



GBIF

Global Biodiversity
Information Facility

SECRETARIAT UPDATE

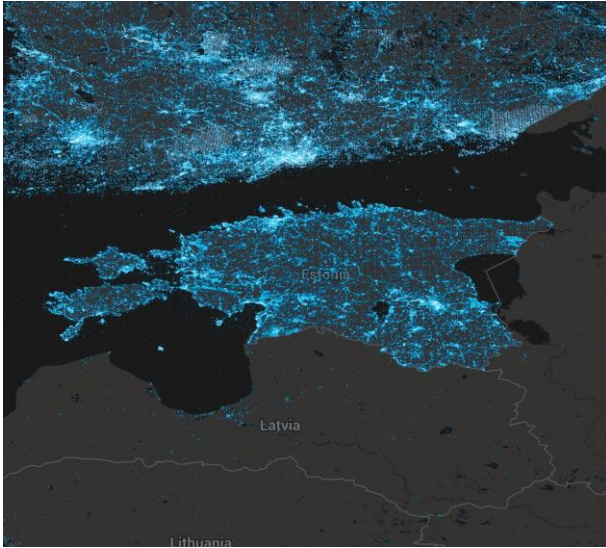
North American Nodes Regional Meeting

Donald Hobern, GBIF Executive Secretary

Berkeley, California

3 June 2018

OVERVIEW



- Refresher on GBIF 2017-2021 Strategic Plan
- Summary of current activities on 2018 Work Programme, with **highlights of key initiatives**

Detailed, line-by-line set of background slides available at

<http://bit.ly/gbif-wp18q2>

2Q 2018 IT DEVELOPMENT PRIORITIES

- Translation for key UI elements of GBIF.org
 - labels, headers, and some system messages.
- Network pages:
 - initial implementation to show multi-publisher networks
- Improve presentation of sampling event data
- Country Reports re-implemented:
 - on-demand feature and more languages
- DataONE:
 - publish GBIF snapshots and finalize API compliance test
- GBIF Pipelines: evaluate Apache Beam implementation
- IPT: security patches (requested by the Smithsonian IT team)
- Occurrence search/downloads API update:
 - support SAMPLING_PROTOCOL, NETWORK_KEY, EVENT_ID and PARENT_EVENT_ID.
- “Include” OTUs in the GBIF Backbone
- Upgrade the Hadoop Cluster to a newer version
- New quarterly roadmaps for informatics & data products teams in CPH
- Will be validated with Science Committee
- Aim to provide ongoing overview on GBIF.org

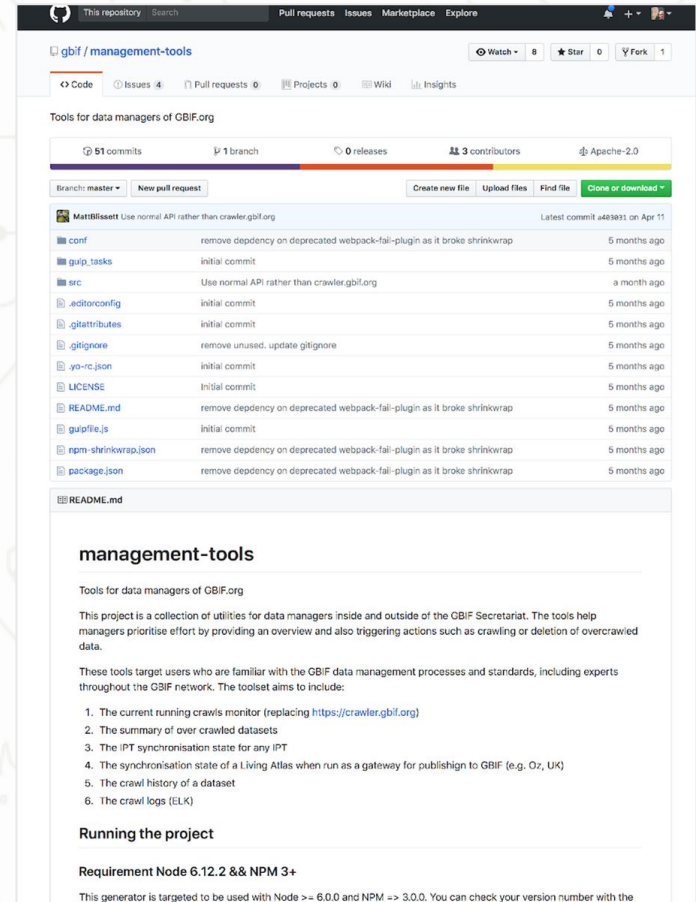
1

EMPOWER GLOBAL NETWORK

1a. Focus on people

Management tools

- Collection of utilities for data managers familiar with the GBIF data management processes and standards
- Trigger actions like crawling, IPT synchronization, and deletion of overcrawled data

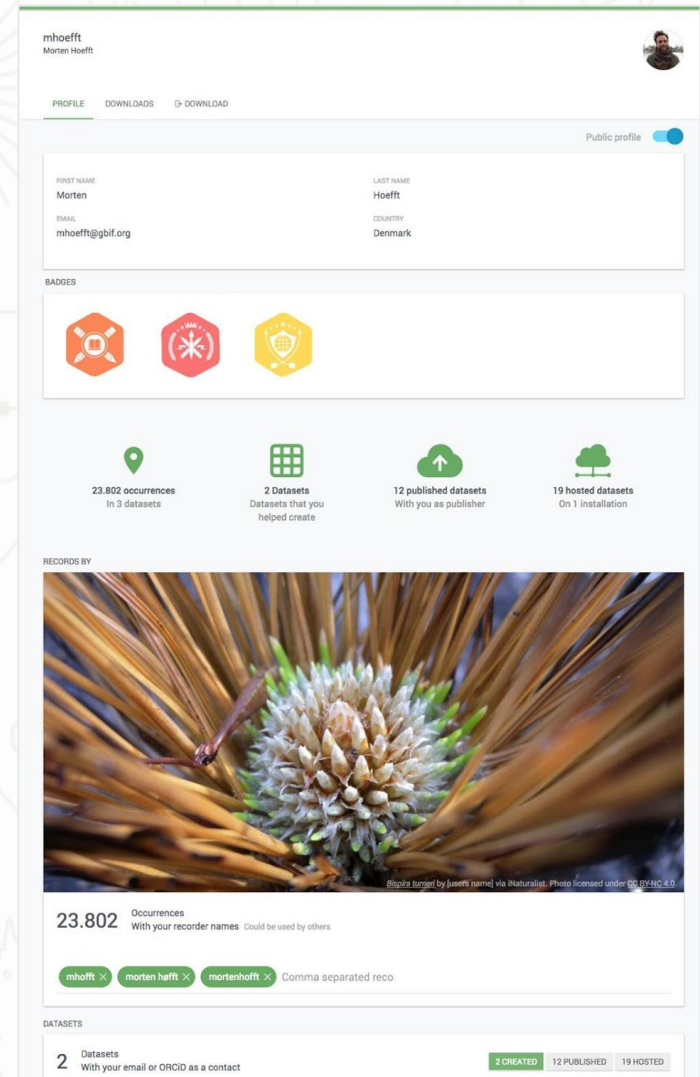


1 EMPOWER GLOBAL NETWORK

1a. Focus on people

Personalization

- Preparing expanded user profile
- Leverage ORCIDs to link opt-in users (**esp. from nodes**) with datasets (as contributor, publisher or host), badges, research papers, occurrences et al.
- Builds on implementation of 'trusted contacts' for management tools



1 EMPOWER GLOBAL NETWORK

1a. Focus on people

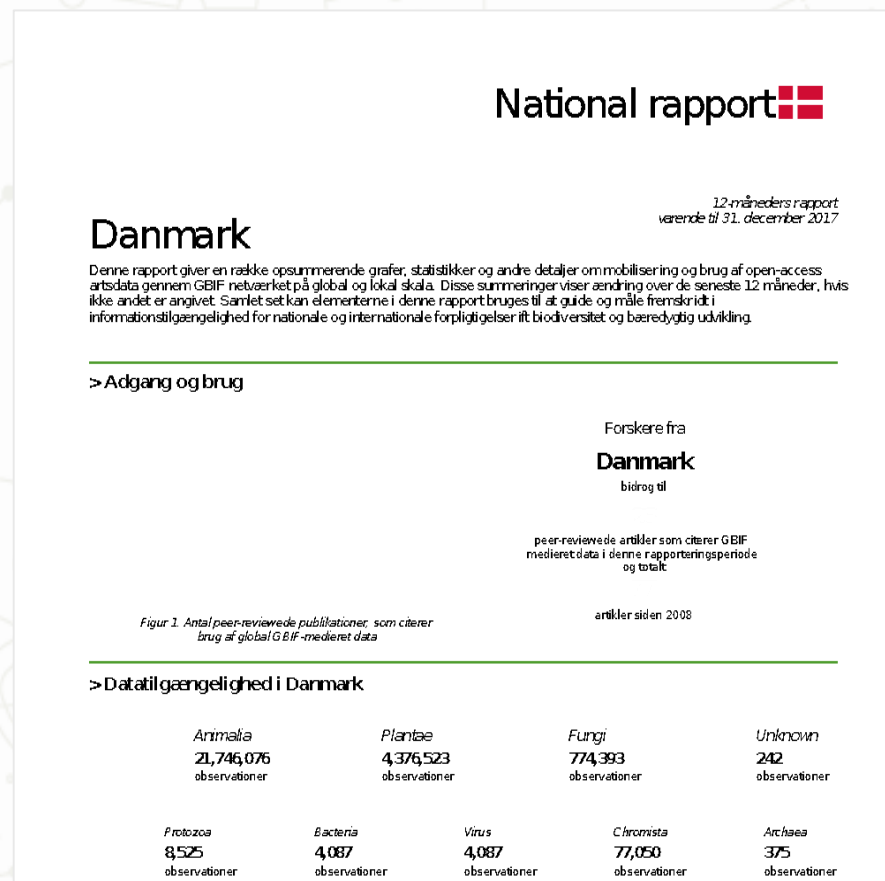
GBIF Community Forum

GBIF						Sign Up	Log In		
all categories ▾ Latest Top Categories									
Topic	Category	Users	Replies	Views	Activity				
Welcome to the new GBIF community forum! This site is for discussing any matters relating to GBIF, including sharing experiences and advice of how to use GBIF-mediated data in research, etc. The first paragraph of this pinned topic will be visible as a welcome... read more			1	807	Apr 20				
Why don't African countries feature in the top countries using data from GBIF?	Data Use		5	40	18m				
Publishers vs. citations	Data Publishing		2	7	1h				
There are 10k downloads per month on GBIF	Data Use		4	12	2d				
Citation of the week	Data Use		6	37	6d				
This just in: Amazonia is the primary source of Neotropical biodiversity	Data Use		1	23	7d				
Editing text on a publisher page	Data Publishing		5	48	14d				
First stop for new data users	Data Use		1	47	24d				

1 EMPOWER GLOBAL NETWORK

1e. Expand national participation

Support for multilingual content on GBIF.org



2

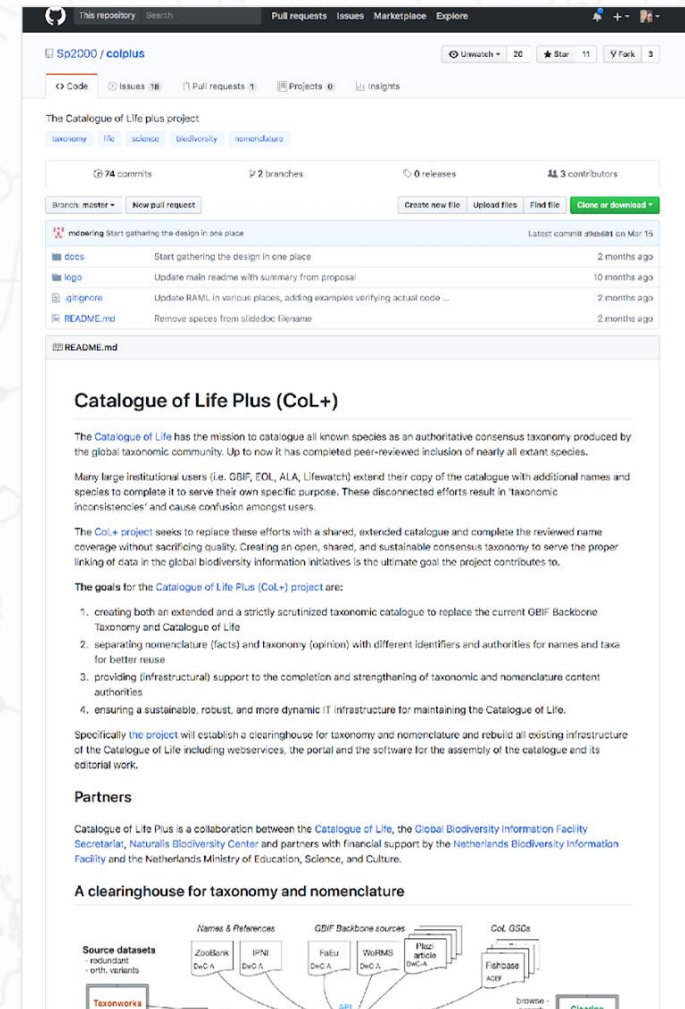
ENHANCE INFRASTRUCTURE

2b. Deliver names infrastructure

Support Catalogue of Life Plus

New COL+ infrastructure

- includes GBIF Backbone Taxonomy
- separates names & taxonomy
- provides avenues for (infrastructural) support to content authorities



<https://github.com/Sp2000/colplus>

3

FILL DATA GAPS

3a. Identify priority gaps

New data analyst & data administrator in CPH

- **John Waller** and **Marie Grosjean** joined data products team on 3 April
- Will (*among other things*) synthesize available tools for visualizing data gaps to support mobilization targets



3

FILL DATA GAPS

3b. Expand data streams

Incorporating OTUs in the GBIF backbone

- Pilot project with UNITE to integrate Operational Taxonomic Units for sequence-based 'dark matter fungi' occurrences
- Builds previous work by GBIF Norway to link to NorBOL sequences

Kingdom	Fungi
Phylum	Basidiomycota
Class	Agaricomycetes
Order	Thelephorales
Family	Thelephoraceae
Genus	<i>Tomentella</i> Pers. ex Pat.
Species	<i>Tomentella bryophila</i> (Pers.) M.J. Larsen, 1974
Synonyms	<ul style="list-style-type: none">= <i>Botrytis bryophila</i> (Pers.) Sacc., 1885= <i>Hypochnus obscuratus</i> P. Karst., 1896= <i>Hypochnus pallidofulvus</i> (Peck) Burt, 1926= <i>Hypochnus subferrugineus</i> Burt, 1916= SH002466.07FU= SH002471.07FU= SH002492.07FU= SH002497.07FU= SH002506.07FU= SH002549.07FU= SH002611.07FU

4

IMPROVE DATA QUALITY

4a. Ensure data persistence

Increasing communication with leading publishers to

- Improve recommendations on **data citation practices**
- Recognize GBIF as a **network of trusted data repositories**
- Increase recognition of GBIF network data as **FAIR data** through RDA, COPDESS, AGU, EGU
- Seek CoreTrustSeal Certification



Gbif.es



Gbif.es / News and Events / Adoption of 'orphaned' Spanish datasets completed

Adoption of 'orphaned' Spanish datasets completed

Last year, the GBIF Secretariat requested the assistance of national nodes in the network to help rescue a collection of 'orphaned' datasets. These datasets no longer appeared to be actively curated or managed, suggesting that their original data hosts no longer had the capacity to host them. In some cases, data hosts have even lost datasets, for example, due to server failure, leaving no other backup than the versions published through GBIF.org.

The [2018 GBIF implementation plan](#) called to have all datasets claimed, curated and re-hosted (activity 4a), and we have recently fulfilled this goal by restoring **94 orphaned datasets from Spanish institutions and hosting them on the GBIF.ES IPT**. Adoption of these datasets facilitates curated updates and corrections and also ensures the ability to reproduce research that is based on these datasets.

We would like to acknowledge the help made by **Lydia de la Cruz** in this regard – former student of the Master in Biodiversity in Tropical Areas and its Conservation, organized by the Menéndez Pelayo International University (UIMP) and the Higher Research Council Scientific (CSIC). Lydia has done the master's practice in the Coordination Unit of GBIF.ES. Thanks to her time and energy invested in this effort and the GBIF support we have adopted a bigger set of orphan datasets than any other GBIF node. More news related to this activity coming very soon.

Links of interest

<https://github.com/gbif/watchdog/wiki/AdoptionPlan#phase-1-2017>

https://www.gbif.org/dataset/search?q=orphan&publishing_country=ES

<https://www.gbif.es/en/proceso-de-migracion-de-las-bases-de-datos-publicadas-a-traves-de-gbif-es/>

4

IMPROVE DATA QUALITY

4b. Assess data quality

- Rebuild of the ingestion pipeline to provide consistent data processing (parsing, interpretation, quality flagging)
- Will align code base of GBIF, Living Atlases—and others?

README.md

Pipelines for data processing and indexing

Status: early development

Vision: Consistent data processing pipelines (parsing, interpretation and quality flagging) for use in GBIF, the Living Atlases project and beyond. Built to scale from laptop to GBIF volumes. Deployable on JVM, Spark, Google Cloud, .

Architecture

The project provides vanilla JVM-based parsing and interpretation libraries, and pipelines for indexing into SOLR and ElasticSearch, built using Apache Beam.

Apache Beam provides a high level abstraction framework ideal for this purpose with the ability to deploy across target environments (e.g. Spark, local JVM) and with many built in connectors for e.g. HBase, SOLR, ElasticSearch etc.

Ingress

Ingress is from Darwin Core Archive (zip files of one or more "CSV"s) or ABCD Archives (compressed XML) only[1]. During ingress data is converted from it's native format and stored as Avro files containing Darwin Core compliant data.

This is depicted below:

```
graph LR; DwC-A[DwC-A] -- Read --> PC[PCollection of verbatim data]; PC -- Persist --> verbatim[verbatim.avro];
```

Ensure all records have occurrenceID

5

DELIVER RELEVANT DATA

5a. Engage academia

Secretariat and Science Committee focused on widening the reach of

- **Young Researchers Award – deadline 15 June**
- **Ebbe Nielsen Challenge – deadline 5 September**



5

DELIVER RELEVANT DATA

5b. Support biodiversity assessment

MOU signed with IPBES

- Expected to emphasize **national-level links** between nodes and IPBES focal points
- Seeking best examples of **node guidance**, i.e. Belgium, Colombia
- **Integrating GBIF-relevant data networks** into UNEP-WCMC project on developing national biodiversity and ecosystem assessments for selected countries (CO, CM, VN, ET, BH et al.)
- Proactive global and national engagement prior to launch of **2019 I&AS assessment**



#GBIF1billion

#GBIF1billion

[Get data](#)[Share](#)[Tools](#)[Inside GBIF](#)[Login](#)

GBIF | Global Biodiversity Information Facility

Free and open access to biodiversity data

[OCCURRENCES](#)[SPECIES](#)[DATASETS](#)[PUBLISHERS](#)[RESOURCES](#)[WHAT IS GBIF?](#)[ABOUT GBIF ESTONIA](#)

Sap-sucking slug (*Elysia viridis*) by v_s_ via iNaturalist. Photo licensed under CC BY-NC 4.0.

Occurrence records

981,456,243

Datasets

38,947

Publishing institutions

1,181

Species

Learn more about the number of species covered by data in GBIF.org.



2018 GBIF Ebbe Nielsen Challenge seeks open-data innovations for biodiversity

9 May 2018



Port Stephens, Australia - a sea hare haven?

4 May 2018



The world's newest country becomes GBIF's newest Participant

18 April 2018



Call for nominations opens for 2018 GBIF Young Researchers Award

Deadline: 15 June 2018

#GBIF1billion

[Get data](#)[Share](#)[Tools](#)[Inside GBIF](#)[Login](#)

#GBIF1billion

Resources for the GBIF network to use in association with the upcoming big data milestone

The GBIF Secretariat is pleased to provide the Participants and nodes of the GBIF network the following resources to assist them in marking and celebrating the achievement of 1 billion species occurrence records.

We currently expect that GBIF.org will reach 1 billion records during the first half, possibly even the first quarter, of 2018. However, the communications team in Copenhagen will prepare materials and conduct media outreach in advance of reaching that milestone.

Comments? Questions? Interest in coordinating your own activities marking #GBIF1billion?

Write to the [communications team](#)

#GBIF1billion campaign goals

- Build awareness in new audiences
- Align existing perceptions with current status and achievements
- Promote network innovations in the mobilization, delivery, use and citation of open data
- Emphasize impacts on scientific research and policy

Social media

**BIODIVERSITY
OPEN DATA
AMBASSADORS**

Suggested process draft webpage & form

Title
Firstname*
Lastname*
E-mail
ORCID*
Organisation*
Country*
ZIP code
City
Street address
GBIF experience
Other relevant experiences
Motivation story 100 words on you and biodiversity data, importance, steps
Keywords. Professional interests.
Languages
Your next and recurrent data advocacy opportunities

Yes, I agree

Agree with key principles

Be ready to direct experience, position, time, role, and other resources to advocate for best practice in biodiversity data.

- Findable, accessible, interoperable, and reusable
- Fully and openly shared on the least restrictive basis possible
- Available with minimum time delay and free of charge
- Require stewardship, citation, authenticity, and quality control
- Recognized as fundamental research products by data publishers, users, authors, reviewers, editors, curators and funders through education and advocacy effort from experts like you

Basic **eligibility** screening at GBIF Secretariat

Approval by the country Node – GBIF participant countries
Nodes Steering Group / GBIF Science Committee

Names and affiliations **visible at GBIF.org**

Welcome letter, links and **resources** sent

Ambassadors **present and promote** best practice
as they see fit / responding to the GBIF Secretariat invitations

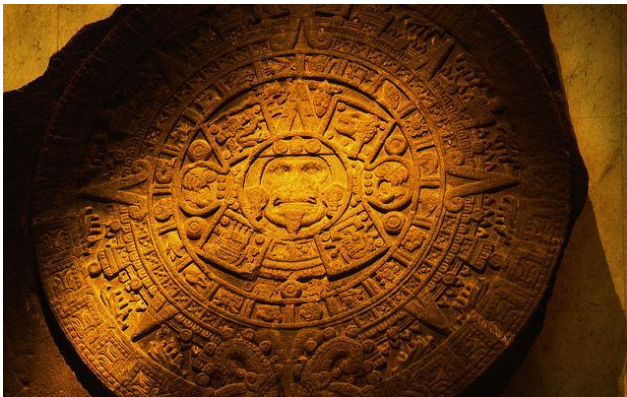
Intensity in ambassador's **control**; no reporting, specific Ambassador forum

Timeline

<i>2016, GBIF S</i>	<i>Original idea by Olaf Banki</i>
<i>Sep 2017, Helsinki Meeting</i>	<i>Presentation & discussion, Global Nodes</i>
<i>Feb 2018, GBIF S</i>	<i>Refining the concept with SC and NSG</i>
<i>Mar 2018</i>	<i>Minutes closed, action plan developed</i>
<i>Apr 2018</i>	<i>Presentation of action plan to SC and NSG, final tweaks, approval</i>
<i>Apr-Sep 2018</i>	<i>Implementation, first steps</i>

- **Webpage: definition, principles, how it works**
- **Webform**
- **First nominations and names**
- **Resources (the kit)**

<i>Oct 2018, Kilkenny</i>	<i>Presentation of progress at GB25</i>
<i>Nov-Dec 2018</i>	<i>Refinement, running the program</i>
<i>2019+</i>	<i>Rolling smoothly</i>



Calendario Azteca tallado en piedra
www.flickr.com/photos/edwinpoon_gz/6791267050

Information & communication toolkit

resources webpage



- Definition
- Criteria
- Application form
- List of ambassadors
- Useful GBIF webpages, always at hand:
dataset classes, citation guidelines, R and Python, library
- GBIF user guide
- GBIF communication strategy
- GBIF guide for Nodes -> Handbook
- GBIF monthly slides
- Live or offline support with difficult questions through distributed helpdesk
- GBIF for science slides (1-, 3-, 10- slide versions and full set of 15 slides)
including advice on access, data quality, post-processing, citation
and on data publishing
- Posters and brochures: layout files to print locally
- Communication space, GBIF forum

Legend ● ● ●
to use

= not started / in progress / ready

NODES

1 EMPOWER GLOBAL NETWORK

1c. Equip Participant nodes

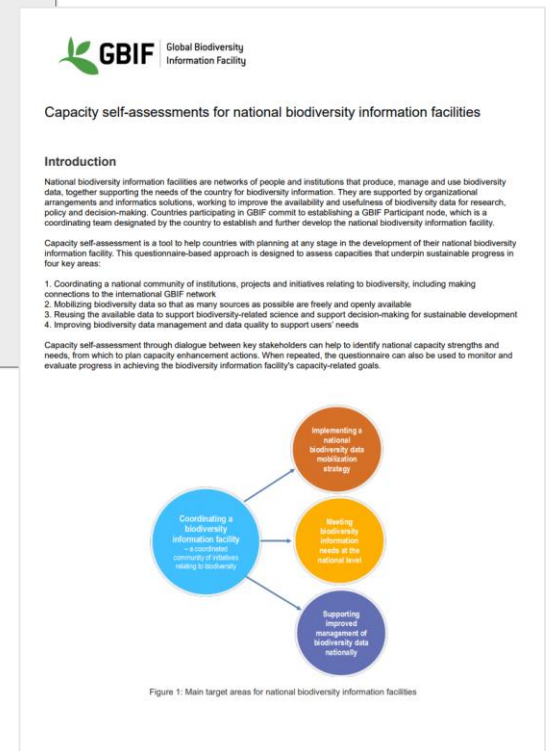
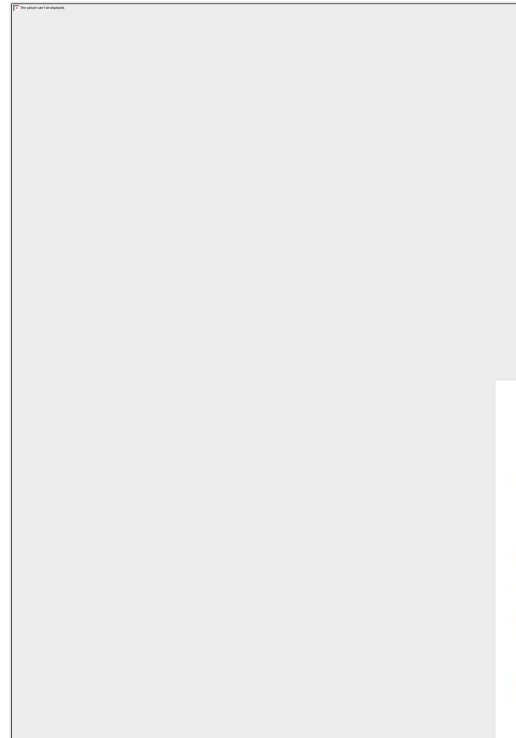
Supporting the establishment of effective Participant Nodes



UPDATES ON KEY GUIDANCE MATERIALS

Guide to establishing an effective GBIF Participant node

Capacity self-assessment guidelines for national biodiversity information facilities



<https://www.gbif.org/document/80925>

<https://www.gbif.org/document/82277>

SEEKING GUIDING EXAMPLES FROM NODES



Help others learn from your experience

How has your node grown, developed and overcome challenges? We'd like real examples from the community

Range of **suggested themes and topics**

bit.ly/gbifnode-guiding-examples

- Coordinating a national biodiversity information facility
- Implementing a national biodiversity data mobilization strategy
- Meeting biodiversity information needs at the national level
- Supporting improved management of biodiversity data nationally
- Collaborating with other nodes/initiatives (beyond the purely national scope)

Contact mraymond@gbif.org for template

VOLUNTEER MENTORS FROM THE COMMUNITY



Trainers and mentors 'selfie' at BID Pacific workshop 2017

- Approach started in the Biodiversity Information for Development (BID) programme
- **Developing a community of practice:** Involving experts from the GBIF community to support the BID funded projects
- Volunteers apply through online form and **trained after selection** in a webinar
- Mentors given roles in supporting capacity enhancement workshops and now assigned to support BID projects

GBIF VOLUNTEER MENTORS 2018



- Next step:
expanding from BID programme
- Recruiting new volunteer mentors:
 - Apply via online form
 - Plans to organize new webinar series for mentors
- Opportunities to support upcoming training events and provide support to other projects (e.g. supported through BIFA)

CURRENT MENTORS FROM NORTH AMERICA

Full name	Institution	Country	Webinar date	Volunteer Mentor Badge Awarded	BID Project Mentor	BID Trainer	GBIF Trainer
David Bloom	VertNet, Project Coordinator; iDigBio, Data Mobilization Specialist	United States	25-Jan-2017	26-Jan-2017	Yes		
Deborah Paul	Florida State University, Digitization - Workforce Training Specialist	United States					
Kate Ingenloff	PhD Candidate, Biodiversity Institute, University of Kansas	United States	12-Mar-2017		Yes		
Katherine Webbink	Field Museum of Natural History -- Information Systems Specialist	United States					
Nelson Rios	Yale University Peabody Museum, Head of Biodiversity Informatics and Data Science	United States	25/10/2017		Yes		
Robert P Guralnick	Professor and Curator at University of Florida	United States	4-Aug-2018				
Sharon Grant	Field Museum - Technology Liasion to Science	United States	13-Jul-2016	14-Dec-2016		Yes	Yes

GBIC 2

1

EMPOWER GLOBAL NETWORK

1g. Coordinate resources



GBIC2

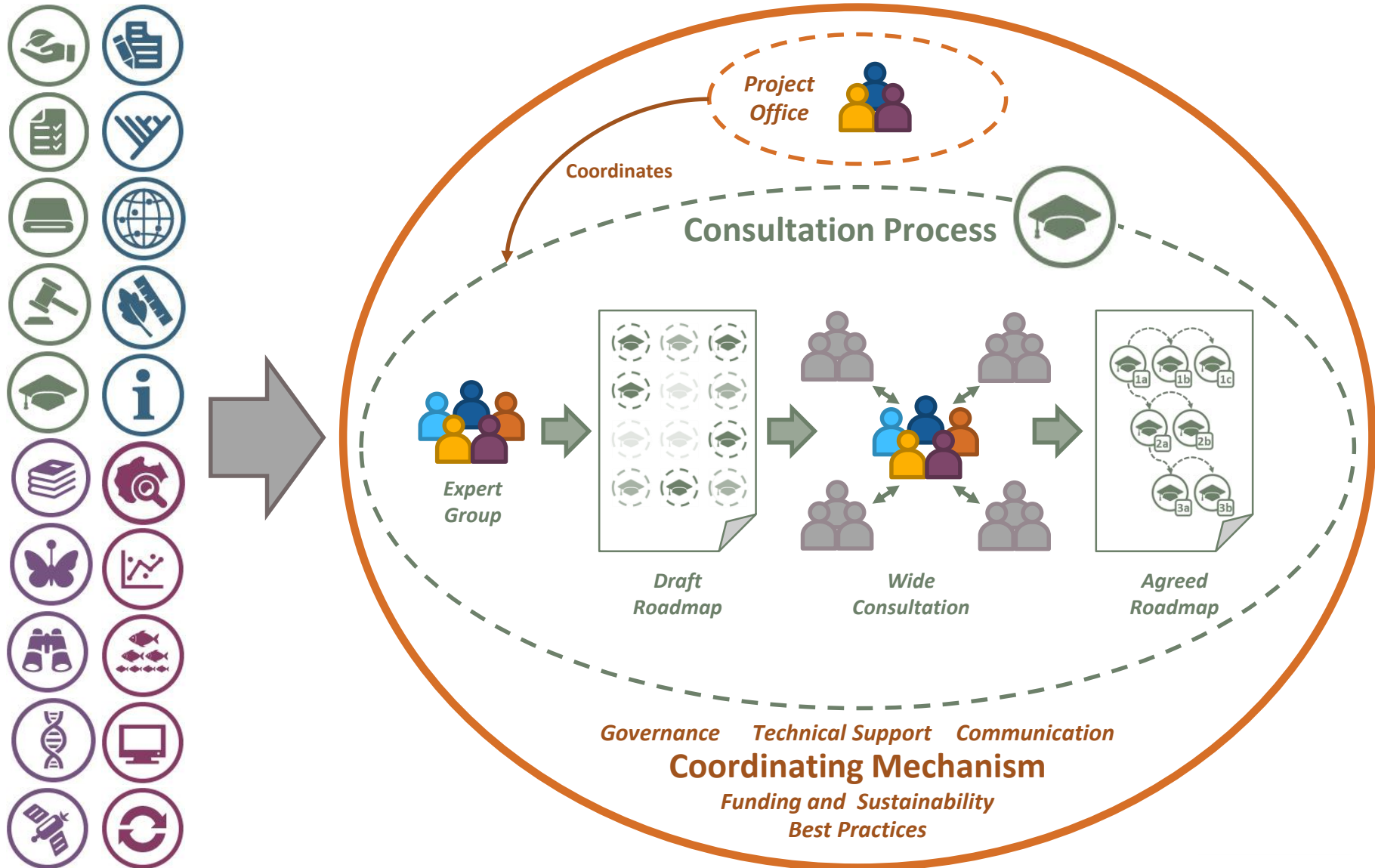
2nd Global Biodiversity Information Conference

Delivering biodiversity knowledge in the information age

24-27 July 2018
Copenhagen



GBIC2: COORDINATING GBIO IMPLEMENTATION



GBIC2: COORDINATING GBIO IMPLEMENTATION

