EU BON MS241: Specification for registry and metadata catalogue

Submitted: 2015-03-27

Milestone MS241 for Task 2.4 (Metadata registry and catalogue), due in month 27 (Feb 2015), was deferred to allow discussion at the EU BON workshop in Joensuu (March 2015). The actual registry and catalogue implementation (D2.3: registry and metadata catalogue) is due in month 39 (Feb 2016) with a service update (MS242: updated registry and metadata catalogue) in month 51 (Feb 2017).

Contents

Introduction

Task 2.4 (Metadata registry and catalogue) is described as follows in the Description of Work (DoW).

Building on the existing GBIF and LTER registry and metadata catalogues, an enhanced and integrated metadata system will be developed for EU BON. The various entities such as networks, projects, sites, and datasets identified in the analysis and mobilization efforts of WP1 will be described in the new registry/catalogue. The entity descriptions should include web service interfaces or other access points, and will also be registered at the GCI and other indexing services. In order to overcome heterogeneity of data, accommodate multilingualism, enhance discoverability and interoperability, and facilitate querying in portals, the use of Knowledge Organisations Systems (KOS; e.g., thesauri) will be explored. (Lead GBIF; UEF, CSIC, Pensoft, MRAC, INPA, IBSAS; Months 9-51)

Central to the efficient functioning of any network is the presence of a registry. This can be thought of as a relational database with human and web service interfaces for the registration and curation of network entities. It stores inter-relationships of entities (e.g., which institution owns which data sets), enabling discovery and access (via cached data and/or stored end points). This information is typically captured in metadata records that are stored in a metadata catalogue (a metadata database).

The requirements as stated in the EU BON DoW recognise the need to draw on already existing registries and catalogues such as GBIF and LTER and to integrate these with the GEOSS Common Infrastructure (GCI). It also indicates the kinds of entities that should be documented, including not only data sets but also related networks, projects and sites. The system developed should also support semantic querying.

Networks and data repositories participating in EU BON are expected to expose their data and metadata through standards-based services, e.g., well known international standards such as OGC CSW and W*S or community agreed *de facto* standard such as the GBIF APIs. The principal task for a registry such as EU BON is then to integrate information across such disparate networks. One solution to this problem is to use a broker, a component which acts as a communications middleman between the source and the target of a communication. By this means, a community can continue to use their own technologies and standards and leave the establishment of interoperability across communities to the broker. The broker becomes responsible for interfacing to the various community standards and mapping them to a common internal model so that all the participating networks become accessible via a single unified interface. The broker can also support well known standards such as OGC CSW.

General registry architecture

The high level architecture of EU BON as proposed by CSIC relies on an Enterprise Service Bus (ESB)[1] that is responsible for orchestrating workflows and connecting various data and service providers [1]. The registry and metadata catalogue form one component of the ESB and can be implemented as a broker catalogue system such as GI-cat[2] (Figure 1) which was introduced to EU BON partners in the Crete meeting [2] [3]. The GEO Data Access Broker (DAB) application, based on GI-cat, connects disparate information resources from different communities that are not aware of each other's interoperability mechanisms. It can be used to harvest content from existing registries and bring them under one view. In the case illustrated in Figure 1, the common data model (view) is based on the ISO 19115 metadata standard.

After reviewing the recommendation that GI-cat be evaluated as the brokering system for EU BON, CSIC determined that, as a specialized broker, GI-cat is a powerful solution for integrating metadata sources under a common data model and for providing an interface from EU BON to the GEOSS registry/catalogue. However, as it is lacking in connectors for common WSDL services and specific input sources needed for EU BON purposes (e.g., GenBank, EU-Nomen, WoRMS, etc), a revision of the architecture is therefore proposed, consisting of a hybrid solution that integrates GI-cat inside a larger ESB based SOA architecture (Figure 2).
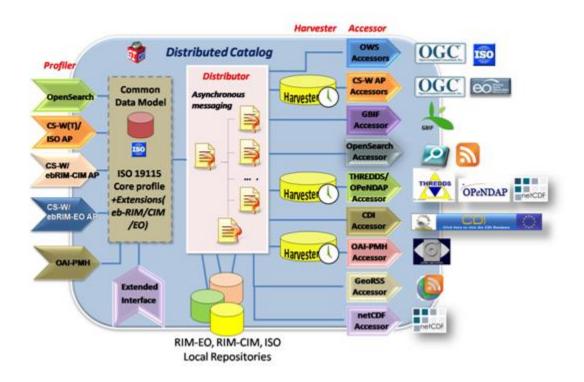

Figure 1. The GI-cat broker system featuring some catalogue query interfaces (right) and several backend mediation components (source: Nativi et al. 2009).

---

[1] http://eubon.cybertaxonomy.africamuseum.be/sites/default/files/EU%20BON_EAI_SOA-presentation.pdf
[2] http://essi-lab.eu/do/view/GIcat/WebHome

The GI-cat broker system as described can thus serve as an essential component of the EU BON's European Biodiversity Portal for registering and connecting the partner systems of EU BON. As described in the EU BON DoW, the European Biodiversity Portal (task 2.5)

> "*will technically integrate the various data sources under one search facility and spatially/temporally oriented user interface. The portal will build on the tools developed in task 2.3, functions developed by task 2.4. It will provide access to full detailed data, geographic visualisations and remotely sensed data. It will be closely linked to the GCI and GEO Portal, and access layers and data from GEOSS sources*".

The broker design does not impose common APIs on each participating partner system. Rather, while communities should be encouraged to adopt, where appropriate, well known and widely deployed standards such as OGC web services, in their absence, accessors for each system will need to be developed and centrally managed within GI-cat. It is important to recognize the burden of this, and thus encourage partners to make use of existing registries (e.g., GBIF, DataONE) to ensure that only a few accessors are actually needed (Figure 3). For example, it may be more efficient for the KNB network to connect to the EU BON broker via GBIF (i.e., GBIF harvests the KNB EML metadata first), rather than directly. There are several established registries (robust, integrated, with helpdesk, secure in funding, etc.) that have been running for years that should be utilised rather than building competitors. Not only does this reduce overlap, it helps ensure EU BON is building on infrastructure that has the investment needed to ensure longevity and robust operation.
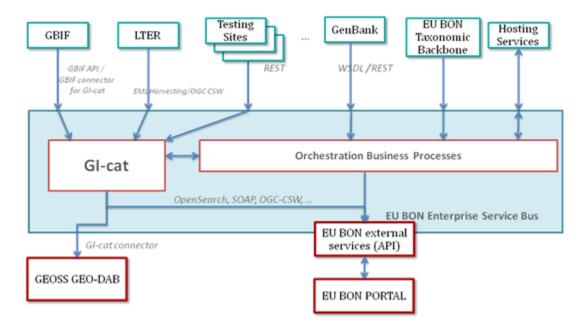

Figure 2. Proposed architecture for EU BON consisting of a hybrid solution in which GI-cat in integrated in the enterprise service bus.

EU BON Portal

As the principal client of the Enterprise Service Bus API, the EU BON portal will incorporate an interface to the registry and metadata catalogue allowing end users to discover and access metadata and associated data. Additionally, providing the metadata holdings of the EU BON network to the GEOSS catalogue is one of the main ways EU BON will contribute to the GEOSS Common Infrastructure (GCI) and the GEO-Discovery and Access Broker. The EU BON portal will therefore need to provide an OGC CSW service that the GEOSS catalogue can access. As a specialized brokering system, GI-cat can provide OGC CSW outputs in addition to its own GI-cat connector (in order to connect to other GI-cat). Other entities on the network may decide to connect their catalogues to GEOSS directly by supporting their own OGC CSW.

For further details on the EU BON portal, please consult MS251 – Specification for the EU BON portal.

Common data model

The common data model currently supported by GI-cat is based on the OGC 19115 geospatial metadata standard. Widely promoted and adopted, e.g., within the EU INSPIRE framework[3], this specification is well suited for describing geospatial resources in general (e.g., temporal and geospatial coverages) but is somewhat lacking in expressivity for certain aspects of biodiversity related resources, e.g., for describing taxonomic coverage and detailed field sampling methodologies. It is therefore recommended that the EU BON registry extend the common data model in GI-cat to incorporate, at least, taxonomic coverage, as would be available in original EML metadata.

Search operation

GI-cat supports both federated search as well as a harvesting/caching mechanism. Harvesting of metadata is the usual choice to optimise performance but the chosen mechanism can be matched to the characteristics of the participating network entity.

Semantic mediation

GI-cat is part of the larger GI-suite[4] that includes the GI-sem component. GI-sem supports semantic queries through query expansion involving external semantic services and aligned ontologies (e.g., GEMET, GEOSS SBA vocabulary[5], etc.). GI-sem will be investigated as part of the EU BON registry to test how heterogeneity of data and multilingualism can be accommodated when searching for data.

---

[3] http://inspire.ec.europa.eu/documents/Metadata/MD_IR_and_ISO_20131029.pdf
[4] http://www.uos-firenze.essi-lab.eu/#!gi-suite/c3ak
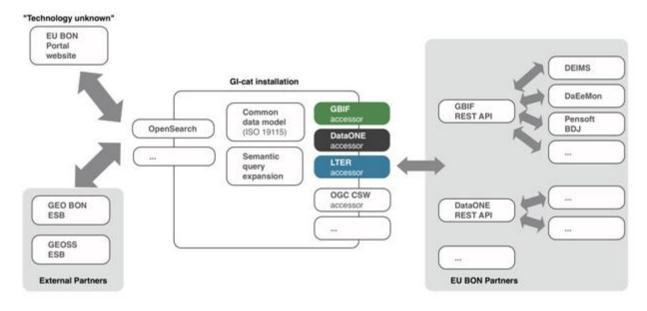[5] http://en.wikipedia.org/wiki/Societal_Benefit_Areas

Figure 3. The architecture for the EU BON registry as deployed using GI-cat.

Network entities

Several specific components and sub-networks have already been identified for integration in EU BON. In addition to GBIF, these include, e.g., the broader DataONE network (including LTER, KNB, ILTER, Dryad), DaEuMon[6] and DEIMS[7]. These are described in the following section. In addition, Annex 1 provides an extended list of network resources which are candidates for connection to the EU BON registry. For each entity, where known, the protocol/access point/source and metadata standard used are noted.

GBIF

The GBIF registry is comprised of a database and metadata store allowing management of the following entities:

- Organizations: basic metadata, contacts and technical access points
- Datasets: metadata (built around the GBIF EML Metadata Profile, but supporting the complete EML specification), contacts, technical access points, tags, machine tags
- Networks: basic metadata and ability to relate datasets with networks.

The registry is accessible through a RESTful API[8] and is connected to many software platforms around the GBIF network. As an existing robust platform, it can serve as a gateway platform for some activities that do not yet have an existing technical infrastructure, e.g., DaEuMon.

---

6 http://eumon.ckff.si/about_daeumon.php
7 http://www.lter-europe.net/info_manage/lter-infobase
8 http://www.gbif.org/developer/registry

Pensoft have already integrated the Biodiversity Data Journal with the GBIF registry.

Within EU BON, GBIF is addressing the following:

1. Harvesting and indexing of the Knowledge Network for Biocomplexity (KNB) metadata catalogue so that KNB resources are discoverable through EU BON.
2. Providing access for authenticated users to the machine tagging API to allow simple annotation of content.
3. Via the registry, orchestrating persistent storage of datasets within DataONE so that persistence is guaranteed and data become discoverable through DataONE and available for DataONE workflows.

DataONE
DataONE[9] provides a REST based API[10] for accessing all services including metadata. DataONE encompasses several member-nodes including KNB and Dryad.

LTER/ILTER
Member nodes of LTER/ILTER[11] use the Morpho/Metacat system for publishing data sets with EML metadata into the KNB network.

KNB
KNB is an international repository for ecological/environmental data sets, and, as such, a key member node of DataONE. KNB is the home site for the Morpho/Metacat publishing system[12].

Dryad
Dryad[13] is an archive for data sets underlying scholarly publications. It is a member node of DataONE.

DaEuMon
The DaEuMon database of monitoring schemes has been identified as a key resource for EU BON. It requires enhancement on two fronts: the coverage needs to be increased from the current 643 entries at least two-fold and the database model needs to be enhanced to cover issues of access interfaces, download of data, terms of use and database details. In conjunction with these developments, it is proposed that the following actions be undertaken:

1. Providing an EML document for each monitoring scheme (database record)
2. Registering each institution within the GBIF registry and through the GBIF REST API[14]

---

[9] https://mule1.dataone.org/ArchitectureDocs-current/overview.html
[10] https://mule1.dataone.org/ArchitectureDocs-current/apis/
[11] http://www.ilternet.edu/
[12] https://knb.ecoinformatics.org/
[13] http://datadryad.org/
[14] http://www.gbif.org/developer/registry

3. Registering each monitoring scheme as a dataset within GBIF through the API, and registering the complete EML document
4. Registering a network "EUMON" within the GBIF registry, and associating all datasets with the network
5. Where available, registering technical endpoints for the data

These steps will allow the DaEuMon database to function autonomously while ensuring that the data are discoverable and accessible through existing infrastructure and available through a well-defined API, which will be integrated within the EU BON and GEOSS infrastructure. Additionally, the datasets will benefit from forthcoming improvements to the GBIF registry such as assignment of DOIs for persistent identification and citation. Furthermore, extended metadata descriptions could be turned into "data paper manuscripts" and submitted to publishers.

DEIMS

DEIMS[15] provides a web client interface for accessing data/metadata associated with ILTER, LTER-Europe and the EU projects EnvEurope and ExpeER. The DEIMS web site does not appear to have implemented its discovery service yet although EnvEurope provides a GeoPortal site where individual metadata documents are available in EML, RDF/SKOS and ISO 19115 (but not via an API/Service).

PlutoF

PlutoF workbench[16] is a biodiversity data management platform. It is used to create, manage and publish research and citizen science projects under which various taxon occurrence data (e.g. herbarium and living specimens, human observations, molecular data, and reference-based occurrences) are collected and mobilized. Data collected and managed in PlutoF, together with projects' metadata, can be accessed through public RESTful API-s, but will also be made available to the research community through GBIF.

EU BON WP 1 Inventory

The inventory of biodiversity data sources developed in WP1 (task 1.1: assessment and evaluation of biodiversity data sources) will be published as a collection of EML metadata documents to GBIF.

ENVRI search portal

The ENVRI project has developed a search portal[17] that provides a unified interface for searching of connected metadata catalogues. It is implemented as a single central catalogue hosted at ESA ESRIN that harvests metadata from participating research institutes. There are options for other parties to install the software but, unlike GI-cat, it is not provided as a deployable package but must, rather, be set

---

[15] http://data.lter-europe.net/deims/
[16] https://plutof.ut.ee/
[17] http://envri.eu/

up from several components (e.g., PostgreSQL, POSTGIS, Proj.4, libXML2, etc). The query interface is based on the OGC CSW OpenSearch standard. The ENVRI search portal can thus be readily connected to the EU BON broker which has OpenSearch built in.

INSPIRE GeoPortal

The INSPIRE directive provides specifications for the Spatial Data Infrastructure in European Community. The INSPIRE GeoPortal provides a road map to authoritative datasets, geospatial services, catalogues outlining content, an application program interface (API) for developers to create value-added applications, and core geospatial applications that can be used throughout Europe.

The INSPIRE Geoportal network main services include:

- Discovery Services, consisting of registries and catalogues to locate spatial datasets and associated services.
- Download Services, for accessing complete dataset(s) or part of a dataset(s) based on OGC WFS 2.0 specification which translates into ISO/DIS 19142 and 19143 standards.
- View Services, based on the OGC WMS 1.3.0 specification which translates into the ISO 19128 ISO/DIS standard.

Searching and catalogue interfaces are based on CSW requests but many features such as compliant GML representation of the dataset are only available if the dataset owner has enabled it.

The GeoPortal supports REST-based harvesting: the harvester can connect to a remote GeoPortal version 9.3.x or 10.x server and retrieve metadata records that match the query parameters specified using the GeoPortal REST API.

GeoServer[18] is open source software that implements many Open Geospatial Consortium standards including Web Map Service (WMS 1.1.1, 1.3.0), Web Feature Service (WFS & WFS-T 1.0.0, 1.1.0, 2.0), Web Coverage Service (WCS 1.0.0, 1.1.1), and Web Processing Service (1.0.0). It offers an INSPIRE extension that allows GeoServer to be compliant with the View Service specifications for the INSPIRE directive.

ESRI has adapted the ArcGIS system to provide ArcGIS for INSPIRE[19], a common platform for GIS applications throughout any European organization. This solution provides discovery, view, and download services for INSPIRE-compliant data, as well as the ability to create metadata and data models for those who need it.

Poland's GeoPortal2 project[20] is an example of an INSPIRE related broker.

Protected Planet

The Protected Planet portal[21], the web interface to the World Database of Protected Areas (WDPA)[22], is an interactive mapping website with comprehensive information on the different types of protected

---

[18] http://geoserver.org/
[19] http://www.esri.com/software/arcgis/arcgis-for-inspire
[20] http://geoportal.gov.pl/en/aplikacje/geoportal-inspire
[21] http://www.protectedplanet.net/
[22] http://en.wikipedia.org/wiki/World_Database_on_Protected_Areas

areas. The WDPA is a joint project between the United Nations Environment Programme (UNEP) and the International Union for Nature Conservation (IUCN).

The server-side components of the portal were developed by the United Nations Environment Programme-World Conservation Monitoring Centre (UNEP-WCMC), while front-end development and visual design was handled by Vizzuality, based on open source technologies, GeoServer and PostGIS.

The WDPA Development Team at UNEP-WCMC has a formal agreement with GBIF, to integrate their network species occurrence data with the shapefiles of protected areas on the WDPA, which assists organisations to visualise the occurrences of species within a protected area.

Plazi

Plazi[23] provides data extracted from the published scientific literature. The cited literature is available via DOI either through a CrossRef DOI or through the Biodiversity Literature Repository, a joint initiative of CERN, Plazi and Pensoft. The data is treatment based, i.e., the part of publications that are linked to a specific taxonomic names usage are extracted, semantically enhanced and, if available, linked to external resources such as ZooBank for taxonomic names. The data include any data published. With an agreement with GBIF, observation data are provided via a Darwin Core Archive. All the records are provided with a unique identifier, the treatments with a persistent http URI, and the literature cited with DOIs allowing to cite each of the data and follow back to its source.

All the software used is open source including, whenever available, TWDG standards, e.g., DwC, Taxon Concept Schema (TCS), or TaxonX and TaxPub, the DTD for the latter based on the National Library of Medicines Journal Article Tag Suite and developed and maintained in collaboration with Pensoft and NLM.

Summary

The EU BON ESB will have to support the metadata APIs of several established communities as well as standard metadata services such as OGC CSW and OAI-PMH. The high level common metadata model, currently implemented in GI-cat as ISO 19115 will need to accommodate EML, in particular regarding taxonomic and field methodology coverage given that it is probably the most widely used metadata standard in the biodiversity and ecological communities.

---

[23] http://plazi.org

References

[1] EU BON D2.1 Architectural design, review and guidelines for using standards. http://eubon.eu/getatt.php?filename=EU%20BON_D2.1_Architectural%20design%20review%20and%20guidelines%20for%20using%20standards_10716.pdf

 [2] First EU BON Training Event , Heraklion, Crete, April 3rd 2014. Information architecture – GEOSS perspective - Lorenzo Bigagli. http://eubon.cybertaxonomy.africamuseum.be/sites/default/files/uploaded_files/Bigagli_slides.pdf

 [3] S. Nativi, L. Bigagli, "Discovery, Mediation, and Access Services for Earth Observation Data", Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of Volume 2, No 4, Page(s):233 – 240, 2009 URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=05200393

Annex 1. Network entities for consideration for inclusion in the EU BON registry

This annex is also available as a separate document on Google Docs (https://docs.google.com/document/d/1c_iuoukHbSEeLeFvV5Kz-1INJMCQ3qBLX_WW3m6MvDw) so that additional entities can be added as they are identified through ongoing requirements gathering.

#	Entity	Protocol / Access Point / Source	Standards used
1	GBIF	GBIF REST API; http://www.gbif.org/developer/summary	DwC; EML
		GBIF OGC CSW; (expected late 2015)	EML, ISO 19115, DC
2	DataONE	DataONE REST APIs; http://mule1.dataone.org/ArchitectureDocs-current/apis/	EML
3	LTER/ILTER	via KNB/ DataONE	EML
4	KNB	via DataONE	EML
5	Dryad	via DataONE	EML
6	DaEuMon	via GBIF?	EML?
7	DEIMS	DEIMS REST API?	EML
		OAI-PMH?	
		OGC CSW?	
8	PlutoF	via GBIF	EML, DwC
9	ENVRI	OGC CSW	ISO 19115

10	WP1 Inventory	via GBIF	?
11	Pensoft/BDJ	via GBIF; BDJ has an RESTful API to import complex manuscripts: http://pwt.pensoft.net/dev; The url resource for downloading an article in XML from BDJ API is http://bdj.pensoft.net/lib/journal_archive.php?journal={journal_name}&date={date}	EML, DwC, DwC-A, OAI-PMH, TaxPub
12	NASA Global Change Master Directory	RESTful web service: http://gcmdservices.gsfc.nasa.gov/mws/capabilities?format=html OGC-CSW: http://gcmdsrv.gsfc.nasa.gov/csw?request=GetCapabilities&service=CSW	Directory Interchange Format (DIF); OGC-CSW;
13	BioFresh metadatabase	http://data.freshwaterbiodiversity.eu/metadb/ check for API	
14	INSPIRE Geoportal	http://inspire-geoportal.ec.europa.eu; GeoPortal REST API, OGC CSW	ISO 19139, ISO19115:2003, GML
15	Protected Planet	http://www.protectedplanet.net no API?	KMZ, CSV, SHP
16	Mercury metadatabase	http://mercury.ornl.gov Supports RSS, Geo-RSS, OpenSearch, Web Services and JSR-168 Portlets	FGDC, Dublin-Core, EML, ISO-19115
17	Global Land Cover	http://glcfapp.glcf.umd.edu/ No API	FTP; TIFF; JPG
18	Bioclim (Bioclimatic Variables)	http://geodati.fmach.it/production/ows_europe_lst	OGC WMS, WCS

19	ASTER Global Emissivity Database (GED); Emissivity, mean land surface temperature, NDVI	NASA; http://emissivity.jpl.nasa.gov/aster-ged; no API? search and download via several portals	HDF5, JPG and KML images
20	AMSR-E/Aqua Daily L3 surface soil moisture; Soil moisture/Water content	NSIDC; http://nsidc.org/data/ae_land3 NO API The Data Pool provides direct FTP access to NSIDC Earth Observing System (EOS) data, including the entire archive of AMSR-E data.	FTP; HDF
21	ECV soil moisture dataset; Soil moisture based on six sensors	ESA; http://www.esa-soilmoisture-cci.org/node/145; no API?	NetCDF
22	Harmonized World Soil Database; Soil map of the world	FAO/IIASA; http://webarchive.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/ ; no API?	BIL
23	ERS/MetOp Soil Moisture; Coarse resolution soil moisture data	TUWien; http://www.ipf.tuwien.ac.at/radar/index.php?go=ascat	
24	ASCAT coastal winds; Swath grid - Winds	EUMETSAT; http://www.osi-saf.org/; no API?	WMO GRIB (ed 2); BUFR formats; NetCDF
25	Corine Land Cover	http://discomap.eea.europa.eu/arcgis/services/Land/CLC2012_Dyna_WM/MapServer/WMSServer?request=GetCapabilities&service=WMS	OGC WMS
26	DLR-AVHRR; Various atmospheric variables (temperature,air)	NOAA/AVHRR; https://wdc.dlr.de/sensors/avhrr/; Uses WDC Integrated Services: Generic workflows for user-/group management, product and metadata handling and publishing via OGC	DLR-specific "DFD" format; raw HRPT and NOAA LAC format.

		compliant interfaces + DOI registration.	
27	NASA MEaSUREs; Vegetation phenology and indexes – NDVI	NASA; https://earthdata.nasa.gov/our-community/community-data-system-programs/measures-projects; several portals for data access	OPeNDAP; FTP; HDF5
28	WorldClim; climate layers	WorldClim; http://www.worldclim.org/ no API	ASCII grids; ESRI grids
29	CliMond; climate layers	CliMond; https://www.climond.org/ no API	ASCII grids; ESRI grids
30	Global Satellite Mapping of Precipitation; Precipitation analysis	NASA; http://sharaku.eorc.jaxa.jp/GSMaP_crest/ no API?	FTP; CSV
31	MODIS Land Surface Temperature dataset	https://earthdata.nasa.gov/data/standards-and-references/earth-data-science-disciplines/land/surface-temperature; download from several data centres	
32	Ecosystem maps	USGS; http://rmgsc.cr.usgs.gov/ecosystems/ ArcGIS REST API for map services.	ESRI; TIFF; etc ArcGIS API services provides: Export map, Identify, Find, Generate KML
33	Mangrove maps	USGS; http://edcintl.cr.usgs.gov/ip/mangrove/	"data will be available soon for download. We are still preparing the final formats of these files"
34	Digital elevation model (MODIS)	http://modis.gsfc.nasa.gov/data/	
35	Global Urban Footprint	DLR; http://www.dlr.de/eoc/en/desktopdefault.aspx/tab	For DLR portals, access to EOC Earth observation data: "OGC-compliant, convenient

		id-9628/16557_read-40454/	data discovery, viewing, and download services are available through the EOC-Geoservice"
36	Forest height	DLR; http://www.dlr.de/hr/desktopdefault.aspx/tabid-2470/3731_read-5661/	
37	Leaf Area Index (LAI) (Spot satellite)	http://land.copernicus.eu/global/?q=products/lai	
38	Global forest change	University of Maryland; http://earthenginepartners.appspot.com/science-2013-global-forest	
39	Sea surface temperature	NOAA; http://www.ospo.noaa.gov/Products/ocean/sst/contour/index.html	
40	GlobCover; global composites and land cover maps from MERSI sensor and ENVISAT satellite mission	http://due.esrin.esa.int/globcover/; no API; downloadable zip file	GeoTIFF
41	SEDAC; socioeconomic data	http://sedac.ciesin.columbia.edu/data/sets/browse; several OGC services: http://sedac.ciesin.columbia.edu/maps/services	WMS; WCS; WFS; WMTS; WM-SC
42	Plazi: treatment, observation and traits data	via GBIF: http://plazi.org no API; HTTP GET to retrieve a DWC-Archive that includes the observation records, treatments, list of taxa included	DwC, DwC-A, OAI-PMH, TaxPub, Taxon

Annex 2: Task assignment for implementing the EU BON registry and catalogue system

This annex is also available as a separate document on Google Docs (https://docs.google.com/document/d/1KW2AAKpTfgtn9_pg8q6moNv69giZKOP4NzhyON0OK10) so that tasks can be added and/or amended as implementation proceeds.

Partner person-month allocations

Task 2.4 is led by GBIF. Seven partners are involved. The person month allocations for those participating are: GBIF (11), UEF (10), CSIC (1), Pensoft (2), MRAC (0.5), INPA (1), IBSAS (14).

Tasks

The following table provides a list of tasks involved in the establishment of the EU BON registry and the partner who is responsible for executing a task.

#	Task	Agent	Due by	Comment
1	Install and configure an instance of GI-cat software	CSIC	done	Evaluation due Sept 2014
2	Assess complexity of accessor development	CSIC	ongoing	Evaluation due Sept 2014; awaiting access to source code
3	Develop, enhance the GBIF registry, metadata catalogue and Restful API	GBIF	ongoing	
4	Harvest KNB EML metadata for inclusion in GBIF metadata catalogue	GBIF	Dec 2015	
5	Provide access for authenticated users to the machine tagging API to allow simple annotation of content	GBIF	June 2015	
6	Via the registry, orchestrate persistent storage of datasets within DataONE so that persistence is guaranteed and data become discoverable through DataONE and available for DataONE workflows	GBIF	Oct 2015	
7	Expose GBIF metadata via OGC CSW	GBIF	Dec 2015	

8	Develop script for harvesting GBIF EML documents via the GBIF Registry API	IBSAS	June 2015	GBIF have Java libraries for this; contact trobertson@gbif.org
9	Develop XSLT for transforming GBIF EML profile to ISO 19115	IBSAS	June 2015	GBIF has an XSLT for EML->ISO; can be used as base for further development
10	Modification of the existing GI-cat accessor for GBIF to allow searching for datasets using the GBIF Registry API, and returning taxonomic information according to an ISO extension or metadata keywords.	CSIC	ongoing	Consider how this overlaps with tasks 7 and 8
11	Develop accessor for DataONE	IBSAS	Sept 2015	As an alternative to developing an accessor, harvest DataONE EML via GBIF
12	Develop accessor for DEIMS	CSIC	ongoing	Use the DEIMS+GeoNetwork OGC CSW to harvest metadata in ISO19115 format, or harvest EML directly.
13	Extend DaEuMon schema (for describing monitoring schemes); run campaign to include more schemes; provide enhanced metadata in EML	UEF	From April	
14	Develop accessor for DaEuMon.	IBSAS	Sept 2015	As an alternative to developing an accessor, GBIF may harvest EML metadata directly from DaEuMon instead
15	Connect ENVRI catalogue based on OGC-CS OpenSearch	CSIC	June 2015	Configure GI-cat to use the OGC-CSW OpenSearch interface implemented by ENVRI
16	Connect any prioritised remotely sensed data catalogues	CSIC	ongoing	Such catalogues are likely to be using OGC CSW; configure GI-cat to use them.

17	Develop interface(s) from broker to GCI, e.g. OpenSearch, CSW, etc.	CSIC	Sept 2015	Use the OGC-CSW interface included in GI-cat
18	Connect metadata in WP1 inventory	UTARTU	June 2015	Publish WP1 inventory via GBIF
19	Accommodate taxonomic information in GI-cat central model	CSIC	June 2015	Use an ISO extension, e.g., the NAP biological extension, or simply store the taxonomic information within the portal.