Global Biodiversity Information Facility

Draft GBIF Work Programme 2014-2016

Event	20th meeting of the GBIF Governing Board, 8-10 October 2013, Berlin, Germany
Agenda item	11. Presentation of GBIF Work Programme 2014-2016 in light of the GBIO
Release date	09-09-2013
Doc. status	Draft ☒ Approved ☐
Action	For approval ☒ For discussion ☐ For information ☐

Meeting documents can be downloaded from the documents folder

GBIF Work Programme 2014-2016

Executive Summary

For continuity with previous years, this document is titled *GBIF Work Programme 2014-2016*. Now that GBIF is a mature organisation, GBIF's annual activities include an increasing focus on delivering stable production-quality services based on existing infrastructure and processes. This is reflected in the level of expenditure on GBIF Secretariat staff and informatics infrastructure, and in the day-to-day activities of GBIF Secretariat staff. This Work Programme can therefore be considered to represent an *Operating Plan* to support GBIF's key capabilities.

The GBIF Work Programme 2014-2016 defines activities planned for 2014 and an outline for how these activities are expected to develop in 2015-2016. The goal is to enable GBIF to deliver critical

components of the GBIF Strategic Plan 2012-2016[1]. The multi-year view will allow GBIF Participants to plan their own activities to benefit from and align with this plan, but deferring detailed planning for 2015 and 2016 will allow more intelligent responses to changing opportunities.

This Work Programme follows the Strategic Plan in organising activities according in three **Activity Streams** based on the three Strategic Imperatives: "Advance the Informatics Infrastructure", "Advance the Engagement" and "Advance the Digital Content". There are many interconnections between these Streams, the most significant of which are indicated in the text. Each Activity Stream is intended to represent part of a set of activities continuing into 2015 and 2016.

The three Activity Streams together aim to enhance GBIF's effectiveness in supporting and communicating all areas of its mission. This work will showcase GBIF as a rich and cost-effective environment for building a globally-connected view of biodiversity occurrence and patterns and will position the GBIF network and infrastructure as the natural context for biodiversity data mobilisation activity. The biggest challenges in front of us relate to the enormous investments still needed in developing expertise and capacity for publishing biodiversity data around the world and in funding the digitisation and mobilisation of information from natural history collections and field research. This Work Programme therefore also includes work to explore establishing GBIF Supplementary Funds focussed on Capacity Development and on Content Mobilisation and to develop the case for contributions to these areas of major importance. This approach is in line with the recommendations of the GBIF Task Group on Financial Sustainability to explore models for separating between funding for core GBIF activities and for programmes of work which build on or extend the GBIF infrastructure.

The deliverables planned for 2014 are as follows:

STREAM 1: ADVANCING THE INFORMATICS – Persistence and validation

- Stable operation of new **GBIF portal** and services as a global infrastructure
- Processes to allow all data sets and data records to have **stable identifiers**
- Tools to support **data citation** needs for data publishers and for data users
- Clear **machine-readable data licences** and guidelines on usage for all GBIF-mobilised data
- Guidelines and supporting tools to assess and improve **metadata completeness** for all data
- GBIF portal upgrades to report **data quality and fitness-for-use** for each data set and species
- Proposal for use of repositories to improve **long-term persistence** of GBIF-mobilised data
- Proposal for expanding GBIF model for **endorsement of data sets**
- Demonstration of mobilisation of **sample-based data** through the GBIF network
- Criteria and process for **evaluating software compliance** with GBIF recommendations

STREAM 2: ADVANCING THE ENGAGEMENT – Communication and implementation

- Comprehensive **communication strategy**
- **Essential documents and information resources** for GBIF stakeholder groups
- Guidance and planning for **national biodiversity information facilities**
- Improved collaboration with **organisational participants**

---

[1] See the Strategic Plan at http://www.gbif.org/orc/?doc_id=2792&l=en

- 2nd **Global Biodiversity Informatics Conference** to advance GBIF-compliant data and infrastructure
- Investigation of model for building **Supplementary Funds** for GBIF-related activity

STREAM 3: ADVANCING THE CONTENT – Relevance and fitness-for-use

- Engagement of expert communities to form **fitness-for-use working groups**
- **Metrics and indicators** for assessing relevance of GBIF-mobilised data
- Identification of **content mobilisation priorities**
- Evaluation of models for creating and curating **reference data sets** within GBIF network
- **Quality and fitness-for-use reporting** through GBIF portal and services
- Support for publishing **metadata for undigitised collections**

Note that throughout this document "GBIF" is intended to refer to the interconnected activities of the whole GBIF network.

Stream 1: Advance the Informatics – Persistence and validation

Key Deliverables

- Stable operation of new **GBIF portal** and services as a global infrastructure
- Processes to allow all data sets and data records to have **stable identifiers**
- Tools to support **data citation** needs for data publishers and for data users
- Clear **machine-readable data licences** and guidelines on usage for all GBIF-mobilised data
- Guidelines and supporting tools to assess and improve **metadata completeness** for all data
- GBIF portal upgrades to report **data quality and fitness-for-use** for each data set and species
- Proposal for use of repositories to improve **long-term persistence** of GBIF-mobilised data
- Proposal for expanding GBIF model for **endorsement of data sets**
- Demonstration of mobilisation of **sample-based data** through the GBIF network
- Criteria and process for **evaluating software compliance** with GBIF recommendations

Overview

GBIF has played a central role in the development and acceptance of informatics solutions for exchange of biodiversity information. Its role as a global coordinator of solutions has led to widespread adoption of consistent data standards, tools and policies across hundreds of projects and developments around the world. GBIF itself benefits from these developments since new biodiversity data sets are increasingly published in forms that already meet GBIF requirements.

Biodiversity informatics has made great progress in the last couple of decades but serious issues remain to be addressed. Although hundreds of millions of data records are now available through GBIF, many of these lack **complete metadata** required to assist users with interpretation and to support understanding of **data quality and fitness-for-use**. Many data lack clear, unambiguous **machine-readable data licenses**. This makes it difficult for GBIF and data users to be confident that particular uses are in line with publisher expectations. Few data publishers are in a position to guarantee the **long-term persistence** and accessibility of their data. Lack of **stable identifiers** for data sets and data records makes it difficult to **cite data** used in an analysis or to give credit to the data publishers. This same issue also limits the ability of the global community to correct and curate these data.

These issues have been explored in previous work programmes and have been addressed through recommendations from a number of task groups. However there has been limited progress in delivering solutions. During 2013, GBIF's informatics development capacity has been occupied in redeveloping the fundamental data management components and portal. These provide the essential foundations for work on these issues with data persistence and with validation of data completeness, quality and fitness-for-use. During 2014, GBIF has both the opportunity and the responsibility as a global coordinator of biodiversity informatics solutions to address these issues, particularly through clear documentation and through updates to the **GBIF portal** and Integrated Publishing Toolkit (IPT).

Integration of suitable data into the GBIF network is hindered in some regions by GBIF's current process requiring Participant endorsement of data publishers prior to including their data. This process works very well in other regions. Closer engagement with expert communities through the **fitness-for-use working groups** in Stream 3 offers the possibility of GBIF using complementary

sources of information to provide a richer assessment of the value of each data set. During 2014, the GBIF Secretariat will lead a consultation with Participants on possible approaches to **endorsement of data sets** that will retain the strengths of the current model and improve scientific oversight but accelerate integration of relevant data where possible.

One of the key issues facing GBIF is the **long-term persistence** of data. Few institutions, and even fewer projects, have funded strategies that promise to maintain online access to data more than a few years in advance. As a consequence the GBIF data index increasingly includes information and summary data from data sets that are no longer accessible anywhere else on the web. This will become an increasingly serious issue. During 2014, GBIF will explore options for providing stable hosting and replication for all GBIF-mobilised data sets, including assessment of potential benefits and risks from GBIF itself managing data repositories, options for alliances with other data publishing networks, and possible technical approaches through peer-to-peer replication via the IPT and other tools. A recommendation will be developed to respond to this issue during 2015.

There is a clear and widely-acknowledge need for users in many fields to be able to understand whether data records have been collected as part of larger projects with consistent sampling methodologies and to use this criterion to discover and download data. Recent workshops have considered options to extend the Darwin Core definition to handle this situation. During 2014, GBIF will work with interested data publishers which hold suitable data to trial the use of Darwin Core extensions for **sample-based data** and to demonstrate acces to these elements through the GBIF portal and services.

Milestones

Item	Due
Documented policy on OccurrenceIDs as **stable identifiers** for data records *At GBIC-2 (see Activity Stream 2)*	Mar 14
Documented criteria for **evaluating software compliance** with GBIF recommendations *Including criteria for data publishing tools, national portals and applications using GBIF web services*	Aug 14
Portal and IPT updated to handle DOIs as **stable identifiers** for data sets *All software to work consistently with DOIs both when issued by data publishers and when issued centrally by GBIF*	Sep 14
Portal and IPT updated to support **data citation** *Consistent tools to access resolvable citation text for all data sets and all data downloads*	Sep 14
Portal reports statistics on data sets with **stable identifiers** and **metadata completeness** *Automated assessment of compliance; establish indicators of progress over time*	Sep 14
All mobilised data sets are associated with a supported **machine-readable data licence** *Updates as required to Portal and IPT, all data sets associated with alicence agreed by data publisher or with most restrictive option*	Dec 14
Portal reports on **data quality and fitness-for-use** for each data set and species *Including first metrics from fitness-for-use working groups in Activity Stream 3*	Dec 14
Demonstration of publishing of **sample-based data** through GBIF network *Mobilisation and discovery of additional data elements from a sample of GBIF data publishers with suitable data*	Dec 14

Additional detail is included in *Work Programme 2014 Additional Detail* on page 14.

Budget

Item	Costs (K€)
Stream 1 (Informatics) **2014 ACTIVITIES**	

Staff salary (incl. On-costs)[2]
 	
Includes supplementary funding (reducing costs for core funds to €889K):
  • *EU BON   €86K*
  • *EMODnet 2   €12K*
	987

Operating Expenses[3] Portal enhancements for quality reporting and sample-based data *Developer contract to expand capability of core Informatics staff*	100
IPT enhancements for identifiers, metadata and citations *Small contract to add complementary features to IPT*	30
Test implementations of sample-based data mobilisation *Work with existing GBIF data publishers holding relevant data to assess data standards and demonstrate mobilisation*	40
Explore options for replicated data repositories *Consultancy to assess options for collaboration with other networks and develop amodel for GBIF adoption*	20
PROPOSED ADDITIONAL ACTIVITIES (if funds increase)	
Pluggable Darwin Core Archive processing architecture *Contract to develop archiecture and interfaces for processing components; small contracts to develop components*	50

2015-2016 plans

During 2015-2016 it is proposed to build on 2014 activities in the following ways:

- Work to assist existing data publishers to resolve outstanding issues with **stable identifiers, machine-readable data licences** and **metadata completeness**
- Further integration of **data quality and fitness-for-use** metrics
- Initial implementation of proposal for repositories to support **long-term persistence**
- Expand mobilisation of **sample-based data** and explore options for improved discovery and visualisation of these data
- Work with biodiversity informatics developers and projects to **enhance softwa re compliance**

---

[2] On-costs are 'staff-related salary costs' such as medical insurance, education expenses, relocation and recruitment costs. These have been pro-rated over all staff. In previous WPs these costs were included under Management costs – see Summary Budget 2014 on page 18.

[3] Staff-related travel is included in Management costs, see Summary Budget 2014 on page 18.

Stream 2: Advance the Engagement – Communication and implementation

Key Deliverables

- Comprehensive **communication strategy**
- **Essential documents and information resources** for GBIF stakeholder groups
- Guidance and planning for **national biodiversity information facilities**
- Improved collaboration with **organisational participants**
- 2nd **Global Biodiversity Informatics Conference** to advance GBIF-compliant data and infrastructure
- Investigation of model for building **Supplementary Funds** for GBIF-related activity

Overview

GBIF depends on significant investments at national and organisational level to advance its mission. Its processes and infrastructure must assist Participants and other stakeholders to maximise the value of these investments, by identifying opportunities for reuse and alignment and ensuríng that all expertise, developments and contributions of data deliver benefits to all.

GBIF has been a significant force for establishing globally accepted standards for sharing and reusing biodiversity information. It has enabled many different communities to collaborate and benefit from one another's efforts. As GBIF becomes more mature with a significant network of contributors and a powerful data infrastructure, it must communicate clearly on the opportunities for GBIF stakeholders to contribute to this shared activity and how they can most fully benefit from the services GBIF can deliver.

This Stream therefore focuses first on a comprehensive **communication strategy** and on delivering **essential documents and information resources** in an up-to-date, clear and consistent form. This work will focus on key groups of GBIF stakeholders and on ensuring that they have the information they need in order to interact effectively with, and contribute to, GBIF. For each document or information resource, decisions will also be taken on the appropriate forms and languages in which the resource needs to be available. This will help all GBIF Participants and other stakeholders to understand GBIF's processes and infrastructure and the opportunities to align their own plans and deliverables with GBIF to achieve greater combined impact. It will assist enable GBIF to communicate its core mission and value proposition and engage new Participants and funding bodies. This documentation will address all key aspects of GBIF's work, including governance, data mobilisation, informatics, data use, and research and policy relevance.

One of the expected outcomes from this work will be increased clarity around the roles and responsibilities of all GBIF stakeholders, bringing new opportunities for GBIF Participants and other partners to plan and report on activities that interconnect with GBIF's mission. During 2014, processes will be set in place to support these connections. N**ational biodiversity information facilities** (BIFs) are the means by which GBIF Participants deliver relevance at the national level, by building solutions that simultaneously address the needs of data publishers, data users, research infrastructures and policy-makers. Progress in developing such BIFs varies significantly between Participants. The Secretariat will work with the Nodes Committee to develop guidance and planning materials and to establish processes for developing national implementation plans and for assessing progress.

GBIF also depends on the activities of its many **organisational participants**. Some of these are focused on data from particular regions or thematic areas. Others deliver biodiversity informatics solutions that underpin or complement GBIF. Others again focus primarily on regional cooperation and capacity development. GBIF needs to focus more clearly on the opportunities for productive collaboration with each of these Participants. The focus in 2014 will be on starting to develop clear memoranda of cooperation with existing GBIF organisational participants, identifying opportunities for better alignment of plans and activities. From 2015, GBIF will seek to incorporate these opportunities into development of its Work Programme.

GBIF has been a catalyst for international collaboration around biodiversity informatics. However, there have been few opportunities for its diverse stakeholders, including biodiversity informatics partners, software developers, node managers and interested biologists to come together and share ideas and plans for mobilising data and delivering informatics solutions. The recent Global Biodiversity Informatics Conference (GBIC), held in Copenhagen in July 2014 proved to be a valuable and cost-effective opportunity for some of these interactions. This Work Programme therefore introduces this as a biennial **Global Biodiversity Informatics Conference**, to support interactions between these many stakeholders. This will be an umbrella for biodiversity informatics organisations to meet and plan future developments. The plan includes budget to support an organising committee and to host GBIC-2 in Copenhagen in 2014. Additional funds will be fundraised to provide assistance with travel funding. The conference is planned to include a technical meeting of GBIF Nodes, with a focus on review and planning around the Nodes Portal Toolkit and existing software developments around the GBIF network. It will also provide a forum for meetings of the Working Groups presented in Activity Stream 3 and for other technical and thematic discussions.

The GBIF Task Group on Financial Sustainability has emphasised the need to expand the funding base for GBIF-related activity and proposed the need for supplementary budgets. Existing core funds support the operation of GBIF as an international, inter-governmental organisation and the activities of the GBIF Nodes network. They also support the development and operation of the global components of the GBIF data infrastructure. However GBIF does not have access to the significant funds required to address the global needs for capacity development and for content mobilisation. This Work Programme therefore includes an activity to explore possible mechanisms for GBIF to set up and operate **Supplementary Funds** in each of these areas, including consideration of the interests and needs of potential funding sources, possible mechanisms for administering projects and for guaranteeing the necessary peer review and oversight of funded projects.

Milestones

Item	Due
First draft of **communication strategy** *Identify key audiences, messages, channels*	Jan 14
Communication strategy *Including document and language priorities and delivery plan*	Mar 14
Global Biodiversity Informatics Conference	Jul 14
Memoranda of cooperation with at least five **organisational participants** *Opportunities for development or follow-up at Global Biodiversity Informatics Conference – to contribute to 2015 plans*	Jul 14
Recommendations on models for **Supplementary Funds** *For consideration at GB21*	Aug 14
Guidance and planning materials for **national biodiversity information facilities**	Sep 14

Item	Costs (K€)
First component delivered: process and resources for national review of content mobilisation opportunities	
Priority documents and information products complete in English *Priority resources for Governing Board, Node Managers, data publishers, research/policy users of biodiversity information*	Sep 14
Priority documents and information products available in two other languages *Those priority resources identified as needing translation delivered in at least two prioritised languages*	Dec 14

Additional detail is included in *Work Programme 2014 Additional Detail* on page 14.

Budget

Item	Costs (K€)
Stream 2 (Engagement) **2014 ACTIVITIES**	
Staff salary (Incl. On-costs) – combined with Stream 3[4]	*986*
Operating Expenses	
Monitoring data use in scientific publications *Contract to monitor and summarise publications, subscription to reporting services*	60
Global Biodiversity Informatics Conference *Costs for organising and planning conference*	50
Training/mentoring call *Activities supporting planning/implementation of national nodes and mobilisation of collection metadata*	40
Authoring/translating key documents *Contracts to GBIF Participants to assist with delivery or translation of priority documents*	40
Explore models for supplementary funds *Consultancy to evaluate possible models and develop proposal for GB21*	20
PROPOSED ADDITIONAL ACTIVITIES (if funds increase)	
Travel assistance for attendees to Global Biodiversity Informatics Conference *Funds to assist with travel for attendees otherwise unable to attend*	50
Integrate participant news/information in portal *Contract to enhance GBIF portal and country pages to incorporate news/information feeds from Participants*	30
Supplementary funding: Capacity development *Any additional funds specifically allocated to capacity development*	Open

2015-2016 plans

During 2015-2016 it is proposed to build on 2014 activities in the following ways:

- Further delivery of **priority documents and information products** identified within the communication strategy
- Delivery of essential documentation and information resources in **additional languages**
- Development of comprehensive mechanisms to support planning for **national biodiversity information facilities** and to report on progress

---

[4] Salary costs for Stream 2 (Engagement) and Stream 3 (Content) staff salary have been combined here for reasons of staff privacy with smaller numbers in Stream 3 – see Summary Budget 2014 on page 18.

- Implementation of identified collaboration activities between GBIF and **organisational participants**
- Third **Global Biodiversity Informatics Conference** in 2016

Stream 3: Advance the Content – Relevance and fitness-for-use

Key Deliverables

- Engagement of expert communities to form **fitness-for-use working groups**
- **Metrics and indicators** for assessing relevance of GBIF-mobilised data
- Identification of **content mobilisation priorities**
- Evaluation of models for creating and curating **reference data sets** within GBIF network
- **Quality and fitness-for-use reporting** through GBIF portal and services
- Support for publishing **metadata for undigitised collections**

Overview

GBIF now operates a global data infrastructure that is increasingly recognised as an essential resource underpinning biodiversity research and science-based policy. This position has been achieved in the first decade of GBIF's existence through adopting a focus on connecting data mobilised by data publishers anywhere in the network. In effect GBIF today depends on publisher-driven data priorities.

GBIF is now in a position to place greater focus on the needs of users and to establish research-driven and policy-driven data priorities. The goal should be 1) to understand questions in different research areas which depend on mobilisation of high-quality biodiversity data, 2) to evaluate the suitability and coverage of existing GBIF-mobilised data to address these questions (fitness-for-use in a specific area), and 3) to develop strategies for improving quality and addressing gaps in these data. Such strategies will assist GBIF Participants and funding agencies in determining how to invest in further data mobilisation. Tools to evaluate fitness-for-use of GBIF-mobilised data for different purposes will also help GBIF to measure progress in meeting the goals for which it was established.

This strategy depends on identifying and working with communities of expertise which have an interest in the application of biodiversity data in a particular thematic area. Such communities can provide guidance on fitness-for-use of biodiversity data for specific research and application uses. They can also assist with review of existing GBIF-mediated data and with developing indicators to track progress in content mobilisation. Where applicable such groups may also assist GBIF with data curation and with identifying priority data sets for future mobilisation.

GBIF has existing relationships with several communities which are well-placed to provide this guidance. Some GBIF Participants and partners represent such communities (e.g. OBIS[5] for marine biodiversity and Bioversity International[6] for agricultural biodiversity). GBIF is also part of some existing collaborations which might be expanded to provide this guidance (e.g. GIASIP[7] for invasive species and GEO BON[8] or IPBES[9] for large-scale monitoring needs). In each of these cases, GBIF will work with these partners to recognise existing groups, or when appropriate develop new groups, to act as **fitness-for-use working groups** and provide guidance on all relevant aspects of GBIF-mobilised

---

[5] http://www.iobis.org/
[6] http://www.bioversityinternational.org/
[7] http://giasipartnership.myspecies.info/
[8] http://www.earthobservations.org/geobon.shtml
[9] http://www.ipbes.net/

data. In other cases, GBIF may explore establishing a separate working group to provide guidance in a specific area of interest.

This Stream focuses on developing a mechanism for recognising or establishing groups to provide this guidance, and on establishing an initial set of working groups. OBIS and GIASIP have been identified as likely opportunities for rapid progress exploring two different models. These working groups will have a set of defined roles including development of **metrics and indicators** for assessing the coverage, completeness and fitness-for-use of GBIF-mobilised data and identification of **content mobilisation priorities** for providing key guidance to different stages of the content mobilisation process, from data capture through to end use. In future years, some of these working groups may develop to assist with identification and curation of **reference data sets** within the GBIF network, filtered subsets of GBIF-mobilised data considered to represent the best data for particular research purposes. Mechanisms and models for this concept will be explored during 2014, for implementation in 2015 and beyond. Working groups may also provide status reports which can be incorporated into GBIF communications and reports.

The aim is to establish two initial fitness-for-use working groups during 2014, based on an assessment of GBIF-aware communities that are best prepared to start this process and deliver impact. These groups will be able to participate in GBIC-2 (see Activity Stream 2).

The GBIF tools and infrastructure will be enhanced to respond to the guidance offered by these groups to incorporate recommended metrics and indicators within GBIF management processes and to use these to provide **quality and fitness-for-use reporting** through the GBIF portal. Such reporting will assist GBIF with evaluation of the effectiveness of its global data mobilisation and tools. This information will also be included within user interfaces and web services to enable users to assess whether data are likely to meet particular requirements.

GBIF task groups have recommended a strategy for mobilising metadata even from collections and projects that have not yet digitised their resources or which are currently unable to share the data. This strategy would increase GBIF's understanding of the range of resources still to be mobilised and may assist with develop of plans and proposals for addressing gaps in existing data. A strategy for mobilising **metadata for unpublished collections** will also enable these institutions to gain visibility and to begin the introduction of biodiversity informatics solutions. During 2014, GBIF will develop a clear policy and materials to guide institutions in sharing such metadata and will upgrade portal interfaces to make these resources discoverable.

Milestones

Item	Due
Process for operating **fitness-for-use working groups** and initial communities selected *Plans and terms of reference developed in collaboration with Science Committee*	Jan 14
First **fitness-for-use working groups** established *Membership established an d initial teleconference*	Mar 14
Strategy and guidelines for **metadata for unpublished collections** *Documentation and guidelines published*	Mar 14
Fitness-for-use working group workshops *At GBIC-2*	Jul 14
Fitness-for-use working group reports and recommendations *Including **metrics and indicators** and **content mobilisation priorities***	Aug 14

Item	Costs (K€)

Recommendations on mechanisms and models for **reference data sets** *For incorporation into 2015 Work Programme*	Aug 14
First **quality and fitness-for-use reporting** for metrics recommended by working groups *Integrated into quality reporting activities in Activity Stream 1*	Dec 14

Additional detail is included in *Work Programme 2014 Additional Detail* on page 14.

Budget

Item	Costs (K€)
Stream 3 (Content) **2014 ACTIVITIES**	
Staff salaries (Incl. On-costs) – combined with Stream 2	*986*
Operating Expenses	
First fitness-for-use working group *Funds to support establishing first working group and to support meetings*	30
Second fitness-for-use working group *Funds to support establishing second working group and to support meetings*	30
PROPOSED ADDITIONAL ACTIVITIES (if funds increase)	
Support for data paper publishing *Funds for small contracts to encourage activities leading to content mobilisation through data papers*	30
Additional fitness-for-use working groups *Funds to support establishing two more working groups and to support meetings*	60
Supplementary funding: Content mobilisation *Any additional funds specifically allocated to content mobilisation*	Open

2015-2016 plans

During 2015-2016 it is proposed to build on 2014 activities in the following ways:

- Continued work with those **fitness-for-use working groups** that have an interest in further work to monitor and improve the quality of GBIF-mobilised data
- Establish **fitness-for-use working groups** to review other thematic areas
- Implement preferred model for **reference data sets**
- Expand focus on mobilising and using **metadata for all natural history collections**
- Develop funding plan and proposals to seek **supplementary funding** for large-scale content mobilisation

Work Programme 2014 Additional Detail

This section provides some additional detail on tasks which will be undertaken to produce the deliverables described above under each of the Streams, and places these activities within the context of the on-going operational activities of the GBIF Secretariat.

This Work Programme introduces a separation intended to clarify the activities and costs involved in maintaining GBIF as a robust organisation. Within each of the three Streams, we have a set of **2014 Activities** which includes both the on-going operational activities required to maintain GBIF capabilities and specific activities aim to enhance these capabilities and make GBIF more relevant and effective. Recognising that budgets may change during 2014, several **Proposed Additional Activities** are also included, with the intention that these should be priorities if they can be funded during the period. These will be a priority for seeking supplementary funding.

Stream 1: Advance the Informatics – Persistence and validation

2014 Activities

The GBIF network depends on publication and hosting of primary biodiversity data sets by its Participants, but a globally integrated view of all of these data is required to support efficient discovery and access by users around the world. This Stream includes all **on-going operational activities** which require funding at a global level to support development and operation of shared informatics components.

- Operation of global central data management components
- Coordinating development of critical software components (including GBIF portal, Integrated Publishing Toolkit and Nodes Portal Toolkit)
- Promotion of community-developed standards for biodiversity data
- Coordination of biodiversity informatics strategy and delivery with other global and national projects and organisations

The following activities are planned for 2014:

- Stable operation of new **GBIF portal** and services as a global infrastructure
    - o Support for users to discover and filter data using quality criteria (including fitness-for-use working group criteria)
    - o Expanded reporting on data access through GBIF network and on use of data in research
    - o
- Processes to allow all data sets and data records to have **stable identifiers**
    - o Ensuring all data sets have a Digital Object Identifier (DOI)
    - o Seeking to ensure that each data record includes a unique occurrence identifier
- Tools to support **data citation** needs for data publishers and for data users
    - o Clear processes for citing data sets and data used in an analysis
    - o Simple processes and tools to enable users to cite data consistently and reliably
- Clear **machine-readable data licences** and guidelines on usage for all GBIF-mobilised data
    - o Seeking to ensure that all data sets have a recognised machine readable data licence
- Guidelines and supporting tools to assess and improve **metadata completeness** for all data
    - o Evaluation and reporting on metadata completeness and quality
    - o Seeking to ensure that the basis of record is clear for each data record (including understanding whether a specimen or other evidence exists)
    - o Criteria from fitness-for-use working groups (see Activity Stream 3)
- GBIF portal upgrades to report **data quality and fitness-for-use** for each data set and species
    - o Standards compliance
    - o Metadata completeness
    - o Presence of key data elements
    - o Automated checks for issues and outliers

- - Criteria from fitness-for-use working groups (see Activity Stream 3)
  - Endorsements of data publishers and data sets by Nodes, fitness-for-use working groups and other stakeholders (seeking broader basis for understanding quality and recognition for data sets)
- Proposal for use of repositories to improve **long-term persistence** of GBIF-mobilised data
  - Agreed criteria (service definitions) for establishing a GBIF-compliant persistent repository (see also Activity Stream 2)
  - Pilot implementations with a selection of GBIF Participants, using IPT or other tools
  - Evaluation of options for managing redundant storage between repositories
  - Establishing a network of recognised GBIF-compliant repositories offering hosting services to data holders
- Demonstration of mobilisation of **sample-based data** through the GBIF network
  - Agreed enhancements to Darwin Core to support elements to enable users to:
  - Recognise occurrence records from a single sampling event
  - Understand which sampling events are comparable across time and space
  - Understand relative abundance of taxa within samples
  - Pilot data mobilisation activity with existing data publishers
  - Evaluation of options for using and presenting these data through search interfaces, maps, etc.
- Criteria and process for **evaluating software compliance** with GBIF recommendations
  - Guidelines and evaluation criteria for software products to mobilise GBIF-compliant data and metadata
  - Including data publishing tools, national portals, tools using GBIF services

During 2014, these Work Programme activities will correspond to the work of 9.0 full-time GBIF Secretariat staff.

Proposed Additional Activities

The following activities are considered priorities if additional funds remain or new funds become available during 2014. Otherwise these are considered candidate activities for 2015-2016.

- Further GBIF portal upgrades to report data quality and fitness-for-use for each data set and species
  - Further criteria from fitness-for-use working groups (see Activity Stream 3)
  - Identifying use of data sets in research publications
- Contract to develop an architecture and application programming interface or service interface definition for pluggable tools to process Darwin Core Archives as part of quality checks, modelling workflows, etc., including:
  - Tools to validate standards compliance
  - Tools to check data consistency
  - Tools to validate content against external vocabularies
  - Tools to augment data with additional fields from external data sets
  - Tools to filter Darwin Core Archives
- Small contracts to promote development of pluggable Darwin Core Archive processing tools

Stream 2: Advance the Engagement – Communication and implementation

2014 Activities

The GBIF network depends on coordination and support for GBIF's Participants and for the implementation of Participant "biodiversity information facilities" (BIFs). Implementing a national or organisational BIF will involve building an effective network linking all those national or organisational stakeholders that have an interest in the mobilisation or use of biodiversity information, including data publishers, data users, related infrastructures, expert communities and funding bodies. This Stream accordingly includes all **on-going operational activities** which require funding at a global level to support capacity development and communication and adoption of common practices to deliver such BIFs.

- Support for activities of GBIF Nodes Committee and GBIF Science Committee
- Support for implementation and operation of biodiversity information facilities
- Support, coordination and advocacy for collaboration at regional and global scales
- Coordination and development of essential GBIF documentation, manuals and guidelines
- Provision of central news and communication services
- Monitoring and reporting on use of GBIF-mediated data in scientific publications

The following deliverables are planned for 2014:

- A comprehensive **communication strategy**, including:
  - Identification of key stakeholder groups, languages messages and communication channels
  - Prioritised plan for delivering essential documents and information resources
  - Prioritised plan for translating materials into Participant and/or UN languages
- **Essential documents and information resources** to support key stakeholder groups, including at least:
  - GBIF Governing Board
  - GBIF Node managers
  - Data publishers
  - Research/policy users of biodiversity information
- Processes to support GBIF **national** participants to develop effective **biodiversity information facilities**, including:
  - Strategy for planning BIFs and for assessing progress in developing a national BIF
  - Process for evaluating content mobilisation opportunities and planning a mobilisation strategy
  - Process for evaluating user communities and requirements for biodiversity data
  - Mechanism for reporting national priorities and progress in implementing a BIF
  - Projects in place to trial these processes with at least two GBIF Participants
- Processes to support GBIF **organisational participants** in planning collaborative activities with GBIF
  - Agreements identifying specific opportunities for collaboration and cooperation between GBIF and at least five GBIF Organisational Participants
- 2nd **Global Biodiversity Informatics Conference** (GBIC-2)in Copenhagen (probably in July 2014, and intended to occur in alternate years) for GBIF stakeholders with interest in:
  - Biodiversity informatics solutions including GBIF data and services
  - Implementation of GBIF Nodes
  - Evaluation of GBIF-mediated content for scientific users (including Stream 3 working groups)
- Exploration to develop a viable model for GBIF to operate **Supplementary Funds** in support of major activities not covered by core funds, including:
  - Consultation with GBIF Participants to develop models for seeking and managing funds and for ensuring scientific oversight
  - Proposal for model for introducing Supplementary Fund for Capacity Development and Supplementary Fund for Content Mobilisation

During 2014, these Work Programme activities will correspond to the work of 6.2 full-time GBIF Secretariat staff.

Proposed Additional Activities

The following activities are considered priorities if additional funds remain or new funds become available during 2014:

- Travel support to ensure broader participation at GBIC-2 (priority for additional fund-raising)
- Contrat to enhance GBIF portal and country pages to incorporate news and information feeds published directly by GBIF Participants

Stream 3: Advance the Content – Relevance and fitness-for-use

2014 Activities

GBIF does not itself publish primary biodiversity data as a centrally-funded activity, but the GBIF network depends on broad acceptance within its community of the benefits and incentives for mobilising data. This Stream includes all **on-going operational activities** which require funding at a global level to support developing and advocating best practice recommendations and supporting international adoption of data mobilisation approaches that are compatible with GBIF's needs. It also includes activities to monitor and report on the success of GBIF data mobilisation in supporting science and policy needs.

- Support, prioritisation and advocacy for biodiversity data mobilization and quality enhancement
- Coordination of assessment of fitness-for-use and quality of mobilised data
- Evaluation and monitoring of global data needs and data gaps for science and science-based policy
- Monitoring and reporting of data usage and impacts from GBIF activities

- Representation of GBIF in key international fora and partnerships for biodiversity science and policy

The following activities are planned for 2014:

- Mechanism for establishing or connecting with **fitness-for-use working groups**, including:
  - Process (with Science Committee and Nodes Committee) to identify priority thematic areas for expert engagement. Possible thematic areas include:
    - Invasive Alien Species (with GIASIP)
    - Threatened Species
    - Conservation Areas
    - Agrobiodiversity
    - Botanical data (with World Flora Online)
    - Marine systems (with OBIS)
    - Mountain biodiversity
    - Collections and taxonomy
  - Template terms of reference to establish (or recognise existing) fitness-for-use working groups to provide GBIF with guidance in a thematic area
  - Engagement with GBIF Participants to identify and select suitable members for fitness-for-use working groups
  - Joint meeting of fitness-for-use working groups at GBIC-2
- Clearly-defined roles for **fitness-for-use working groups** within GBIF data mobilisation processes, including (in relation to the Expert Working Group's thematic area), including:
  - Identification of key data-driven questions
  - Recommendations on enhancements to data standards and classes of data in use in GBIF
  - Recommendations on criteria and algorithms for evaluating data quality, fitness-for-use, coverage and completeness
  - Recommendations on **content mobilisation priorities** (both priority data resources and priorities for improving already mobilised data)
  - Identification and curation of **reference data sets**
  - Development of data mobilisation strategies
  - Reporting on progress in GBIF Annual Reports
- Enhancements to GBIF data infrastructure and processes to support and benefit from **fitness-for-use working group** recommendations (see also Activity Stream 1), including:
  - Enhancement of data indexing processes to support additional data quality **metrics and indicators** (including fitness-for-use, coverage and completeness) for different thematic areas
  - **Quality and fitness-for-use reporting** to support informed use of GBIF-mediated data and to evaluate progress in content mobilisation
  - Feedback to data publishers on possible steps to improve data quality and relevance for particular thematic areas
  - Feedback to GBIF Nodes on data resources identified as priorities
- Processes to support data-holding institutions in planning collaborative activities with GBIF
  - Process for natural history institutions to publish **metadata for undigitised collections**
  - (In Activity Stream 2) Processes to assist national biodiversity information facilities to review content mobilisation opportunities

During 2014, these Work Programme activities will correspond to the work of 1.8 full-time GBIF Secretariat staff.

Proposed Additional Activities

The following activities are considered priorities if additional funds remain or new funds become available during 2014:

- Funds to issue call for small contracts to encourage content mobilisation through data papers
- Funds to support establishment of additional fitness-for-use working groups

Summary Budget 2014

The following table summarises the 2014 budget in the context of this Work Programme. Further information is presented in the *GBIF Budget Proposal 2014*. The following should be noted:

1. All figures representing costs and funding are in thousands of Euro.

2. Costs and information on funding for activities has been split between **2014 Activities** and **Proposed Additional Activities** (priority activities that would be funded if revenue is higher), following the breakdown used elsewhere in this document.

3. **2014 Activities** includes activities planned for 2014, including all three Streams and costs for the Governing Board and committees, for operating the global Secretariat and for management and support of GBIF staff. The **Governance** and **Secretariat** lines follow the *GBIF Budget Proposal 2014* - see that document for further explanation. All salary lines include on-costs (insurance, recruitment costs, etc.) which are included within the GBIF annual budget and in GBIF financial reports as part of the Management costs.

4. Salary costs for Stream 2 and Stream 3 (and hence total costs for these streams) are combined for reasons of staff privacy with smaller numbers in Stream 3.

5. Operating expenses (staff travel and other operating expenses) incurred internally within the Secretariat at this stage are in Management costs rather than a particular Stream. At present, the financial system is not designed to attribute them to Streams.

6. Within **2014 Activities**, the items are ordered according to recommended priority (most important first). If available funds are too low to undertake all of these activities, it is proposed that the last items in the list should be the first that are deferred to future years. If on the other hand additional funds become available, the first items in **Proposed Additional Activities** are recommended for uptake.

7. Each line in **2014 Activities** indicates the expected source(s) for funding the activity, either from **Core** GBIF funds or from **Supplementary** funding from other sources. At present supplementary funds are identified only for two of these lines. If additional supplementary funds are made available and can be applied to any of the listed activities, it is proposed to transfer corresponding core funds to delivering more of the **Proposed Additional Activities**.

Item	Stream	Costs (K€)	Funding (K€)	
			Core	Supplementary
2014 ACTIVITIES				
Staff Salaries (Incl. On-costs)				
Management and support staff	-	527	527	-
Stream 1 (Informatics) *Supplementary funding from:* ▪ *EU BON €86K* ▪ *EMODnet 2 €12K*	1	987	889	98
Streams 2 & 3 (Engagement &Content)	2/3	986	986	-
Operating				
Governance *Governing Board, Committees, ENP and YRA*	-	150	150	-
Secretariat *Facilities and operating costs, including travel costs* *Supplementary funding from:* • *Copenhagen University €23K*	-	343	320	23
Stream activities				
Portal enhancements	1	100	100	-
Monitoring data use in scientific publications	2	60	60	-
First fitness-for-use working group	3	30	30	-
Global Biodiversity Informatics Conference 2	2	50	50	-
IPT enhancements	1	30	30	-
Test implementation sample data	1	40	40	-
Second fitness-for-use working group	3	30	30	-
Training/mentoring call	2	40	40	
Authoring/translating key documents	2	40	40	-
Explore models for supplementary funds	2	20	20	-
Explore options for replicated data repositories	1	20	20	-
TOTAL 2014 ACTIVITIES		*3453*	*3332*	*121*
Total Governance, Secretariat, Management and support		*1020*	*997*	*23*
Total Stream 1 (Informatics)		*1177*	*1079*	*98*
Total Streams 2 (Engagement) & 3 (Content) combined		*1256*	*1256*	*-*
PROPOSED ADDITIONAL ACTIVITY (if funds increase)				
Pluggable DwC-A processing architecture	1	50	-	-
Travel assistance for attendees to Global Biodiversity Informatics Conference	2	50	-	-
Support for data paper publishing	3	30	-	-
Additional fitness-for-use working groups	3	60	-	-
Integrate participant news/information in portal	2	30	-	-
Supplementary funding: Content mobilisation	3	Open	-	-
Supplementary funding: Capacity development	2	Open	-	-