

Milestone MS232

Version: 1
Date: 2015-05-11
Authors: Larissa Smirnova, Patricia Mergen, et al.
Document reference: MS232



Technical workshop, review of documents, test versions of data sharing tools (M27)

STATUS: FINAL

Project acronym: EU BON
Project name: EU BON: Building the European Biodiversity Observation Network
Call: ENV.2012.6.2-2
Grant agreement: 308454
Project Duration: 01/12/2012 – 31.05.2017 (54 months)
Co-ordinator: MfN, Museum für Naturkunde - Leibniz Institute for Evolution and Biodiversity Science, Germany

Partners:

- UTARTU, University of Tartu, Natural History Museum, Estonia
- UEF, University of Eastern Finland, Digitisation Centre, Finland
- GBIF, Global Biodiversity Information Facility, Denmark
- UniLeeds, University of Leeds, School of Biology, UK
- UFZ, Helmholtz Centre for Environmental Research, Germany
- CSIC, The Spanish National Research Council, Doñana Biological Station, Spain
- UCAM, University of Cambridge, Centre for Science and Policy, UK
- CNRS-IMBE, Mediterranean Institute of marine and terrestrial Biodiversity and Ecology, France
- Pensoft, Pensoft Publishers Ltd, Bulgaria
- SGN, Senckenberg Gesellschaft für Naturforschung, Germany
- VIZZUALITY, Vizzuality S.L., Spain
- FIN, FishBase Information and Research Group, Inc., Philippines
- HCMR, Hellenic Centre for Marine Research, Greece
- NHM, The Natural History Museum, London
- BGBM, Botanic Garden and Botanical Museum Berlin-Dahlem, Germany
- UCPH, University of Copenhagen: Natural History Museum of Denmark, Denmark
- RMCA, Royal Museum of Central Africa, Belgium
- PLAZI, Plazi GmbH, Switzerland
- GlueCAD, GlueCAD Ltd. – Engineering IT, Israel
- IIEP, Institute for European Environmental Policy, UK
- INPA, National Institute of Amazonian Research, Brazil
- NRM, Swedish Museum of Natural History, Sweden
- IBSAS, Slovak Academy of Sciences, Institute of Botany, Slovakia
- EBCC-CTFC, Forest Technology Centre of Catalonia, Spain
- NBIC, Norwegian Biodiversity Information Centre, Norway
- FEM, Fondazione Edmund Mach, Italy
- TerraData, TerraData environmetrics, Monterotondo Marittimo, Italy
- EURAC, European Academy of Bozen/Bolzano, Italy
- WCMC, UNEP World Conservation Monitoring Centre, UK

This project has received funding from the European Union's Seventh Programme for research, technological development and demonstration under grant agreement No 308454.








EU BON

EU BON: Building the European Biodiversity Observation Network
Project no. 308454

Large scale collaborative project

MS232**Technical workshop, review of documents,
test versions of data sharing tools**

Milestone number	MS232
Milestone name	Technical workshop, review of documents, test versions of data sharing tools
WP no.	WP2
Lead Beneficiary (full name and Acronym)	RMCA
Nature	Workshop held
Delivery date from Annex I (proj. month)	2015-02-28 (M27)
Delivered	yes
Actual forecast delivery date	2015-05-11 (M30)
Comments	Participant list can be found in the annex of this document

Name of the Authors	Name of the Partner	Logo of the Partner
Larissa Smirnova	RMCA	
Patricia Mergen	RMCA	
Hannu Saarenmaa	UEF	
Éamonn Ó Tuama	GBIF	
Francisco Antonio García Camacho	CSIC	

In case the report consists of the delivery of materials (guidelines, manuscripts, etc)

Delivery name	Delivery name	From Partner	To Partner

Summary of the Milestone

The milestone MS232 reports back the results of the EU BON informatics workshops which was held March 18-20, 2015 back to back to CETAF ISTC annual meeting in Joensuu, Finland (see full program and presentations: <http://digitalium.fi/en/content/eu-bon-and-cetaf-joint-informatics-workshop>). The event was organised by the University of Finland (UEF, <http://www.uef.fi/>) and the Digitalium, (<http://digitalium.fi/en>) - EU BON consortium member and WP2 Leader, in collaboration with CETAF ISTC and other EU BON work packages. It aimed to demonstrate the new EU BON developments and prototypes, try out new ideas and share experience between partners. The main outcomes of the meeting are draft documents on information architecture (more particularly registry and portal) and data sharing tools and action plans for the next year. 36 people from the EU BON consortium and beyond (16 countries) attended the workshop (see participant list in the Annex).



Introduction

The annual technical workshop is a next step in achieving the objectives set by the project:

- Establish an **information architecture** for the EU BON project that will be compatible with the global GEO BON, INSPIRE, other European projects, and the LifeWatch research infrastructure
- Develop data integration and interoperability between the various networks, and with **new generation of data sharing tools** enhance linking between observational data, ecosystem monitoring data, and remote sensing data
- Develop new web service interfaces for data holdings using state-of-the-art **standards and protocols**. Register the networks on the GEOSS Common Infrastructure (GCI) using harmonised metadata
- Develop a **new portal** to enable fast access to EU BON integrated data and products by researchers, decision makers and other stakeholders

The workshop was dedicated to the demonstration of newly developed or adopted EU BON products, discussion and drawing up the action points for the next period. It has been also worked on upcoming deliverables and papers. Confinement to CETAF ISTC workshop allowed to gain synergies with the IT representatives of major EU institutions and to outreach to a non EU BON community considered as important stakeholders of the projects outputs.

The workshop included several presentations:

- Presentation of Digitalium (Riitta Tegelberg)
- CETAF summary of state of play (Anton Güntsch)

- Data sharing tools (Patricia Mergen)
- EU BON Registry (Éamonn Ó Tuama)
- EU BON Portal (Antonio Garcia)
- EU BON Citizen science gateway (Veljo Runnel)

Presentations, post-conference documents and pictures are available here: <http://digitarium.fi/en/content/eu-bon-and-cetaf-joint-informatics-workshop#post-conference>

Achievements and current status

Drafting the MS232: Test versions of data sharing tools

An overview of MS231 “Specifications of data sharing tools” (http://eubon.eu/getatt.php?filename=Milestone_MS231_Specifications%20of%20data%20sharing%20tools_11247.pdf), submitted in May 2014, was given and the importance of the document was emphasized as it collects extensive information about:

- currently available tools for the biodiversity research community,
- recommendations towards the main requirements needed to build new releases of data sharing tools for EU BON data providers.

Following the recommendations of the EU reviewers and Scientific Advisory Committee to make the content more accessible to the various users the milestone will be published as a paper. This will be initiated in spring 2015.

The action points were defined as follows:

- Use the main part of the document for a publication. Annex 1 is already available on the Data mobilization helpdesk as a searchable repository (<http://eubon.cybertaxonomy.africamuseum.be/data-sharing-tools-repository>)
- Pensoft Biodiversity Information Journal (<http://biodiversitydatajournal.com/>) has been identified as suitable for the publication. The requirements/guidelines of the journal will be checked.
- Draft outline of the paper will be made available in April 2015. It is envisaged to use the collaborative writing tool of Pensoft. An agreement on the list of co-authors will have to be reached; logically these will be the authors and contributors to MS231.
- Reshape the current text by adding citations. Up-to-datedness of the milestone (especially the recommendations and conclusions) should be checked. Use cases with examples of usages are envisaged for the next deliverable. The final draft to be checked by native English speaker, with the objective to keep the text user friendly for a not IT literate audience.
- The preferential goal is to have it published before November 2015 (second EU review of EU BON project) or alternatively to have it submitted and accepted by then.

It was also agreed that the tools repository on the Data Mobilisation Helpdesk would be regularly updated. This includes new information and releases about the listed tools, but also adding additional tools as they are used or tested by the community.

Beyond the repository of data sharing tools, the Data Mobilisation Toolkit also contains information on the other components of EU BON architecture such as registry, citizen sciences products, monitoring schemes and standards. This information should also be checked and updated by all partners on regular basis.

Another issue discussed was writing the deliverable 2.2 on data sharing tools. It should be ready due February 2016 (month 39). The deliverable should take MS231 in account, but focus on implementation of the selected tools, latest data standards for biodiversity observations, their interlinkage with the EU BON data portal (EBP) release, and positioning among the other technical components. The deliverable should include the description of the installations and implementations of the data sharing tools, analysis of their usage and usability with the use cases from the test sites (WP5). Synergies with other WPs are also important, so for example it should be discussed with WP3 and WP4 - what is the proper data format they need to run their models and with WP6 and 7 – what data is expected for the policy and decision makers and if chosen tools can provide such data in suitable formats.

The action plan for writing the deliverable:

- See if the deliverable can be written from the start in the format of a publishable paper.
- The deliverable will be focusing on tools effectively used within EU BON and will demonstrate it on use cases from test sites (mainly in Europe, but also taking into account the non-European test sites, i.e. Brazil).
- As gaps have been identified with East European data sources, thus these stakeholders should be addressed.
- Interactions and interoperability with GEO BON has to be addressed, but also with other networks such as GBIF, TDWG, COL, LTER LifeWatch etc.

Drafting the MS241: Metadata registry and catalogue specification

The penultimate draft version of MS241¹ was presented by Éamonn Ó Tuama (GBIF, 2.4 Task Leader). Registry is a database for documenting and curating descriptions of the network entities. It will be build on existing GBIF and LTER registry and metadata catalogues, should have an interface to GEOSS Common Infrastructure (GCI), and possibly support semantic querying. The only remaining and important issues are Network entities (networks, projects, sites, datasets). The table of resources contains 42 entities by now, EU BON partners are invited to add resources to that table. Detailed task allocations were presented².

The action points were defined as follows:

- Update the list in Annex 1 of MS241 -> will be a living part of specification (on google doc)³
- Implement the actions assigned to each partner of the Task, at the timeline given
- Contact WP3-4 to identify what data is needed to achieve their goals. This is an ongoing task and some inputs to Annex 1 have already been provided.

Drafting the MS251: Specification for the EU BON portal

MS251 was presented by Francisco Antonio García Camacho (EBD-CSIC, 2.5 Task Leader). The portal is developed by IBM (with CSIC supervision) and the prototype is available at <http://161.111.232.210/EBPCoreWeb/>.

Main focus was put on the integration of GBIF (API v1.0) and LTER-EU searches. For the taxonomic filter PESI EU-Nomen & WoRMS were used.

¹ As of 27-03-2015, MS241 has been submitted and is available at:

http://eubon.eu/getatt.php?filename=EUBON-MS241-registry-specification_12216.docx

² https://docs.google.com/document/d/1KW2AAKpTfgtn9_pg8q6moNv69giZKOP4NzhyON0OK10

³ https://docs.google.com/document/d/1c_iuoukHbSEeLeFvV5Kz-1INJMCQ3qBLX_WW3m6MvDw

Ten specifications were presented. Several specifications are already implemented in the prototype. Several important issues remain and were points of extensive discussions:

- User roles
- Search functionalities
- Quality control

User roles: up to now three groups of users were identified (general public, researcher and policy maker). It was emphasised that the “search” should be adapted to the particular user needs. Only user-friendly interface can be proposed. Search criteria should be also adapted to user roles, this could be a subject for a new survey. Also workflow for data supply is important to understand.

Search functionalities, filters: these will include filtering by Data types (as defined in DoW), by temporal or taxonomic coverage.

Several issues were discussed, for example what tool should be used for geospatial representation; different layers for integration and representation; content management.

Data quality: the important issues of data quality of datasets were discussed and the EU BON portal should provide an annotation tool for data quality. It was agreed that data quality is a complex issue and difficult to define on several reasons:

- What is quality? Reliable data? Completeness? Fitness for use?
- Lack of standards for annotations and quality control
- Criteria are difficult to define (rating depends on purpose you will use the data for)

The general action points were defined as follows:

- The requirements of the stakeholders and site-specific requirements should be added
- Feedback from other WPs should be surveyed
- Next version of the document will be prepared and sent to consortium before end April.

Citizen science gateway (Veljo Runnel, UTARTU)

The results of the EU BON (2014) survey „Assessing Citizen Science involvement in biological research“ were presented and the importance of citizen science (CS) in modern data landscape was emphasised. There are pros and cons in the assessment of citizen science data, but no doubt that the engagement of citizen scientists is and will be even more important in the future. Surprisingly, 78% of respondent assess the quality of data as satisfactory. Citizen science data are widely used by local decision makers, so availability of such data is an important advantage of data mobilisation network. GBIF has already started with indexing citizen science data (e.g., from eBird, iNaturalist) and the development of tools such as EU BON IPT tool will only ease the data sharing from various biodiversity monitoring applications.

There are still challenges in data providing from CS initiatives:

- Access unavailable data
- Data standards
- Data duplication in different portals and databases
- Language barriers

The module for managing citizen science data on the PlutoF platform was demonstrated. The next developments will include:

- Best practise case analysis
- Guides for effective citizen science project management repository
- Integration with EU BON portal
- Collaboration with major citizen science organisations as ECSA (European Citizen Science Association)

Challenges and further/future developments

Hosting service:

An important question raised during the technical workshop was a need for a hosting service for data. If EU BON chooses centralized tools, then data should be hosted by the portal, but if a distributed model is chosen, this may not be necessary. A central cache or index is however still needed with keeping at least some metadata centrally. In case of GBIF, the data remains with the providers but the problem can occur if countries, not associated to GBIF (East Europe for example), want to provide their data. Hosting service is not foreseen in DoW and thus there are no funds and no person months foreseen for that. Different possibilities were discussed, and it was concluded that for the long-term sustainability it is best to use the existing services like GBIF, LifeWatch, GEO BON and on the longer run to check with the ERIC or future ESFRI EU infrastructures.

Mobilisation of data:

The important question to be addressed is what data is going to be mobilised? EU BON has to collaborate with other initiatives like GBIF, LTER or LifeWatch on that. It was agreed that EU BON should also support GBIF and fill in the gaps by trying to mobilise data from East Europe. Another major activity is collaboration with EuMon, because EuMon has metadata of biodiversity monitoring schemes, which is the focus of EU BON and GEO BON, and is thus a priority for the next period.

In conclusion, the importance of this workshop should be emphasized as it allows correlating the technical work done by different tasks and work packages, to plan the future steps and to coordinate goals and objectives to minimise the efforts and maximise the output. Coming together helps project partners to clarify their ideas, discuss with colleagues different possibilities, demonstrate prototypes and recent developments and it increases in general the capacity building of the project.

Annex: list of participants (* - EU BON partner, ** - EU BON Associate partner)

	Name	Country	Institute
1	Aaike De Wever*	Belgium	Royal Belgian Institute of Natural Sciences
2	Alan Paton	UK	Royal Botanic Garden Kew
3	Allan Zirk*	Estonia	University of Tartu Natural History Museum Senckenberg Research Institute Frankfurt
4	Andreas Allspach*	Germany	(SGN) CETAF Consortium of European Taxonomic Facilities
5	Ana Casino**	Belgium	
6	Anton Güntsch*	Germany	Freie Universität Berlin
7	Cedric Chaveriat	France	French Foundation for Biodiversity Research
8	Christian Langer*	Germany	GEO BON
9	Dare Talvitie	Finland	Luomus French Foundation for Research on Biodiversity (FRB)
10	Aurelie Delavaud**	France	
11	Éamonn Ó Tuama*	Denmark	GBIF
12	Elsbeth Haston	UK	Royal Botanic Garden Edinburgh
13	Florian Wetzl*	Germany	Museum für Naturkunde Berlin
14	Florian Wolf	Germany	iDIV
15	Francisco Antonio García Camacho*	Spain	CSIC
16	Franck Theeten*	Belgium	Royal Museum for Central Africa
17	Gregor Hagedorn*	Germany	Museum für Naturkunde Berlin
18	Hannu Saarenmaa*	Finland	UEF Digitalium
19	Heimo Rainer	Austria	NHM Wien
20	Henry Engledow*	Belgium	Botanic Garden Meise
21	Israel Peer*	Israel	GlueCAD - Eng. & Biodiversity IT
22	Jere Kahanpää	Finland	Finnish Museum of Natural History
23	Jiri Frank	Czech Republic	National Museum Institute of Botany, Slovak Academy of Sciences (IBSAS)
24	Karol Marhold*	Slovakia	
25	Konstanze Bensch	Germany	Botanische Staatssammlung München
26	Larissa Smirnova*	Belgium	RMCA
27	Marian van der Meij**	Netherlands	Naturalis
28	Marie-Elise Lecoq	France	GBIF France - MNHN Institute of Botany, Slovak Academy of Sciences
29	Matúš Kempa*	Slovakia	
30	Patricia Mergen*	Belgium	Royal Museum for Central Africa
31	Miguel Fernandez	Germany	GEO BON
32	Sophie Pamerlon	France	GBIF France - MNHN
33	Simon Chagnoux**	France	MNHN CNR (National Research Council) - LifeWatch ITA
34	Stefano De Felici	Italy	
35	Teodor Georgiev*	Bulgaria	Pensoft Publishers
36	Veljo Runnel*	Estonia	University of Tartu, Natural History Museum