/or museum exhibitions, two major challenges were identified. One is incentives partially covered by providing processing fee for the conversion of the collection data. The other is importance of bridging the language barrier between English and Japanese. In this respect, S-Net shows a significant functions.

To popularize the use of biodiversity data, two meetings mainly for local data providers and one meeting for public are being convened. Providing opportunities of use of S-Net in high school lectures are being planned.

We still feel that significant use case of biodiversity data are few, and continue to increase activities in this aspect. We recently engaged with "Japan Search", a national cross-sector

portal for federated search. By providing data to Japan Search, unexpected “chemical reaction” with natural historical specimens and cultural properties may happen.

In terms of future direction, we expect more contribution to the digital archive, and promotion of data exploitation, and pay more attention to improve data quality.

4. Project funding periods for the Norwegian GBIF Node

Dag Endersen, GBIF Norway

The GBIF Nodes management training at the start of the 2019 global nodes meeting focus on mentoring the establishment of effective nodes for new GBIF participant countries. However, we should also remain attentive that longterm and established nodes can also be vulnerable to disruptions. This lightning talk will present the use case of the longterm Norwegian Node currently heading towards a funding disruption. Norway joined GBIF in 2004, 15 years ago, after a careful process coordinated by the Norwegian Ministry of Education and Research and the Research Council of Norway. The following year in June 2005 the new GBIF Node for Norway was established with a clear and longterm mandate and hosted by the UiO Natural History Museum at the University of Oslo. However, the chosen funding mechanism was not longterm but divided into a series of project periods. At first, the node budget was approved for a period of three years (2005-2007), followed by two five-year periods (2008-2011 and 2012-2016) and another three-year project (2017-2019). The rationale stated by the Research Council for reducing the most recent project period to three years was an intention for the GBIF node to move to a new long-term operational model. A similar process towards a long-term operational model has recently been successfully implemented in Finland and is currently in progress in Sweden. However, as of September 2019, no longterm model is yet in place and the GBIF node is heading for 2020 without any node budget in place. This lightning talk will briefly present some of the alternatives that have been explored for finding a permanent operational model for GBIF-Norway.

5. The Effective Node: Doing the greatest good with minimal funding

David Bloom, Vernet

The creation and support of a regional or thematic node can be a challenging process. This lightning talk will focus on the creation of the VertNet Node, how a small node with the right people can make a big impact (225 dataset, 84 publishers, 10 countries with more on the way), and how it has all been done with little to no funding.

6. The achievements of The Nigeria Biodiversity Information Facility

Omokafe ugbogu, NgBIF

The Nigeria Biodiversity Information Facility (NgBIF) is new and had won two National projects under GBIF: Capacity Advancement for the Nigeria node of GBIF (BID BID-AF2017-0210-NAC) and Nigeria Node Mentoring (CESP2018_012) and one Small project : Species diversity, abundance, banking and barcoding of Odonata of Southern and Eastern Nigeria (BID_AF2017_0311_SMA) won University of Lagos. Some of the activities of NGBIF can be found on this link. Nigeria IPT is hosted by GBIF France with this link . Under these projects, 30,309 Data were mobilized and published. In addition, 17,105 data occurrences mobilized under JRS sponsored project titled " Capture of Primary Biodiversity Data for West African Plants " were published on our IPT. Our activities under GBIF have brought Nigeria Biodiversity to a greater height, while some Biodiversity holders have been trained in Data mobilization and publication skills on GBIF platform.The Nigeria Node participated in the Regional and International activities of GBIF.

7. Progress on the national Swedish Biodiversity Data Infrastructure (SBDI)

Anders Telenius, Sweden

Based upon F.A.I.R. principles and using open-source software developed in international collaboration within the Global Biodiversity Informatics Facility (GBIF) and Living Atlases (LA) communities, and on tools developed within the Biodiversity Atlas Sweden and Swedish LifeWatch infrastructures the Swedish Biodiversity Data Infrastructure (SBDI) will establish 1 January 2021. The consortium, consisting of 11 national partner institutes will be the key national e-infrastructure mobilizing data from a wide range of sources into a single Swedish biodiversity data layer, and provide access, analysis and visualization services offering the research community rich opportunities for innovative, interdisciplinary research on biodiversity and ecosystems. SBDI contributions to the LA developer community will focus on Swedish areas of excellence: system integration, near-real-time data mobilization, marine biodiversity, natural history collections data, systematic monitoring programs, biotelemetry, microbial diversity (prokaryotes, unicellular eukaryotes, and microscopic fungi), palaeoecology and molecular biodiversity data.

Session 3: Thematic groups

1. BID and Beyond

Introduction to the thematic break out group at the Global Nodes Meeting 2019

Facilitator: Jean Ganglo
Contacts at GBIFS / co-facilitators: Maheva Bagard Laursen (mblaursen@gbif.org), Mélianie Raymond (mraymond@gbif.org)

Refer to the BID impact summary flyer prepared for the meeting in Brussels in November 2019.

The Biodiversity Information for Development (BID) programme was to be in its final year, but is now anticipated to continue with a funding top-up from the EU. This will enable further biodiversity data mobilization and use actions in the ACP regions.

In a meeting in Brussels in November, we aim to explore with funders and partners how we can expand BID's impact in the future. BID could be an umbrella for many capacity development actions building on the tools, training materials and approach developed by the GBIF community in this first phase.

In preparation for this meeting, and to help shape the direction of future capacity development actions in the GBIF community, we would like to discuss the following two areas with nodes in the breakout session at the global nodes meeting:

1. Scoping demand for expanding BID

What types of end uses for biodiversity data could projects target in future?
Should we explore the links to meeting information needs under the Sustainable Development Goals (e.g. on human health, climate change, food security)?
How can we use these kinds of projects to best strengthen national and regional infrastructures?
Where should we expand the BID approach?
What information needs are best addressed on the regional level?
Do we need to target other communities working with different data types to integrate the BID approach e.g. in training on mobilizing DNA barcode data?
How can we overcome the challenges- and lag time- in going from the primary biodiversity data to influencing decision making?

2. Implementation models for expanding BID

How will nodes be involved in expanding the BID approach?
Is there further scope for reusing BID-developed training and materials in node-level collaborative projects and programmes?
Can nodes help approaching funders to gain support for further BID actions?
Can nodes help identifying opportunities to integrate data mobilization and data use training into other programmes?

The BID approach has relied strongly on the community of practice. How can we ensure this is scalable?

Discussion notes BID and Beyond

1. **What types of end uses for biodiversity data could projects target in future?**

Should we explore the links to meeting information needs under the Sustainable Development Goals (e.g. on human health, climate change, food security)?
How can we use these kinds of projects to best strengthen national and regional infrastructures?

1 category of end user is policy makers and decision makers
Also university students making research products

Target thematic data mobilization to fill data gaps
E.g. Benin students working on public health, e.g. Ebola, target public health sector for data mobilization
Risk maps for diseases can support the decision makers

Public health students can help process the data into finished products to be used by decision makers e.g. ecological niche modeling
Students important target - data generation and data use

Include policy makers within the project consortia
South Africa data mobilization grants include a policy making component for 2 -3 year projects. Supporting institutional collaboration connecting to policy makers. E.g. Marine data used to support decision making around marine protected areas. Operation Phakisa. Bringing different sectors together to look at decision making including marine data. This takes time.

Angola working with SABONET programme - inventory of plants - good to influence decision makers. The decision makers could establish two new conservation areas. Studies, and institutional collaboration. Traditional knowledge programme - on medicinal plants, species validated. **Strengthen the national nodes, providing IPT and other tools to make the work of the node easier**.

Two categories of end users: students and policy makers
Type of data to be mobilized: thematic data mobilization, climate change, medicinal plants, health
Capacity building for nodes to be efficient in the projects

Timelines was too short for the whole process of influencing decisions.
Projects are not restricted to using the data that they are mobilizing for that project - they can use other datasets in influencing decisions.

Some grants allocated specifically to data use to influence decisions

Tackle data mobilization to fill gaps.
Data-science-policy interface takes time. It's a long process. But the projects can take the information to the table.

National institutions don't have the capacity to process the information into finished products. Students can be helpful in the process. Build capacity in processing the information for use. This requires in depth capacity building. E.g. Masters programme

Where should we expand the BID approach?

DNA barcodes could be explored with a few projects.
Be broad in the calls. Try to decide based on the projects received.
Focus on SDGs as a thematic areas
End goal is that the nodes get more established and mobilize more data.

IPBES as a data-science-policy interface and can help show the relevance.

The projects must address data mobilization in thematic areas related to SDGs to 2 categories of end users: students and decision makers

Project coordinators should make the efforts to put the information on the table of the decision makers. Capacity building at the node level.

Foundational biodiversity information programme in SA. They engaged with the decision makers to find out the information that they needed. Encourage nodes to act as knowledge brokers. Nodes could play a role in interfacing with the decision makers.
Species that are relevant for socioeconomic outcomes.
Mobilize datasets that are near completion.

Demand-driven data. Involve the decision makers from the beginning to understand their needs.

2. How will nodes be involved in expanding the BID approach?
Is there further scope for reusing BID-developed training and materials in node-level collaborative projects and programmes?
Can nodes help approaching funders to gain support for further BID actions?
Can nodes help identifying opportunities to integrate data mobilization and data use training into other programmes?
The BID approach has relied strongly on the community of practice. How can we ensure this is scalable?

GBIF is tackling how to capitalize the approach developed in BID

BID to support regional collaboration between nodes as a means to support ongoing project activity

Need to look for other donors to support activity at the national and regional level

2. Beyond one billion

Introduction to the thematic break out group at the Global Nodes Meeting 2019

Facilitator: Chihjen Ko
Contacts at GBIFS / co-facilitators: Andrew Rodrigues (arodrigues@gbif.org), Andrea Hahn (ahahn@gbif.org)

NOTES: http://bit.ly/31tMmsT

This 1 h long session addresses the state of data mobilisation and approaches to dealing with data gaps in GBIF. Despite GBIF's position as the largest global aggregator of primary biodiversity data, data mobilisation is still predominantly opportunistic resulting in a bias towards, for example, birds, recent observation data in the Northern Hemisphere. Furthermore, the inclusion of new types of data, such as sequence or tracking data, may introduce new biases (see parallel session on data types). Nodes are the primary agents of data mobilization and as such coordinate efforts to ensure representative data coverage, and they feed back the needs of data users to the Secretariat.

Previous and ongoing activities include interacting with thematic communities eg agrobiodiversity and invasive species, to identify data needs, the "suggest a dataset" tool to capture candidate datasets for mobilization, funded regional data mobilization programs like BIFA, BID and BioDATA, enhanced data useability through issue detection and flagging and further data analysis for gridded data sets, outliers, and data density measures. Are these tools adequate for guiding mobilization, and if not, what components are missing?

Data mobilization planning within the GBIF network, under conditions of limited resources, should ideally consider the three areas of (1) data gaps (where are data missing in time, geographic and taxonomic dimensions), (2) documented needs for data (what purposes are data needed for, and why are these more important than others), and (3) information about additional available data resources (where could we get missing data from). That still leaves open the question of scale: at what level (global, national, regional, local; taxonomic granularity; data completeness and fitness for use) is this information useful to Nodes? At the joint Science and Nodes Steering Group committee discussions in February 2019, a number of recommendations were made. These were:

- A call for national checklists
- More distribution data with taxonomic resources
- Improvements in usage tracking mechanisms to identify unserved needs (search parameters used in unsuccessful search sessions)
- Integration of private sector data (environmental impact assessments)
- Data content estimator to evaluate the data richness of an area
- More guidance on data mobilization tools

Looking forward, are these recommendations comprehensive or are there additional recommendations that can be made? How can GBIFS and the global nodes community

organize and help to facilitate these targets? How do we capture data needs at different levels that drive planning and data mobilization efforts? How do we evaluate data coverage within the context of needs?

How would nodes use this type of input to support mobilisation prioritization in practice? What kind of decisions would it support? What kind of workflows do you have where you would like a prioritization to consult?

Alternative breakout group notes link: https://docs.google.com/document/d/1OfEhCfVhpi7HN9M3LA5bH9jbqM-Dy1qtHLptQ4ZxkHY/edit

Notes2: Beyond one Billion

Date: 19.10.2019

Introduction: see https://docs.google.com/document/d/1FZBBtO7Q1VRkqSlW5hUKFajeEURfOqRNwolCg0ayvrg

Group participants:

Sinh Nguyễn Văn, VN

Jean Paul Kubwimana, ARCOS

Elja-Leena Laiho, FI

Veljo Runnel, EE

Knut Anders Hovstad, NO

Jo Judge, GB

Maxim Shashkov, RU

Maxime Compremanne, BE

facilitator: Chihjen Ko, TW

GBIFS co-facilitators: Andrew Rodrigues, Andrea Hahn

Recommendations

1. Working with GBIF towards policy agreements through stakeholder/funders/ministry to make it a requirement of projects/contracts. Data management plan/budget.

2. Awareness raised for DwC to address the richness.

3. Work with GBIFS addressing taxonomic gaps, small grants to 1) identify the taxonomic gaps; 2) to connect to experts who are specialized of the targeted gap; 3) train so people are knowledgeable to the taxonomic group and share the knowledge.

Please take note of the discussion using this document. Just find a spot and start to write in chronological order. Thanks.

Introduction by the chair

What can we do to Address or to close the gap we have in biodiversity data.

View1:

1. check on existing check list and identify the gap. As a way to measure the current status
2. More distribution data with taxonomic resources
3. Improve the usage of current data e.g:
4. Integrate private sector data e.g EIA data , more guidance on data mobilization tools

How can GBIF secretariat and node community can work together to address this? How do we prioritize that?

Views of participants

Geographical gap: there countries that are lacking data (you can see that when you look on GBIF portal. Probably there is data that is not published.

proposed action:

1. contacting data holders and get the check list of project done and see where they took data see who was involved in that topic and then contact them. Eg: of a project in cyberia
2. Another way to get to those data is to discuss with funding bodies so that it becomes mandatory for any funded project to share data through GBIF.
3. Or approach research publishing journals .
4. To have a dialogue with owners in case there is a case of sensitive data.

It can work. Example of Taiwan: ***the chair to elaborate more on this example***

Taxonomic gap : *cause is*

Lack of expertise in different taxa,

Actions:

1. use citizen science so that when upload a specimen, experts can confirm
2. Measure current state and decide on what is needed.
3. Lack of National checklist for example. It is possible to extract the national checklist from GBIF and see what taxa is under represented.
4. Encourage Universities to increase the effort to teach those understudied taxa.
5. Norway has a small funding to encourage people to develop taxonomic knowledge. GBIF can also do that and make it easier to apply.
6. See a way to formulate recommendations to IUCN and partners regarding the red list of species

How can GBIF secretariat and node community work together to address this?

Data publication is not included in proposal design. If GBIF find resources to fund research project and make it mandatory to share raw data,

Policy side*: GBIF member state have to be requested to insure provision in the policy to share data*

If states acepte to become members of GBIF they should also get something. This is a kind of partnership. Like this GBIF can have a say to influence the data sharing policy??//

For those projects funded to governments (through the Ministry of Environment for example) fund disbursement should be made in accordance to whether the data were shared or not

Much effort is needed to raise the awareness of people about the DwC principles.

7. Streamline the information in the contract.

Africa Group

Agenda

1. Organization of regional meeting
2. Regional engagement and participation
3. Fundraising issue: what will be the structure of application
4. AOB: Regional coordination

1. Regional Meeting

Togo to host the regional meeting . question: How is Togo ready to host the meeting?

Togo node manager has approached the Ministry and collaborator and they agreed to host the meeting as the

They started working on the topics to be discussed. The main idea being data use, the second being the fund (node expect fund coming from the secretariat and additional fund to be raised by the ministry of higher education.

The budget was elaborated. They included also the budget to be given by the Ministry.

Requested to separate budget to know what the secretariat will offer.

Regarding the content to be discussed , togo requested to draft it and share with members so that they can exchange about it.

Proposed date of the meeting: the 2 nd week of August 2020.

Elaborate in advance the list of participants. The representative to do so.

Secretariat: there is a line provided to facilitate regional meetings . but the NSC will meet to discuss and agree on how to share the budget.

2. Regional Engagement

Engagement of different node is being done but we realised that the participation is declining. What is the problem?

1 when we send communication, there is no feedback. ???

For Example: the meeting in Cameroon: we decided to open a platform for communication. Whatsapp. For example. We shall be sending emails as usual but also whatsapp to remind people, and easier communication. Participants were requested to share their phone numbers

Ideas from the group:

Lawrence: the problem is not much on the communication but on people in charge

Normally it is related to the decision making process. Node manager have to be proactive and where needed communicate with the regional representative in case there is something he or she can not handle him/her self.

3. Data repositories

Introduction to the thematic break out group at the Global Nodes Meeting 2019

Facilitator: Anabela Plos
Contacts at GBIFS / co-facilitators: Laura Anne Russell (larussell@gbif.org), Marie Grosjean (mgrosjean@gbif.org), Tim Robertson (trobertson@gbif.org)

NOTES: http://bit.ly/GNTrepos

During this one hour session we aim to discuss the state of data repositories in your immediate network.

In the early days of GBIF, data was typically shared through connectors that exposed databases on the internet (e.g. BioCASe and DiGIR-based tools). While this approach still serves some communities well, there has been a growing trend within GBIF - and in other disciplines - to use open data repositories to archive and share data. The GBIF Integrated Publishing Toolkit (IPT) is one example of repository software that has been used by many to connect and share views of data through GBIF. Now around 10 years old, we believe it is a good time to reflect on the IPT and consider how it should evolve in the future to best serve our collective needs. This must consider what is going on in wider communities around open science data along with regional and national needs for using and providing infrastructure.

This session seeks to hear from the Nodes on their needs, experiences and challenges and therefore we encourage participants to come prepared to share their ideas. This is a broad domain, and will surely yield differences across our network. Some guiding topics may include:

- How important is the IPT for your activities?
    - Do you have specific wishes for future editions?
    - Would you like the IPT to accommodate more types of data?
    - Does running and IPT represent a significant burden (time or money) to you?
- How much does co-locating institutional data play a part in your network (e.g. sharing of IPTs)?
- Is offering a data repository for your community a significant part of your work?
    - Does this represent significant burden with e.g. data security and backup?
- Do you have a requirement to use regional or national infrastructure provided outside of the GBIF network, such as the European Open Science Cloud?
- Are you required to operate infrastructure as part of your Node mandate?
- Are you exploring open repositories such as Zenodo as a mechanism to archive and share data?
- Do you anticipate that funders will require projects to deposit data in certified repositories in the near future? (CoreTrustSeal is an example of a certification)
- Would you like the IPT to evolve into a tool that helps users shape their data into a standard format, but then deposit that into an open repository (e.g. Zenodo)?

- Does data quality control and data review fit into your data deposition processes, and how?
- Should GBIF.org evolve to offer hosted repository services for your use?

For Reference, Work programme text follows

Priority 4: Improve Data Quality

Activity 4a: Ensure data persistence

Tasks

1. Identify and verify datasets within GBIF network without current owners
2. Publish reference instances of these datasets within hosted IPTs
3. Develop processes and mechanisms for adoption of orphaned datasets by suitable agencies or experts

2019 Progress

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

2019 Participant contributions

Biodiversity Heritage Library: Make progress on adding BHL Europe data to BHL. Minimal progress has been made. Developed a plan for a rapid ingest of BHL Europe materials.

iDigBio: iDigBio partnered with the Society of Herbarium Curators to conduct a Strategic Planning for Herbaria short course. The goal was to produce a strategic plan for each represented herbarium, including vision, mission, stakeholders, strategies, goals, objectives, evaluation, and sustainability, among other things. iDigBio has been working with the community on alternative data storage solutions, such as CyVerse.

Mexico: In process, mentoring and collaboration with UNIBIO-Instituto de Biología UNAM publisher for reactivate data publish and rescue orphaned datasets.

Norway: GBIF Norway contributed to best practices guidelines and implementation of specimen-level DOIs in collaboration with the UN Food and Agriculture Organization in coordination with the GBIFS (see also activity 3a, 4b, and 5b).

2020 Work items

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

2020 Participant plans

Biodiversity Heritage Library: Implement the plan for rapid ingest of BHL Europe materials.
Canadensys: We will follow the recommendations from GBIF in order to ensure data persistence.
iDigBio: iDigBio is working to improve its data mobilization efforts and workflows, including moving towards IPT as the preferred publishing mechanism. iDigBio will continue to work with the community on alternative data storage solutions and strategies.
Naturalis Biodiversity Center: Persistent identifiers for specimen related data will be implemented in ELViS.

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.

Priority area 2

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

Activity 3e: Liaise with journals

Tasks

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

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

Iii. Integrate support for data journals into hosted IPT infrastructures and data rescue processes

2019 Progress

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

DATA REPOSITORIES - breakout session - 2019-10-19

Questions from the agenda:

- How important is the IPT for your activities?
    - Do you have specific wishes for future editions?
    - Would you like the IPT to accommodate more types of data?
    - Does running and IPT represent a significant burden (time or money) to you?
- How much does co-locating institutional data play a part in your network (e.g. sharing of IPTs)?
- Is offering a data repository for your community a significant part of your work?
    - Does this represent significant burden with e.g. data security and backup?
- Do you have a requirement to use regional or national infrastructure provided outside of the GBIF network, such as the European Open Science Cloud?
- Are you required to operate infrastructure as part of your Node mandate?
- Are you exploring open repositories such as Zenodo as a mechanism to archive and share data?
- Do you anticipate that funders will require projects to deposit data in certified repositories in the near future? (CoreTrustSeal is an example of a certification)
- Would you like the IPT to evolve into a tool that helps users shape their data into a standard format, but then deposit that into an open repository (e.g. Zenodo)?
- Does data quality control and data review fit into your data deposition processes, and how?
- Should GBIF.org evolve to offer hosted repository services for your use?

For DISSCO, no requirements to use a specific repository but requirement to use repositories that are certified. The repository has to meet some standards: how sustainable? Is it persistent? Etc.
It is not clear whether GBIF can be considered as such repository but it seems to move towards that and is in the process of certification.

The Netherlands doesn't have a national repository. If anything, it would be in the realm of the European Open Science Cloud. There are currently two repositories with different goals and it isn't clear how to deposit standardised data. Collectively, there is no message at the moment on where scientists can put data. If GBIF offers something, we should make sure that data can be picked up from this repository.

In France it is a similar situation. Many possibilities but no clear message. GBIF France is often asked if GBIF is a repository. There is no clear answer on where to deposit data (not just biodiversity data). Lots of research institutions recently chose to use Data verse.

Portugal is in a similar situation. Portugal promotes and recommends GBIF as a FAIR compliant repository. Who is in the process of getting certified? Just GBIF secretariat hosted solution or also national IPTs will be certified too, as part of the global network?
IPT so far is more a tool to format data but should it become a repository system? IPT is important also in the process of engaging and training the data publishers.

In DISSCO, two different types of storage seem to be needed:

- Long term storage -persistent, cheap storage for high volume raw data, no need for direct access
- Live evolving data - direct access needed, less need for raw data.

Long-term storage seems to be fueled by national interest in keeping the data that has been generated by national funding (e.g. digitisation programmes for specimen). Data that are used in research on an everyday basis is more an international effort and it makes more sense to store that in international repositories as Zenodo or GBIF..

In Argentina, there are a lot of repositories. All of the data produced by public funds must be public. The ministry believes that Data Cite is a good system. Everyone should have a repository available to them.
For publication, GBIF Argentina privileges Open access.
At regional level, not one solution.

The secretariat has some money set aside to improve the IPT. Any wishes?

- More data types: for example such as traits
- The IPT is a bit old-fashioned, it could be rethought to be more modern and intuitive

Species 2000 / Catalogue of Life, no use of IPT right now. There are systems allowing to convert data to Darwin Core Archive. The Darwin Core Archive itself is limited and more relational models are being developed (see the Catalogue of Life Data Package: https://github.com/Sp2000/coldp). The Catalogue of Life has data agreements with the taxonomic data custodians publishing taxonomic checklists. In the coming period, in the framework of the Catalogue of Life Plus endeavor a consultation will be carried out with the COL data custodians in moving towards the CC licenses that GBIF provides. On the long run offering repository services powered by GBIF for taxonomic checklist would become important for the Catalogue of Life.

PLAZI, compiles a lot taxonomic publication including the taxonomic treatments as Darwin Core Archive, which is the data transfer format to import the DWCA in GBIF (see datasets aka taxonomic publications https://www.gbif.org/participant/369) All specimens shared via PLAZI are linked to the original publication. PLAZI has its own checking and quality control system.
All the elements generated by PLAZI are stored in Zenodo, that is taxonomic treatments,figures and the source article with enriched metadata. Currently exploring how to store DISSCO data as well.
For each treatment we have a DOI to which can be added specific files, such as specific XML, RDF, and for which we developed a set of custom metadata. The taxonomic treatment is a subtype of Datacite publication type.
The idea is that it promotes Zenonod and make a biodiversity repository. Cooperation with Zenono will allow us to fine tune what we need with biodiversity data. It is important to be specific in what we.
The long term is to publish taxonomic concepts on GBIF and link back to the data itself - why not make the Codes adopt an article that states that a name is only available it the name is in the GBIF taxonomic backbone with a link to the treatment, that is that GBIF and COL+ have on the respective species page the data supporting a new species (example).

The LAC region, for papers, work with “la referencia” and they work with Zenodo as well.

Where the dataset is actually hosted is not a problem right now in Portugal but this might become a concern in the future. GBIF Portugal’s IPT is important in the context of international cooperation. For example with Angola. SO the question is to know whether or not it matters where the data sits. Partners don’t necessarily want files to be outside of their own countries.
Data type and other network, have made huge progress in metadata documentation but not data.
Wildest dreams is to have an IPT which are able to ingest and generate different standard format schemas and is able to handle different formats, and support mapping between these different schemas.
We would really like to have other format and compatibilities.

DISSCO is not (yet) using IPT but many partners do. It would be good to maintain the IPT in the future and developing it, as one of the factors of success of GBIF was lowering the barrier of sharing data and IPT plays an important role. DiSSCo might use IPT to share datasets from DiSSCo partners in countries that do not yet have a GBIF Node. A future IPT could have frictionless data as a supported format for serialisation, to be more flexible on using data in different schemes than just Darwin Core. Ideally it should support any scheme. We would like a less unidirectional system as well as data updates will occur more and more after publication, rather than before publication.

In the province of Mendoza (Argentina), one developer is creating custom tools.
GBIF Argentina, teaches how to use the IPT. Sampling-event data is difficult to publish right now in the IPT. It would be great to improve type and quality of data (everyone agrees).
One of the largest datasets published to GBIF from the Netherlands is a vegetation sample based dataset. This dataset cannot be retrieved appropriately from GBIF.org for scientific purposes. The business case towards the scientists is therefore not working. Traditionally, the Netherlands had several distributed IPTs. There is a trend that IPT management is however being pushed back to NLBIF. For a lot of institutes, it is causing issues to have the IPT on their own servers; often also the expertise on the IPT and DwC-A is not present and/or cannot be maintained by partner institutes. Most data within the Netherlands is now being published through the infrastructure present at Naturalis. The Netherlands wants to continue its investments into the national biodiversity informatics infrastructure, but would like to explore further how it can base this national infrastructure more on the global effort, and concentrate its national biodiversity infrastructure developments more on those priorities that are complementary to the global effort. One of the things that the Netherland would like to explore with the Secretariat is around improving sampling event.

In DISSCO, daily changes without a dataset model. Only changes are recorded and you can go to a representation of the data at a certain time with a time.-stamp rather than storing versions of datasets.

Zenodo

The ASEAN Centre for Biodiversity will continue to maintain and promote the IPT as the primary tool to organize biodiversity data across the ASEAN region. ACB has a dedicated server and it is open for everybody who wants to publish their biodiversity data.

Are images required on GBIF? Not at all. In addition to that, GBIF doesn’t store images but only exposes linked images. Only system publicly exposing images online.

In Norway, there is a government company supporting long term repository for research data. But this is not necessarily for daily use.
Some researchers are using IPTs but this is highly heterogeneous. There is a tendency to gravitate towards national IPT.
GBIF is an international repository for daily use and there are different national repo solutions for data use.
The IPT should support more data standards especially when it comes to sample, ecological and trait data. It is not a problem to publish sample plots but it is difficult for researchers to get the full data. The current model in not adapted because it is simplified.

4. Hosted Portals

Introduction to the thematic break out group at the Global Nodes Meeting 2019

Facilitator: Piotr Tykarski
Contacts at GBIFS / co-facilitators: Tim Robertson (trobertson@gbif.org), Morten Høfft (mhoefft@gbif.org)

This 1 hour long session explores the concept of fully-managed data portals operated by GBIF.

Many nodes and thematic groups successfully operate data discovery portals. These use either bespoke tools or reuse the Living Atlas codebase providing rich services tailored to local needs. While these solutions exist, there are increasing calls for GBIF to provide a fully managed solution. The motivation for this comes from different groups including:

- Those without resources to develop and operate a portal
- Those with experience of the costs involved in operating a portal who wish to direct funds to community and data mobilisation activities
- Capacity enhancement programmes looking to demonstrate the result of activities to the funders

In response to this, the GBIF 2020 work program commits to exploring the concept of running hosted portals by piloting three hosted portals targeting: A national portal, a BID Programme portal and a thematic portal providing specimen and collection discovery.

Initially, we anticipate being able to offer a portal solution offering the following:

- General content
    - Customisable home page
    - Basic dashboard
    - Ability to author content pages (About us, how to be involved)
    - News feed, blog
- Occurrence, dataset and publisher search, browse, map, download etc filtered to that data in scope for the portal (e.g. a country)
    - Observations
    - Preserved specimens
    - Material cited in taxonomic treatments
    - Barcode data (BOLD etc)
    - Sample based data
- Citation tracking
- Simple styling (define colour scheme, logos etc)
- Language support

The content would all be driven off public data available in GBIF.org and using the same backend APIs, and any data publishing effort within GBIF would appear automatically in the relevant portals. The portal owners would have responsibility for content in the news, blog and “static” pages and the ability to edit these.

As developments on GBIF.org improve and better linkages between sequences, citations, people, funders, literature etc appear these will be reflected in the portals too. This would bring familiarity to users as functions improve when they visit different sites and the data will be consistent in representation across portals.

During this session we are interested in discussion around this concept in general.

- Is it something that might be of interest to your node to apply nationally, thematically, regionally, other?
- Do you need data organised to national catalogues (e.g. a taxonomy)?
- What features and data entities would be necessary to support for a minimal product to be attractive?

5. Data in GBIF based on genetic sequences

Introduction to the thematic break out group at the Global Nodes Meeting 2019

Facilitator: David Jennings (djennings@flmnh.ufl.edu)
Contacts at GBIFS / co-facilitators: Kyle Copas (kcopas@gbif.org), Joe Miller (jmiller@gbif.org), Marie Grosjean (mgrosjean@gbif.org), Thomas Stjernegaard Jeppsen (tsjeppesen@gbif.org), Dmitry Schigel (dschigel@gbif.org)

Background

Most nodes are already familiar with the role of GBIF to integrate data covering scientific names and checklists, occurrences derived from specimens in natural history collections, and - as an increasing component - human observations, including data from citizen science projects and networks. Building on these strong foundations, GBIF is developing to accommodate data streams from additional domains and data originators - what might be called the 'data frontiers'. Three main frontiers stand out: ecological / sampling event data, remote sensing data, and **data derived from genetic sequences**. This breakout group is to focus on the genetic sequence-based data at two levels: i) global, international data streams and ii) national initiatives with existing or potential links to GBIF Nodes.

Homework

First of all, before coming to the Global Nodes Meeting, you need to know / remind yourself what is going on in GBIF with the genetic sequence data. When preparing for the meeting, please write down your questions and bring them to Leiden with you.

Minimum read

(you gain an understanding of what is happening now, but not much of details):

- Abstract of the GBIF talk at biodiversity_next https://biss.pensoft.net/article/37036/.
- FAQ (How) can I publish molecular/sequence/DNA based data to GBIF? https://www.gbif.org/faq?question=how-can-i-publish-molecular-data-to-gbif
- Three news items
    - UNITE and OTUs https://www.gbif.org/news/2LrgV5t3ZuGeU2WIymSEuk/adding-sequence-based-identifiers-to-backbone-taxonomy-reveals-dark-taxa-fungi
    - MGnify and eDNA https://www.gbif.org/news/6ewyUhBpRYammYWI2CgsM4/biodiversity-infrastructures-to-crosslink-metagenomics-and-species-occurrence-data
    - BINs from BOLD are now part of the backbone https://www.gbif.org/news/2UfGq1L6iXbSu0ElamvVlH/gbif-introduces-new-version-of-the-backbone-taxonomy
- It is also recommended that you attend Jerry Lanfear's session ST15 "Molecular biodiversity evidence in time and space: data linkages and standards" on Wednesday 11:00-12:30

Extended read

Start from the minimum read above, but here your can dig deeper into the topic and see examples of the actual data publishing:

- UNITE OTUs (SHs) are in the backbone, latest UNITE v8 is part of the latest Sep 2019 GBIF backbone. https://www.gbif.org/dataset/61a5f178-b5fb-4484-b6d8-9b129739e59d. UNITE “sequence occurrence” data coming
- BioWIDE occurrence data indexed with UNITE SH names and Latin names and more fungal datasets are expected to follow https://www.gbif.org/dataset/3b8c5ed8-b6c2-4264-ac52-a9d772d69e9f.
- INSDC data flow to GBIF set, some issues fixed, some remain / https://www.gbif.org/dataset/ad43e954-dd79-4986-ae34-9ccdbd8bf568. GBIF and ENA agreed on the technical solutions to fix this data stream (estimated November 2019).
- ‘ENA -> MGnify data flow set and working https://www.gbif.org/publisher/ab733144-7043-4e88-bd4f-fca7bf858880/metrics based on the Adapters connecting BOLD and GBIF infrastructure that synchronise data by Thomas
- BOLD BINs in the backbone see news, 80% consensus. Description in the backbone dataset. BOLD occurrence data reindexed https://www.gbif.org/dataset/040c5662-da76-4782-a48e-cdea1892d14c = public BOLD barcodes with coordinates ver. June 2019.
- Sequence ID tool allows users to get UNITE and BOLD names on sequences. Play with it with any FASTA file you have at hand, fungal or arthropod data are recommended https://www.gbif.org/tools/sequence-id.
- PLAN New adapter Mgnify-like solution will increase the N of BOLD occurrence even more and it will make it automated, regularly updated.
- PLAN More OTU libraries will be targeted for the next backbone(s), including on prokaryotes (SILVA), protists (UniEuk), nematodes, diatoms, and possibly more.
- PLAN Objective to better represent specimen entities and their links to people (e.g. ORCID ID), treatments (i.e. material cited) and sequences (e.g. BOLD) in GBIF.org 2020

BONUS

Discussing standards is not planned for the breakout group, but can warm you up for biodiversity_next. This is for standard nerds only.

- GGBN extension used currently, MIxS extention under consideration for eDNA datasets in GBIF Norway and GBIF Sweden). GGBN https://terms.tdwg.org/wiki/GGBN_Data_Standard MIxS https://press3.mcs.anl.gov/gensc/mixs

Questions for the breakout discussion

- Do you know national initiatives working with the genetic sequence data from the wild species? Examples include national barcode of life facility (BOL), national ELIXIR nodes, GGBN related project.

- Are in you in regular contact with these people,
  do you know their names, roles, location / affiliation contact information, and attitudes towards working closer with a GBIF Node?
- Can you estimate the % of species in your country that have some associated sequence data, is it sequenced using the national or external material?
- Are the national data management systems for genetic sequence data, or all such data is exposed exclusively through the international portals, such as NCBI or iBOL?
- What is going on in your country related to eDNA and DNA based research (molecular ecology) and environmental monitoring? Do you know the key research groups and agencies involved in generation and management of such data. Where is the data?
- Are the experts in your country who are familiar with managing data using molecular data standards?
- Can you recommend reference libraries of the OTUs for inclusion into the future GBIF taxonomic backbones? Please indicate i) taxon ii) name or link of the library, and iii) what is the expected impact on indexing occurrence or samping event data.
- What are the key partners GBIF and GBIF Nodes should be working with in this field, and why?
- How do you see the balance in your future plans for your Node between working with the traditional and the frontier data streams?
- Do you need support from outside your country expertise, what kind and when?
- Do you have a data mobilization plan and are sequence data part of it?
- What are the most urgent guidance and documentation needs on publishing sequence data compared to the existing documentation? 1) https://github.com/gbif/ipt/wiki/howToPublish#instructions 2) https://www.gbif.org/publishing-data 3)

We are four people in the Secretariat that actively deal with genetic sequence data:

Joe Miller - phylogenetics and phylogeographic, supervision
Marie Grosjean - data publishing solutions, GBIF helpdesk
Thomas Jeppesen - IT development
Dmitry Schigel - metabarcoding, general coordination of the topic

DATA in GBIF BASED ON GENERIC SEQUENCES

Facilitator: David Jennings

Note Taker: Michele Marcotte

GBIF is all about integration of different sets of data

Data from Genetic Sequences

Global Data Stream

Local and link to projects

Create recommendations for the Secretariat

Kyle Copas – to summarize what GBIF:

- Pilot this work
- Not talking about publishing sequence
- Taking sequence as evidence as occurrence
- Taxonomic concept – eDNA alongside with observation
- Expand the boundaries of classification system
- Use of libraries is essential
- Magnify network of bioinformatics institutes – largely from marine sampling, data stream, linean taxonomy; UNITE fungal DB to provide OTU (genus or family level); other libraries to represent OUT; as initial pilot
- National Nodes are up to – to increase the representation
- BioDivNet – Wednesday on this topic
- US example - Assessing the level of specificity – OUT – 10 species names can you tell me where they come from – pollen based – works well from Mexico – well characterized

Donald Hobern:

- GBIF – Bucket of life bins
- Improve the transparency – web services
- BOLD – tighting more with
- IBOLD with the EMBRAVE platform
- Interacting with these bins
- Data points attached to the clusters, placemat of the names and where
- Clusters are formed – of data points at one points
- Representation of the history of the clustering, how the data is understood
- GBIF – trace history – more problem with genetic data
- All of GBIF data – synonyms problem – Poa – ID 20 years ago; with genetic data we can recompute; variability to iterate effectively
- Coverage of barcode of species; 30% that is barcoded; mammals, insects, full genome sequence is the answer for plants - taxon specific
- Genome sequence of species

- UK – full genome – shorter read or barcode; cheap reproducible ways to implement; two together with full genome
- Collaboration in sourcing samples; genome sequencing to be more referenced; different groups with DNA
- Specimens + barcoding sequence together

US:

- Genbank and specimens linkage – out of 300K$ only 100K$ could be linked; Complexity of the problem

Japan:

- barcoding – distinction between species – fungal taxonomic – hidden in plant roots not ID – do have the barcoding – barcoding and specimens are linked together – presence of occurrence and use for quality control of specimens – many fungal specimens are unidentified – connect with sequencing of specimens
- Imperfect fungi – no name
- SeqProject – Confident – Methodologies

Canada: SeqDB and BioMob initiative;

Japan: European Names; cryptic species; misidentified species; linean names; correctly identified; DNA sequencing; objective data

Donald Hobern: many context; be able to use molecular or bins as proxy for taxa; realistic for diversity; methodologies or methods

US: manage data; not to develop methodologies; bins – species hypotheses

Kyle Compas:

- OTU (species based on hypotheses; this looks like; species related in this genus
- Within the UNITE system, GBIF brings the check list; accommodation of the UNITE
- Interpretation for themselves – black box; bin algorithm more intelligible; level of separation between clusters; intuitive bridge from taxonomy; one bin; specimen is the type for the bin; algorithm is transparent
- GBIF – to do BLAST – cluster; NCBI and BOLD and do the mapping; no oversee

US:

- eDNA – cat fish in the Great Lakes – News paper article – exact species in that place – hogh confidence or not
- Conference for e-DNA; methodologies are not standardized

Recommendations:

- Insist that Natural History Collections Specimens be linked to barcoding data because it allows the species identification of great value to the member countries, then if partial and full genome become available they should be integrated.
- Barcoding should be linked with well curated specimens

- It is important to note that sequences are coming from DNA samples but it is important to keep specimens and samples available in case technologies evolve and DNA needs to be redone and for reproducibility purposes. It is also important to have DNA kept in biobank.
- It was outlined that some users want to find all evidences and many users do not want to clean the data; hence developing confidence level with specific threshold (to be determined) is important
- GBIF should also work on perhaps have certified dataset established in relation to fitness for use but it was mentioned that it is very difficult to scale up
- Can GBIF develop qualifiers on molecular data based on specified criteria (that would need to be determine)?
- For occurrences and field samples particularly those on which we do e-DNA, practices should be well documents and there must be validation of methodologies should be available, it is not the role of GBIF to work on this

Sent to:

a.heughebaert@biodiversity.be

djennings@flmnh.ufl.edu

Revised version sent again.

Session 4: Data use Cases

GNM15 Lightning talk sessions

Introduction

Two plenary sessions are open to lightning talks:

- Session2: Node stories
  national/thematic/sub-regional/regional collaborations...
- Session4: Data Use cases
  Data curation, fitness-for-use, policy driven data mobilization...

With these talks, nodes managers have the opportunity to present their activities or projects to their peers. The aims of these talks are to:

- Update the Nodes Committee on the diversity of Nodes activities worldwide
- Exchange best practices
- Foster future collaboration and avoid duplication of effort
- Seed ideas for session 3 (thematic groups) and session 5 (regional groups)

Format

Presentations will take the form of lightning talks, not more than 5 minutes long.

> *"The goal of lightning talks is to articulate a topic in a quick, insightful, and clear manner. These concise and efficient talks are intended to grab the attention of the audience, convey key information, and allow for several presenters to share their ideas in a brief period of time."*[9]

Node managers are therefore kindly invited to stick the time limit, to focus on the essential and to avoid details. If time allows, questions from the assembly will be answered in plenary.

See also general guidelines

Procedure

Step 1

If you are interested to present your work, send the **title** and **abstract** of your talk and your **preferred session** before October 6th to André Heughebaert . Be aware that the number of slots will be limited by the duration of the session. You will get a confirmation if your talk is on the session agenda.

Step 2

The presentation slides, if any, will be made available to the organizers at least three days before the meeting. (Please send your presentation slides to gb26@gbif.org).

Data use cases

Lightning talks

1. Integration of molecular data into Biodiversity Atlas Sweden
2. Enhancing biodiversity data use to inform decision making in Benin and the rest of Africa
3. Expanding the workflow from scholarly published data to GBIF
4. Mobilizing sampling event marine data
5. GBIF Data Usage in China
6. Frictionless Darwin Core

---

1. Integration of molecular data into Biodiversity Atlas Sweden

Anders Telenius, Sweden

Although massively parallel sequencing methods have revolutionized the collection of biodiversity data from environmental samples, metabarcoding data are rarely accessible for re-use. To enable interpretation of fields and values not originally designed for environmental samples we will 1. provide a guide with specific pointers for metabarcoders, 2. supply a pipeline for automated processing of raw reads into denoised, taxon-annotated Amplicon Sequence variants (ASVs) and 3. provide the necessary structures for integration of ASV observations into the Swedish BioAtlas aggregating biodiversity data and making them freely available on-line. 4. Species observations in the Bioatlas are currently mapped against the GBIF taxonomic backbone which unfortunately has poor coverage of some organisms, such as procaryotes, and we thus aim to complement the GBIF taxonomic backbone with identifiers for Swedish ASVs and higher taxonomy from selected external databases.

2. Enhancing biodiversity data use to inform decision making in Benin and the rest of Africa

Jean Ganglo (Benin)

In order to overcome the challenge of limited data mobilization and data use of its national ever-growing partners, GBIF Benin organizes every year at least two workshops to train national partners in basic skills in data digitization, data curation, and data publishing. Many institutions therefore registered on GBIF site and are publishing their data to become more and more visible in the international landscape. Digital Accessible Knowledge (DAK) are therefore accumulating on the country page of Benin. In order to value those accessible data, the node manager of GBIF Benin initiated a regional master program in biodiversity informatics to enhance data use in order to inform decisions on biodiversity conservation and public health. The regional master program actually comprised many African countries such as Côte-d'Ivoire, DR Congo, Madagascar, Togo, and Benin. The first batch of the students are actually doing their research works in various fields including spatial distribution and ecological niche

modeling of animal and plant threatened species, invasive alien species, vector-borne diseases (Lassa fever, Buruli ulcer) etc. Most of the results will be released in December 2019 in form of information and useful tools to support in Benin and the rest of Africa, decision making in biodiversity conservation and sustainable uses as well as in public health.

3. Expanding the workflow from scholarly published data to GBIF

Donat Agosti (Plazi)

A huge untapped resource of biodiversity data is hidden in a corpus of 500M scholarly printed pages, continually expanding by over 1,000 journals, monographs and books. This includes the entire taxonomy including all the synonymies, data about names and specimens and increasingly explicit links to specimens and DNA data. Over the last few years, GBIF and Plazi developed a workflow to upload data liberated from taxonomic publications resulting in over 27,000 datasets. This includes among others 42,000 names that are only provided by Plazi and covering new described species near time at the moment when they are published. The special value are the links from a taxonomic name to the taxonomic treatment and linked figures, that is the data about the particular usage of the name in the publication provided by the author. Another potential value are links to specimens and DNA sequences cited in the text. These data are deposited at the Biodiversity Literature Repository (http://biolitrepo.org), where for each deposit rich metadata is added and a DataCite Digital Object Identifier (DOI) minted. Whilst GBIF is integrating this data source more closely - see e.g. the “Taxonomy” section on GBIF’s home page - progress on Plazi side will be reported in regards of adding increasingly more data.

4. Improving data availability with sampling event and extended measurement or facts: Examples from OBIS-USA

Abigail Benson (USA)

During biological sampling events, measurements are routinely collected about the event along with the biological observations. For example, the same sampling event might collect event measurements like water temperature and salinity as well as biological measurements like abundance and weight. Keeping these measurements together is important to be able to assess how species might be responding to changes in their environment and to be able to make predictions into the future. The Ocean Biogeographic Information System (OBIS) has been utilizing sampling event with extended measurement or fact extension to include both environmental measurements and biological measurements for aggregated data. As a node to OBIS, the Ocean Biogeographic Information System-USA (OBIS-USA) node has made the transition to this new data model and found it beneficial for fully expressing the data. Information that was frequently left out when aligning marine biological sampling data with occurrence core, can now be included in our submissions to OBIS and GBIF via the Integrated Publishing Toolkit. OBIS-USA has several use cases to that demonstrate how additional information has been made accessible through this data model.

5. GBIF Data Usage in China

LUO Maofang (CAS)

According to data usage statistics from GBIF Secretariat, there has been an intensive increase of peer reviewed-publications in English by Chinese authors during the past five years. In this short talk, we will briefly show the GBIF data usage in Chinese articles. Comparing with peer reviewed-publications in English, there are more publications using GBIF data in Chinese. More than 700 articles using GBIF data had been published in more than 300 journals or postgraduate dissertations, among which, the top three journals are Biodiversity Science, Acta Ecologica Sinica and Chinese Journal of Ecology.

6. Frictionless Darwin Core

André Heughebaert (Belgium)

Frictionless specifications are light weight, still they offer essential metadata description such as data types, primary or foreign keys, integrity constraints… making data packages self-explanatory and re-usable. Frictionless Darwin Core is a simple Python conversion tool transforming any DarwinCore archive into Frictionless Data Package. Complying with these specifications offer an immediate access to a bunch of Frictionless ready software. Using GBIF downloaded data or published Darwin Core archives was never so easy. Data Package libraries are available for all modern programming languages : Go, Java, Javascript, Julia, Matlab, PHP, Python, R, Ruby,... If you prefer, simpler webapps and Command Line Interfaces(CLI) are also available to manage/validate Data Package, eg Goodtables.

Session 5: Regional groups

Session5 Guidelines

Introduction

The participants will split into five discussion groups of about 10 people each.
Participants are expected to join their Regional group. For practical issues, less represented regions will be grouped together.

Discussions will focus on Regional priorities and opportunities.

Each group will have 60’ of discussion animated by a facilitator (the regional representative or its deputy). A notes keeper will be designated.

Procedure

The discussion should allow all participants to share their opinions, best practices, concerns and identified barriers with the objective of defining a consensual approach for further improvements around the group subject. The facilitator will briefly introduce the subject(5’) and suggest some initial questions to kickoff the debate.
The facilitator will make sure that all participants have the opportunity to speak and that the discussion stay on the subject. All participants are equal and should behave respectfully.
The facilitator and the notes keeper are strongly encouraged to share their notes and presentation in the meeting documents repository.

When the groups will reconvene in plenary, the notes keeper, or another representative, will summarize the group discussion and present their outcomes.This will feed the **conclusions** session.

Expected outcomes

Group are invited to come with set of realistic **recommendations** and/or suggested **actions** to improve Nodes efficiency in the subject selected by the group. These outcomes might have a local, regional or global scope.

- Recommendations can be addressed to Nodes, Secretariat, Nodes Committee or any other GBIF standing Committee
- Actions will eventually be described with responsible, timeline, credible budget and possible funding

GNM15 - Regional Group Asia

Status & observation
1. Fluctuating participation:
   - 5 yr limit to become Voting.
   - India, Indonesia, Pakistan, The Philippines once joined but dropped to observer status.
   - Japan in its fifth year.
   - Viet Nam joined in 2018.
   - Cambodia is observing and keen.
   - Malaysia, recently published a [dataset] (https://doi.org/10.15468/my24sh)

2. BIFA has funded crucial data activities in the region in recent years
   - BIFA getting new contribution from new countries
   - Enriched data coverage (geographically, taxonomically)
   - Improved institutional participation in the region
   - Revitalised somewhat dormant node activities
   - Provided opportunities to engage ministries' interests

3. Recent inspirational growth in Siberia
   - Dmitry Schigel, Russian speaking
   - Norwegian fund for Asia (Dag Endresen)
   - Enthusiastic group, publishing data, and now observing Governing Board

4. Funding
   - BID, extend to 2023
   - BIFA, 2020 secured

Year 2020 regional events
1. Even year, funding for regional meeting available from core fund, to be announced.
2. Ideally to hold between May and Sep, preferably earlier and friendlier reporting.
3. To colocate with BIFA training, save fund.

Focusing on the region
1. GBIF 20yr review supports Asian focus
    - Identify the coordination hub (people, location)

2. We need:
    - A Dmitry-ious role in Asia
    - Engagement of peers from other regions to co-invest Asia
    - A task force (with nodes)
        1. To understand the region better
            - Relevant researchers and their research topics
            - Analyse data gaps: e.g. taxonomic, species under-recorded, etc.
            - Gap of expertise/education

        2. To define strategies and models for advocacy and outreach. An idea being: Recognising it takes time to bring a govt on board, an idea is to keep utilising BIFA fund or alike to build up appealing results of activity/data to a volume, as funders pleased, so that it

supports conversation at the Participation level

3. To identify and utilise existing activities with other regions
4. To utilise/build on interests from the region, e.g. Malaysia
5. To identify funding opportunities, e.g. Norwegian BioDATA fund
6. To promote policy requirements for research data to deposit in designated place
    - Dealing with sensitive data
    - Data from private sector

7. To identify unpublished data
8. To improve the quality of data from citizen science

<!--stackedit_data:
eyJoaXN0b3J5IjpbLTE0MzcwMTQxMTMsNzMwOTk4MTE2XX0=
-->

European Nodes Meeting at GB26 in Leiden 19 October 2019

EC opportunities for EU Nodes:

Germany: Explore **EU** opportunities for funding regional GBIF activities/network can be visited again.
Portugal: **DiSSCo** will cost additional money for the country - need to be given consideration if linking GBIF and DiSSCo. EU Commission is expressing a positive attitude (BID etc) to GBIF and could be an opportunity to follow.
Netherlands: DiSSCo is complementary to GBIF, not competing.
Poland: Encouraged by national level to join DiSSCo and will try to link initiatives. Afraid of **confusing national funding** levels [*with respect to overlapping scope and tasks*].
Ireland: This discussion is similar to the discussions we have had before. The former "*threat*" was **LifeWatch**. A key tool we need is the value proposition. Examples of tools: Clear soundbites, clear use cases. A clear message of how GBIF is different from DiSSCo etc.
Spain: **EOSC** (European Open Science Cloud) -- country-level interests in Spain to link EOSC and GBIF nationally. EOSC is not yet providing any services - at the moment only making an inventory of projects. Asked to fill in a large set of forms to describe services provided.
France: France has had contact with EOSC.
Portugal: Yes, EOSC is still only an inventory of services - however, EOSC will link relevant services. **Important for EU Nodes to be represented in EOSC collectively together** - not each country node separately. GBIF Secretariat cannot be expected to be the body to represent us in EOSC, we need to form a collective solution driven from our Nodes.
Ireland: Do we have the capacity to make a clear statement of how we want to approach EOSC?
Portugal: Most EU Nodes (70%) that are funded through the Ministry of Science will soon meet pressure [on the national level] to link with EOSC.
Norway: We all need [to communicate] our strengths, but also need to coordinate how we collectively communicate our strengths. Proposed to ask for someone to volunteer to write a value statement [for the entire European regional level in GBIF].
Kyle, GBIFS: **EEA** is a critical participant that could be the bridge to EC opportunities. EEA is an important channel to activate in this picture. At the EU Nodes meeting in Oslo a draft regional strategy was drafted. There is a gap [of expectations] between GBIF Secretariat located in Europe -- but there is no staff at GBIFS [allocated] to coordinate activities in Europe. There will be a need to address funding if this function should be established. Signals on this topic from the regional level will be important guidance for the work at the GBIFS.
France: We have discussed previously if we need a person [in Europe] with funded work time to monitor calls and coordinate responses to funding opportunities.
Ireland: The case for an Asia office was made because of GBIFS resources to deliver relevant/appropriate services in Asia and language barriers.

Norway: Focus on the appropriate channels to communicate through. Could address the INSPIRE metadata standards [as a common topic/service].
Portugal: Has been looking at who is funding EU Nodes -- 18 by the science ministry [EU Nodes includes 19 voting and 2 accosiate]. **Approaching the science funding opportunities is a more important focus than the EEA/environment ministry side**.
Nederlands: EEA is a different organization [from GBIF], but important for us to be relevant for -- eg Names and CoL issues.
Germany: Who to address: **ESFRI INFRASTRUCTURE** programs. The environment side has operational objectives to meet and deliver -- and is less available to be involved in funding research infrastructures (ESFRI). EUROPEANA cultural side was trying to build services.
Poland: Previously we tried to coordinate a **COST** action on behalf of the EU Nodes collectively. Will such a proposal work for us. Each country node is very different with very different needs [, mandate and capacity]. Such a proposal could contribute to making us more alike -- more as a [coherent] service that we can deliver across Europe.
Netherland: COST is not a good call for this purpose.
Germany: **Living Atlas** was argued as a unifying solution to approach EU funding as a cross EU Nodes service/project.
Portugal: COST is appropriate for a strong established community and we are not yet mature to approach such a mechanism. Only asking for money is not the solution.
Ireland: **We need to make a case for what we want to achieve**. Making an IT tool is not necessarily the thing.
Spain: To achieve this we need a professional representative/coordinator. Maybe a professional to help us with identifying the approach.
Germany: The need for a Living Atlas of Australia mirrors very many of the same needs as we have in Europe. And many of the Living Atlas portals already implemented are already proof of concept.
Portugal: Need to look at ESFRI mechanisms and objectives. Expanding our regional GBIF collaboration eastwards might be a key approach.
Poland: Eastwards and southwards (Balkan, Greece etc). We could manage to explain a scientific need for this objective in a COST action proposal. Other suggestions for EU program streams to approach.
Norway: Is not the lack of participating European countries a gap that the EU would respond to. Broader coverage of membership could deliver other things than we can do with these gaps.
GBIFS, Kyle: Connecting to certified **data repositories** endorsed by EU bodies could be a topic with need. Sharing data between thematic infrastructures. Mandatory data archiving in data repositories - and extracting biodiversity data from such archived data packages is an opportunity.
Norway: Data flow mechanism for eDNA through GBIF could provide an opportunity.
Ireland: Approach a professional to assess opportunities.
Plazi: **How to make data accessible could be an opportunity**. Make a system with defined metadata and protocols. Experience with DiSSCo is that is **easy to deposit data, but much harder to make use of that data**.
France: https://biodiversity.europa.eu/

Next year regional meeting

Luxembourg will explore opportunities to host the venue in 2020.
Poland will try in two years for 2021.

Minutes captured and edited by deputy regional representative for Europe (Dag Endresen)

Global Nodes Meeting
Leiden - 19 October 2019

LatinAmerican and The Caribbean Regional Meeting

Due to the low attendance of the nodes of the region and with the particular situations of the nodes present, the meeting focused on the description of the past activities of the nodes in the region, doubts and possible strategies for 2020.

The following topics were discussed:

- NSG organization and node committee
- Active nodes in the region
- Regional meeting 2020
- CESP granted
- BID Caribbean and forms of participation
- Activities to bring new nodes to the GBIF network

Organization of the NSG and committee of nodes: questions were presented and answered regarding the conformation of the NSG, the results of the last election, obligations, activities and time consumed by the regional representatives. In addition, the modality of the meetings involving node managers (regional nodes and global meetings) was described.

Active nodes in the region: the situations of the nodes of the region were described (voting, participant, future nodes)

Regional Meeting 2020: Although there were offers from Colombia, Ecuador, Mexico and Brazil, the decision on which node will host the meeting has not yet been made. Due to the commitments already acquired at the regional level (CESP), the date of the same must be between May and September 2020. More information is expected regarding the list of participants that the Secretariat expects and the amount of funds allocated for the meeting (and thus be able to evaluate the amounts of cofinancing).

The need identified by GBIF Chile for a replication (after the translation of the materials) of the training given at the global node meeting was indicated.

CESP granted: Questions about the CESP granted in the region were described and answered, from 2015 to 2019. Even when asked about the need / desire to identify a topic for the next CESP2020 call, no topic was identified.

BID Caribbean: the program, the 3 meetings held, participating nodes, problems and possibilities of participation of other nodes (mentor, coach) were described. The geographical extension for the second edition of the IDB was reported.

Activities to bring new nodes: Anabela Plos is in communication with Jaime Rodríguez to bring Bolivia to the GBIF network, since although we were in contact with Miguel Fernández (NatureServe), there was no progress of any kind in the formation of a regional node The

contact with José Luis Cartés and Hugo del Castillo (Guyra Paraguay Foundation) was also started to bring the GBIF network to Paraguay.

A workshop was held in David (Panama) during the month of July, with 40 participants, both from the Autonomous University of Chiriquí and the University of Panama and staff of the Ministry of Environment, PNUD and Ciudad del Saber (City of Knowledge). The trainers were Leonardo Buitrago and Ricardo Ortiz Gallego (SiB Colombia) and Anabela Plos (GBIF Argentina). A replica of the same course is planned for 2020 in Panama City.

Global Nodes Meeting
Leiden - 19 October 2019

North American Regional Meeting:

Node managers for the region meet each year, not every other year, in order to reinvigorate the region and develop regional projects.
Plans for next meetings:

- Digital Data Conference (Bloomington, Indiana, US) on 1-3 June 2020
- Biodiversity Summit (Alexandria, Virginia, US) on 20-25 September 2020

Best to do it during Digital Data Conference this year because it will give us time to report back to the NSG prior to the GB
This year, GBIF has a budget for the organisation of the regional meeting (more information to follow after NSG meeting and budget validation)

Extending the call for BID:

The region thinks that the extension of the call to other areas is a good idea, especially for Africa and South America.
The region is interested in participating in the new projects from this new call.

Regional offices:

The idea is coming from recommendation in the 20 years review of GBIF.
Regional offices would be built in order to encourage participation in some regions (idea of 'joint lab'). It's a good way to unify a region and to leverage co-funding (access to new source of funding).

Discussion around a North American portal:

We need to share the information and to develop a coordinated way to work together.
There are many different pieces of the puzzle between our countries/organizations, but we share a continent and we need a way to combine information. We want the same representation between our different countries/organizations but there are some issues, like the fact that the geospatial standards are not the same in the US and Canada.
The answer might come from GBIF and the development of a regional portal, with some flexibility. This regional portal will not answer all of our needs, and we will have to add some information (e.g., counties).
One of the solutions is to provide value-added datasets to GBIF (not only raw data), but how can GBIF include this type of data?

Recommendations:

- We see a great value in a region facilitator (GBIF office in different regions), but it needs to be someone dedicated to the task (not one of the regional representatives or node managers). We are conscious that North America will not be the first region to have a GBIF office, so we will need to discuss that topic again once one (or more) regions have gained experience with that project.
- The North American meeting will be held in 2020 during the Digital Data Conference, and it will be a good opportunity to offer some training about node management, based on the recent Global Node Training.

Session 6: Conclusions

Nodes Strategy 2020-2021

DRAFT

Rationale

During the 15th Global Nodes meeting, the Nodes Committee discussed Thematic and Regional groups. The identified priorities were refined by the NSG to prepare a Nodes Strategic Plan for the two coming years. This plan will guide the global efforts of the Nodes until 2022 when the new GBIF Strategy will be in place.
These strategic objectives will be developed by the Nodes Committee with the support of the Secretariat under the supervision of the NSG. Regional reports (in 2020 and 2021) will be used to show progress against these strategic objectives. During our next Global Nodes meeting (at GB28, by the end of 2021) we will assess the progress towards these strategic objectives. Different regions might have different approaches and activities to reach the strategic objectives. All objectives are of equal importance.

The Objectives

The Nodes strategic objectives for 2020-2021 are directly linked to GBIF Strategic priorities:

1. Investigate Regional coordination (Priority 1: Empower Global Network)
2. Improve data relevance (Priority 3: Fill Data Gaps, Priority 4: Improve Data QualityPriority 5: Deliver Relevant Data)
3. Explore new data types (Priority 3: Fill Data Gaps)
4. Lower technical threshold (Priority 2: Enhance Biodiversity Information Infrastructure)
5. Strengthen our network (Priority 1: Empower Global Network)
6. Redefine participation (Priority 1: Empower Global Network)

1. Investigate Regional coordination

Until now, GBIF regionalisation mostly rely on voluntary efforts of Nodes Managers and Regional Representatives to the NSG. Most regions expressed the desire to go a step further and investigate a more structural coordination of Regional nodes.
Nodes are encouraged to :

- Investigate possible forms of regional coordination (as already done by Africa)
- Define possible tasks under this coordination
- Assess the implications in terms of governance and relation with the Secretariat
- Identify potential alignment with other regional initiatives
- Suggest possible funding scheme

2. Improve data relevance

Some biodiversity data are more relevant to science and therefore decision making than others. Collections specimens and long-term monitoring surveys are more relevant than

opportunistic unvalidated observations. Nodes should focus their activities and affect their resources on what matters the most : publishing complete and relevant data.

3. Explore new data types

Occurrences data and Checklists are, and will remain, central to our data mobilization activities. However, nodes often encounter other kinds of emerging data such as: sampling events, abundance data, camera traps, DNA sequences,... Publishing these new data comes with challenges as standards and tools probably need to adapt. Nodes will report on these new data types as well as DwC evolutions and/or extensions necessary to publish them.

4. Lower technical threshold

We have some wonderful tools for data publication, storage and visualisation.
However, a demand for lite weighted solutions lowering technical barriers on both sides of the pipeline (data publishers and users) is a requirement we should all addressed.
While it is clear that such new tools will be developed and maintained by the community at large, the nodes are in a good position to gather requirements, design mockup and prototype such low tech, simplified solutions of hosted portals and data repositories.

5. Strengthen our network

All nodes have limited resources. Technical resources can easily be solved through cloud based solutions. Human resources, i.e. a skilled and stable Node team, are more difficult to address. Nodes shall seek to improve their human resources with an ideal team of 4 FTEs per country node. A Node team would typically include a Node Manager, an IT, a data officer and an administrative. Recognizing that this is a combined responsibility with the Heads of delegation, Node managers will report on progress to establish this ideal team.

6. Redefine participation

The current participation status (of countries and organizations) does not always reflect the level of effort in data publication, in data use and in support to community. Nodes should reflect on that and suggest ways to recognize and expose the value of national/themaic efforts beyond the formal official/financial participation to GBIF.