Workgroup 8: Deliverables

# The Aim of WG8

According to the Detailed Implementation Plan of GEO BON, the *“WG focuses on data integration and interoperability to help coordinate, standardize and manage data collected by a variety of disparate institutions and individuals for many different purposes. The WG has a mandate that is somewhat different from other GEO BON Working Groups. It is not directly aimed at generating certain products about biodiversity, rather, it will focus on building permanent structures and linkages that will support the delivery of such products.”*

This aim was defined in 2010, and still stands. In this document, we scope the delivery mechanism of the required permanent structures and linkages. The aim of the WG is to put the most important pieces in place by 2015. This can only happen by working through existing, funded thematic, regional and national initiatives such as AP BON, CRIA, DataOne, EU BON. GBIF, LifeWatch, SAEON, and by linking them together through the GEOSS Common Infrastructure, as appropriate.

# Framework

We propose a framework for guiding the deliverables in WG8 that takes two major objectives into account:

1. Support for a portfolio of ***use cases*** aiming at producing certain content deliverables such as EBVs. The use cases that describe the current and future expectations for observations and data processing that a number of user communities such as GEO BON Working Groups have of technology;
2. These use cases need to be supported by an information infrastructure, comprising a number of ICT deliverables that cover the flow of data from observations to end products. . It also contains a range of ***support deliverables***, ranging from helpdesk services to ontologies and vocabularies to conceptual models, white papers, and architecture frameworks.

Within this framework, it will be possible to identify contributions and deliverables from participating organizations and individuals, and assess the extent to which it addresses one or more use cases for the delivery of content.

# Portfolio of Use Cases

The portfolio of use cases described here define the scope of interaction of a variety of user communities with information technology aimed at the production, preservation, discovery, and application of scientific data.

There is at least one use case for each GEO BON Working Group. They are described in annexure C.

# A Framework for Interoperability and Integration

Below, we frame each use case with regard to a generic workflow and life cycle of information. This spans through the phases of description, discovery, assessment, access, analysis, and reporting, as follows:

* Ensure that scientific data and services are described properly, preserved properly, and discoverable;
  + *Meta-data standards implied*
  + *Harvesters, brokers, and meta-data interoperability implied*
  + *Persistent identifiers implied*
  + *Protocols and standards for data exchange/ uploads (IPT, Tapir, BioCASE, DiGIR, etc.) implied*
  + *Preservation standards and formats implied*
  + *Tools and approaches to make searches more efficient (vocabularies, ontologies, dealing with massive meta-data collections, …)*
* Once discovered, its utility, quality, and scope can be understood, even if the data sets are huge;
  + *Implies: Visualisations, feedback on quality, quality metrics and standards, viewing search results in relation to larger spatial, temporal, and ontological/ taxonomic coverages, ability to dynamically extract 'thumbnail' views of large datasets, …*
* Once understood; it can be accessed freely and openly;
  + *Implies: standardised services, licenses and policies, …*
* Once accessed, it can be included into distributed processes, and collated - preferably automatically, and on large scales;
  + *Implies: persistence of mash-ups and mediations, web context documents, web processing services, standards and guidelines for grid computing, …*
* That due recognition is afforded to the creators of the data and services;
  + *Implies: data publication and citation, linking to scholarly articles, …*
* Once processed, the knowledge gathered can be re-used.
  + *Implies: defining and storing templates and examples of finished work, processes, mash-ups, …*

All against a backdrop of a move to extend formal meta-data with linked open data, probably based on RDF, and the construction of knowledge networks.

# A Framework Example

We are using the deliverables submitted by GBIF (Annexure A) as an example of how the process to be followed by Workgroup 8 can be approached.

* The first step is to create a matrix of use case elements and content elements – a generalized framework that can be used to assign deliverables and assess the scope of such deliverables.
* Step 2 places the GBIF deliverables in context within the matrix.
* The matrix can also be used to prioritise deliverables, taking both the importance of use case elements and the content deliverables into account.

Annexure B provides an example of the assignment of GBIF deliverables to the framework matrix.

# What Workgroup 8 Should Do

Given that a framework approach can be adopted, we can define the tasks for Workgroup 8 within it:

1. Discuss, agree, and prioritise a portfolio of use cases.
2. With each use case, define the use case elements and the implications for standards, technology, and specifications.
3. Compile a set of content deliverables (data, services, white papers, ontologies, …) that supports the use cases and their delivery of the agreed Essential Biodiversity Variables. The other workgroups are busy defining their deliverables, and this has a direct impact on the scope and priorities defined for these. Our inter-workgroup discussions should focus on this.
4. It is worth noting that the EBVs imply a wide range of data, meta-data, and service formats or community standards that are not limited to traditional biodiversity observation data, and may include many or all of the following initial list:
   1. Meta-Data:
      1. Darwin Core and species observations;
      2. ISO 19115/ p2 describing and vector/ raster spatial data sets;
      3. EML.
   2. Services:
      1. GBIF services;
      2. OGC WxS spatial data services;
      3. NetCDF and HDF-4/5 data services;
      4. Sensor observation services.
   3. Search and Discovery Protocols:
      1. OGC Catalogue Services for the Web (CS/W);
      2. OpenSearch;
      3. SPARQL ;
      4. Open Archive Initiative (OAI-PMH).
   4. Mediation, Distributed Processing and Automation:
      1. Kepler;
      2. OGC Web Context Documents;
      3. OGC Web Processing Services;
      4. Taverna;
      5. Vistrails.
5. Based on the above, construct and maintain the framework matrix, and perform the following tasks:
   1. Assign deliverables contributed by organisations and individuals;
   2. Identify gaps, synergies, and overlaps;
   3. Conceptualise and define additional deliverables (work packages or projects) to address gaps or synergies (“opportunities”), bearing use case and content priorities in mind;
   4. Communicate these opportunities to GEO BON workgroups, GEO, and the wider community.
6. Build permanent infrastructure and linkages. Do that by registering the deliverables and their services in the available registries such as those of the GCI and GBIF, and build integrative components such as reusable workflows and portals on top the deliverables.

# Template for Deliverables

The template to be used for Workgroup 8 deliverables needs to be different from those of the other workgroups, given that it does not address content directly. The format used by GBIF (Annexure A) is more appropriate and can be used as a basis for agreeing a Workgroup 8 template.

# Annexure A

Deliverables identified by GBIF – Submitted by *Éamonn Ó Tuama*

GEO BON Deliverable from GBIF (1)

|  |  |
| --- | --- |
| Deliverable name | GBIF Registry |
| Deliverable type | Application: Core informatics infrastructure |
| Deliverable description | Work is currently underway to expand the GBIF Registry to ensure its suitability as a central discovery tool for datasets holding all classes of biodiversity data, through collaboration with other global and regional biodiversity informatics projects. The GBIF registry provides a vital role in integrating all the components of a global, distributed biodiversity network. Functions include:   * Allow the registration and discovery of a growing amount of entities: Institutions, networks, datasets, schemas, vocabularies, etc. * Provide the means to direct clients on how to access network resources * Accurately model the complex relationships between entities, to enable correct attribution (e.g. recognizing data hosting partnerships) * Provide a reliable identifier “minting” service, allowing distributed systems to connect on common resources. * Provide network monitoring services, to (e.g.) provide alerts on new resources, or technical failures (servers going offline) * Offer search capabilities through indexing of metadata * Enable external classification on registered objects through the use of tagging (both private and public tagging) |
| Targeted user groups | Data publishers (providers); scientists (using biodiversity data); 3rd parties constructing customized biodiversity portals, e.g., GWOS |
| Partners | GBIF Participants; global and regional biodiversity informatics projects |
| Inputs required | Populating the registry requires inputs from all major/relevant biodiversity networks |

GEO BON Deliverable from GBIF (2)

|  |  |
| --- | --- |
| Deliverable name | Integrated Publishing Toolkit (IPT)  <http://www.gbif.org/informatics/infrastructure/publishing/> |
| Deliverable type | Software application |
| Deliverable description | The IPT provides a software platform to facilitate the efficient publishing of biodiversity data on the Internet, using the GBIF network. The IPT provides a flexible solution for publishing different types of data through the use of the Darwin Core Archive (DwC-A) format. Work is underway to extend the “core” types to include a “sample” core in addition to the currently supported “occurrence” and “taxon” cores. A star schema allows the association of additional data with these cores through use of DwC extensions. The system thus has capability to meet the needs of GEO BON for publishing/delivery of a variety of biodiversity content types. |
| Targeted user groups | Publishers (providers) of biodiversity data |
| Partners | Biodiversity community |
| Inputs required |  |

GEO BON Deliverable from GBIF (3)

|  |  |
| --- | --- |
| Deliverable name | GBIF data index  <http://data.gbif.org/> |
| Deliverable type | Data product |
| Deliverable description | The GBIF network serves as a publication and integration mechanism for thousands of data sets on the recorded occurrence of different species in time and space. All of the included data records are indexed and increasingly validated to support efficient search, discovery and download.  As of November 2012, this data index includes over 389 million records, of which more than 340 million have associated coordinates. GBIF already provides access to this index via the GBIF web services listed below, but can also explore alternative paths and mechanisms to deliver views and subsets of the data for use in GEO BON products. |
| Targeted user groups | Parties engaged in modeling and visualizing species distributions |
| Partners | Natural history collections, ecologists, citizen scientists, environmental impact assessors, etc. |
| Inputs required |  |

GEO BON Deliverable from GBIF (4)

|  |  |
| --- | --- |
| Deliverable name | GBIF web services  <http://data.gbif.org/tutorial/services> |
| Deliverable type | Web services |
| Deliverable description | The GBIF portal offers a range of web services that can be used by other portals and applications to directly access XML formatted GBIF data. The services currently available include:   * Taxon data service - (<http://data.gbif.org/ws/rest/taxon>) * Occurrence record data service -(<http://data.gbif.org/ws/rest/occurrence>) * Occurrence density data service - (<http://data.gbif.org/ws/rest/density>) * Dataset metadata service - (<http://data.gbif.org/ws/rest/resource>) * Data publisher metadata service - (<http://data.gbif.org/ws/rest/provider>) * Data network metadata service - (<http://data.gbif.org/ws/rest/network>   The GBIF Registry itself is implemented as a service-oriented architecture. All interactions with the registry is via a RESTful API, and the GBIF portal itself is the first and principal client of the registry. A revised and expanded set of REST based web services including JSON output will be available with launch of new GBIF portal in 2013 (see:  <http://dev.gbif.org/wiki/display/POR/Webservice+API> ) . |
| Targeted user groups | 3rd parties wishing to build thematic portals |
| Partners |  |
| Inputs required |  |

GEO BON Deliverable from GBIF (5)

|  |  |
| --- | --- |
| Deliverable name | GBIF Nodes Portal Toolkit (NPT)  <http://community.gbif.org/pg/groups/3507/nodes-portal-toolkit-npt/> |
| Deliverable type | Modular software application |
| Deliverable description | The Nodes Portal Toolkit (NPT) is envisioned as a suite of informatics tools designed to help GBIF Participants to deploy, maintain and extend biodiversity data portals (at the national, regional or thematic levels). The NPT will support the setup of customized biodiversity data portals. It employs a strategic extensible approach that supports the means for the GBIF community to progressively add functionalities, allowing participants to design, manage, improve and share their developments for enhanced biodiversity data portals. This addresses the challenge of meeting a heterogeneous set of reporting requirements from very simple (e.g. simply embedding a distribution map with a bit of textual context) to very complex, such as advanced geospatial processing, etc. |
| Targeted user groups | Scientists |
| Partners | GBIF Participants |
| Inputs required | Programmers, developers required to assist with module development |

GEO BON Deliverable from GBIF (6)

|  |  |
| --- | --- |
| Deliverable name | A Knowledge Organization Framework (KOS) for biodiversity informatics  <http://community.gbif.org/pg/groups/21382/vocabulary-management/> |
| Deliverable type | Framework |
| Deliverable description | Common vocabularies that describe the properties of biodiversity resources are necessary for the effective integration of biodiversity data derived from different sources. Thus, the development and uptake of such vocabularies, including thesauri, ontologies, etc. - collectively referred to as Knowledge Organization Systems (KOS), is a prime concern for biodiversity informatics. GBIF has already begun to explore issues around collaborative vocabulary development and management ensuring that work on KOS is fully addressed in its work programme. Working within the Taxonomic Databases Working Group (TDWG) it has led on establishing a vocabulary management task group (VOMAG) to continue the community-wide development of a KOS infrastructure for biodiversity informatics. |
| Targeted user groups | Biodiversity practicioners / domain experts at large |
| Partners | TDWG |
| Inputs required |  |

GEO BON Deliverable from GBIF (7)

|  |  |
| --- | --- |
| Deliverable name | GBIF vocabulary management tool  <http://terms.gbif.org> |
| Deliverable type | Web application |
| Deliverable description | A prototype Semantic Wiki based application intended for the evaluation by the TDWG/GBIF Vocabulary Management Task Group (VoMaG) in collaboration with WP 4 of the Virtual Biodiversity Research and Access Network for Taxonomy (ViBRANT). The aim is to provide a user-friendly and easy to use platform for the collaborative development of vocabularies including definition and translation of basic concepts and public discussion, and as a potential front-end for discovery/exploration of biodiversity terminology. |
| Targeted user groups | Biodiversity informaticians and biodiversity scientists (domain) experts; translators (from English to other languages) |
| Partners | TDWG; ViBRANT |
| Inputs required | Community inputs to term definition/translation |

GEO BON Deliverable from GBIF (8)

|  |  |
| --- | --- |
| Deliverable name | GBIF Online Resource Centre (ORC); <http://www.gbif.org/orc/> |
| Deliverable type | Web site |
| Deliverable description | The GBIF Online Resource Centre is a global online service that provides easy access to documents, best practices, tools and links relevant for GBIF Participants and their Biodiversity Information Facilities. This new system allows advanced functionality such as:   * Support for diverse resource types * Wide thematic scope * Different ways of accessing resources * Enabling community contributions * Different levels of resource access * Multilanguage support |
| Targeted user groups | New and experienced users of biodiversity resources |
| Partners | GBIF Participants |
| Inputs required |  |

GEO BON Deliverable from GBIF (9)

|  |  |
| --- | --- |
| Deliverable name | Capacity building – outreach: a global collaboration platform for biodiversity informatics |
| Deliverable type | Community network |
| Deliverable description | The GBIF network support training, capacity building, communication and outreach necessary to enable the widest possible participation, geographically and amongst thematic communities in establishing Biodiversity Information Facilities for data publishing and sharing through the GBIF infrastructure. A regional focus helps to ensure engagement and empowerment and offers potential synergies with the plans for GEO BON regional groupings. |
| Targeted user groups |  |
| Partners |  |
| Inputs required | Liaison between GBIF and GEO BON regional organizing committees. |

# Annexure B

Example Framework Matrix - Mapping GBIF contributions

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | Set of **Essential Biodiversity Variables** | | | |
| **Use Case and Use Case Elements** | | WG1 | | | WG2… |
| DataSets | Services | Support | … |
| **Proper Description of Data and Services** | |  |  |  |  |
|  | Conceptual Model: Important Entities |  | GBIF-1 | GBIF-1 |  |
|  | Meta-Data Standards | GBIF-2 | GBIF-4 |  |  |
| **Discovery of Data and Services** | |  |  |  |  |
|  | Harvesters and Brokers | GBIF-3 | GBIF-1 |  |  |
|  | Service Availability Monitoring |  | GBIF-1 |  |  |
|  | Guideline architecture: hierarchical or federated? |  | GBIF-5 |  |  |
|  | Improved utility: semantic interoperability | GBIF-6 GBIF-7 | GBIF-1 GBIF-4 | GBIF-6 |  |
| **Data/ Service Preservation** | |  |  |  |  |
|  | Protocols and Standards for Archiving/ Repository Uploads |  | GBIF-2 |  |  |
|  | Persistent Identifiers |  | GBIF-1 |  |  |
|  | Format preservation and integrity protocols |  |  |  |  |
| **Assessment of Data Set or Service Utility and Quality** | |  |  |  |  |
|  | Standardisation of Systematic Quality Metrics |  |  |  |  |
|  | Usability Metrics |  |  |  |  |
|  | Cost and/or time metrics |  |  |  |  |
| **Assessment of Scope and Utility** | |  |  |  |  |
|  | Data visualisations and abstractions | GBIF-3 | GBIF-4 |  |  |
|  | Assessment of coverage (temporal, spatial, semantic) | GBIF-3 | GBIF-4 |  |  |
| **Free and Open Access** | |  |  |  |  |
|  | Standardised Data Policies |  |  |  |  |
|  | Standardised License Regimes, including agreed conditions |  |  |  |  |
|  | Community consensus on funding sources, open access |  |  |  |  |
| **Automated or Assisted Collation of Distributed Data** | |  |  |  |  |
|  | Mediation and persistence of collations |  |  |  |  |
|  | Guidelines in respect of data dimensionality/ scales/ resolutions |  |  |  |  |
| **Automated and Distributed Processing** | |  |  |  |  |
|  | Web processing services, workflow orchestration |  |  |  |  |
|  | Distributed process execution |  |  |  |  |
| **Data Publication and Citation** | |  |  |  |  |
|  | Long-term availability and persistent identifiers |  |  |  |  |
|  | Interfaces for access, visualisation, and citation | GBIF-3 | GBIF-4 |  |  |
|  | Guidance on citation indices af open access |  |  |  |  |
| **Knowledge Persistence and Re-Use** | |  |  |  |  |
|  | General support of communities of users |  |  |  | GBIF-8 GBIF-9 |
|  | Concepts and standards supporitng knowledge networks |  | GBIF-5 | GBIF-5 |  |
| **…** | |  |  |  |  |
|  | … |  |  |  |  |
|  | … |  |  |  |  |

# Document Information

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | | | |
| **Version** | **Date Issued** | **Authors** | **Description of Major Changes** | |
| G372.2.1 Framework – WG8.docx | 19-11-2012 | W Hugo | Inception |
| G372.2.2 Framework – WG8.docx | 27-11-2012 | W Hugo  H Saarenmaa  Éamonn O Tuama | Consolidated comments from contributors  Included GBIF deliverables as examples | |