Report and assessment of training activities and final versions of training manuals

M45

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 Research on Evolution and Biodiversity, Germany
Delivery date from Annex I:	M45 (August 2016)
Actual delivery date:	M45 (August 2016)
Lead beneficiary:	MRAC
Authors:	Kim Jacobsen, Larissa Smirnova, Patricia Mergen (MRAC, Royal Museum for Central Africa), Isabella Van de Velde (RBINS, Royal Belgian Institute of Natural Sciences)
	Pavel Stoev (Pensoft, Pensoft Publishers Ltd, Bulgaria)
	Anke Hoffmann, (MfN, Museum für Naturkunde - Leibniz Institute for Research on Evolution and Biodiversity, Germany)
	Hannu Saarenmaa (UEF, University of Eastern Finland, Digitisation Centre, Finland)

This project is supported by funding from the specific programme 'Cooperation', theme 'Environment (including Climate Change)' under the 7th Research Framework Programme of the European Union		
Dissemination Level		
PU	Public	✓
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

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

Table of Contents

EXECUTIVE SUMMARY

INTRODUCTION

The EU BON Training program is a comprehensive program to benefit biodiversity data providers and data users, and enhance the overall impact of EU BON products and tools. The training program is specifically dedicated to data and metadata integration strategies (including registration of data with GEOSS), use of standards, and use of data tools developed or adopted by EU BON. It thus forms an important contribution to the long-term impact of the project.

This document reports on the 5 training events organized by Work Package 2 (Data integration and interoperability) during the EU BON project. Training events organized by other work packages and in collaboration with other initiatives are also referred to.

PROGRESS TOWARDS OBJECTIVES

A needs analysis identified 4 objectives for WP2: (1) *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; (2) Develop data integration and interoperability between the various networks, and with new generation of data sharing tools to enhance linking between observational data, ecosystem monitoring data, and remote sensing data; (3) 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; (4) Develop a new portal to enable fast access to EU BON integrated data and products by researchers, decision makers and other stakeholders.*

From this it became apparent that a comprehensive training program would greatly benefit biodiversity data providers and data users, and improve the overall impact of EU BON products and tools. Thus a 5th objective of WP2 was, namely the development of a training program to *Ensure global coordination of development efforts and adoption of the results: through international data interoperability task force, supported by a helpdesk infrastructure and training program.*

The EU BON training program has consistently worked towards the fulfilment of these objectives, particularly objective 5. Four training events, attended by 104 participants and trainers, had already been completed at the time of submission of this deliverable and one remained planned for October 2016.

Initially, training events promoted introductory topics for consortium members, gradually outreach was also extended to a more global audience, including non-consortium members and users from outside Europe, thus increasing data mobilisation and promotion of the EU BON network.

ACHIEVEMENTS AND CURRENT STATUS

Overall the training received very positive feedback. The topics were found useful, providing a good overview of existing infrastructures. The IPT tool, ARPHA, PlutoF and the GoldenGate Imagine tools were singled out as particularly useful. The practical format was also appreciated. Most of the participants found that the training experience will be useful for their work, and the hands-on training sessions were greatly valued.

FUTURE DEVELOPMENTS

The 5th training event has been postponed to October 2016, following an invitation by the International Long-term Ecological Research (ILTER) Network to give a talk, during the European Best Examples session, at the ILTER 1st Open Science Meeting, which will take place on 9-13 October 2016 in the Kruger National Park, South Africa. This meeting, with more than 500 participants, provides a unique opportunity to showcase the achievements of the EU BON project.

Furthermore, a continued outreach of training materials can be found on the EU BON training Helpdesk, accessible via the EU BON Portal.

1. INTRODUCTION

The EU BON Training program (Task 2.8) is dedicated to data and metadata integration strategies (including registration of data with GEOSS), use of standards, and use of data tools developed or adopted by EU BON. The target audience includes EU BON consortium members, but also data users and data custodians on a global level, thus contributing to the long-term impact of the project (DoW).

2. NEEDS ANALYSIS AND IDENTIFICATION OF MAIN OBJECTIVES

Work Package 2 focuses on data integration and interoperability. A needs analysis, consisting of: (1) interactions between consortium members (meeting in Brussels on 12 September 2014 attended by MRAC, RBINS, and UEF), (2) an online survey to identify training needs[1], and (3) observations made in the margin of the biodiversity data gap analysis (D1.1), identified 4 objectives:

1. *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,*
2. *Develop data integration and interoperability between the various networks, and with new generation of data sharing tools to enhance linking between observational data, ecosystem monitoring data, and remote sensing data,*
3. *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,*
4. *Develop a new portal to enable fast access to EU BON integrated data and products by researchers, decision makers and other stakeholders.*

From this pre-implementation analysis it became apparent that there was also a need for a comprehensive training program, which would benefit biodiversity data providers and data users, and improve the overall impact of EU BON products and tools.

The 5th objective of WP2 was the training program to *Ensure global coordination of development efforts and adoption of the results: through international data interoperability task force, supported by a helpdesk and training program.*

This deliverable reports on the 5 training events organized by Work Package 2 (Data integration and interoperability) during the EU BON project. At the time of writing, one training event was pending (October 2016). Training events organized by project partners, not strictly WP2, are referred to marginally with reference to the relevant

[1] https://www.surveymonkey.com/r/eubontraining

report (deliverable or milestone). All training events (WP2 and related) are considered in the recommendations and conclusions.

This activity aligns also with key actions of CBD towards the 2020 targets[2]: *"Strengthening and promoting the further mobilization of and access to data by, for example, encouraging the use of common informatics **standards** and protocols, promoting a culture of **data sharing**, investing in digitization of natural history collections and promoting citizen scientists' contributions to the body of biodiversity observations".*

3. IDENTIFICATION, ADAPTATION AND DEVELOPMENT OF TOOLS

To support data mobilization from biodiversity observation networks, the EU BON project adapted existing tools to handle and publish metadata, occurrence data and ecological data. In total, 30 data sharing tools used in the natural history domain were extensively evaluated (MS231).

In absence of proper guidance, the abundance and complexity of data standards and software tools designed for biodiversity information (Smirnova *et al*. 2016) may sometimes be perceived as an obstacle, rather than a resource for the diverse group of practitioners who wish to mobilize biodiversity information. For example, monitoring metadata and data differ from occurrence data, and all are defined by their own set of needs and subsequent parameters, not to mention the complexities related to taxonomic or geographic particularities.

Similarly, the diversity of end users (e.g. volunteer scientist, experienced biodiversity data manager or decision maker) demands an equally complex array of tools. User needs may be related to further analysis of the data, to enable summarized outputs or to facilitate publication.

While no tool can single-handedly address all user needs, their individual characteristics allow selection and presentation of the tools in the form of a work-flow (D2.2; Smirnova *et al.* 2016). Such a workflow can provide guidance, facilitating data mobilization and closing data gaps.

Selected tools are:

GBIF IPT[3]: Tool to publish and share biodiversity data sets and metadata through the GBIF network, with a focus on the new version of the GBIF IPT v.2.3. The IPT was adapted in collaboration with the EU BON to enable a workflow using sample-based data in support of Essential Biodiversity Variables (EBVs). It is also compatible with the DwC-A model.

Spreadsheet tools: 1) GBIF Spreadsheet processor is a web application that supports publication of biodiversity data to the GBIF network using pre-

[2] https://www.cbd.int/sp/targets/
[3] http://www.gbif.org/ipt

configured Microsoft Excel spreadsheet templates; 2) DataUp tool is the tool developed by DataOne to help environmental scientists to upload files to a DataOne repository for data management.

The ARPHA Publishing Platform[4]: Narrative (text) and data integrated publishing workflow to mobilize, review, publish, store, disseminate, make interoperable, collate and re-use data through the act of scholarly publishing.

TreatmentBank[5]: A platform to store, annotate, access and distribute taxonomic treatments and the data objects within.

Metacat[6] and Morpho[7]: Metacat is a repository that helps scientists store metadata and scientific data, particularly from ecology and environmental science, enabling query, search, analysis and effective use of the data sets managed or created by others. Morpho is an application designed to facilitate the creation and publishing of data and metadata to Metacat repositories.

PlutoF[8]: An online service and repository to create, record, manage, share, analyze and mobilize biodiversity data from a range of sources such as natural history collections and citizen science. This tool is developed by the University of Tartu Natural History Museum, dedicated to citizen science.

Training manuals and additional resources are available for each of the selected tools. **Annex 1** provides an overview of links and characteristics of existing manuals.

4. ORGANIZATION OF THE TRAINING EVENTS AND COLLABORATING NETWORKS

All training events were prepared and organized by the Task leader MRAC together with other WPs (WP1, 8) and DEST[9] (the Distributed European School of Taxonomy). The EU BON partners from WP1, WP2 and WP8 facilitated different sessions on data sharing and data publishing and hands-on sessions. Close collaboration with GBIF[10] (Global Biodiversity Information Facility), DataOne[11], GEO BON[12]/GEOSS[13] and LTER[14] (Long-term Ecological Research Network) contributed to the successful implementation of the training program, increasing outreach for EU BON products and tools.

[4] http://arphahub.com/
[5] http://plazi.xuul.org/resources/treatmentbank/
[6] https://www.dataone.org/software-tools/metacat
[7] https://www.dataone.org/software-tools/morpho
[8] https://plutof.ut.ee/
[9] http://taxonomytraining.eu/
[10] http://www.gbif.org/
[11] http://www.dataone.org/
[12] http://geobon.org/
[13] http://www.earthobservations.org/geoss.php
[14] https://lternet.edu/

5. EU BON TRAINING PROGRAM

5.1 TRAINING EVENTS

A comprehensive training curriculum has been developed with a focus on data and metadata integration strategies, use of standards and data sharing tools for institutional data and IT managers, researchers, citizen scientists and monitoring programs. All training events included presentations, demonstrations using case studies and hands-on sessions. Below a summary of each event is provided. **Figure 1** shows classes and interactions between participants and trainers at some of the EU BON training events. A detailed overview of all topics covered with a short description is given in **Annex 2**.

Figure 1: EU BON training events (photos by EU BON partners)

- The **1st training** was held on 3 April 2014, following the General Meeting of EU BON in Crete, Heraklion. This training event was focused on EU BON members and aimed to encourage a basic understanding and shared vocabulary on Information architecture and Data standards used by EU BON. Several topics were covered, including (1) Integrating biodiversity networks through Software Oriented Architecture, (2) Data standards: publishing sample-based data, (3) Information architecture - GEOSS perspective, (4) Data sharing and repositories in GBIF, (5) Data flow and modelling in virtual laboratory, and (6) Data sharing and repositories in DataONE.

- The **2nd training** was held on 19 March 2015, back-to-back with the joint CETAF/EU BON informatics workshops[15]. The event was organized by the University of Eastern Finland (UEF)[16] and Digitarium[17] (WP2 leader), in collaboration with CETAF ISTC and other EU BON work packages. The event was dedicated to data publishing using the new EU BON IPT. Additionally, introductory courses were given on the link with GEOSS/GEO BON, on the current state of EU BON software architecture to support the EU BON Biodiversity Portal, and on the Darwin Core[18] sample data model. The practical part of the training included the demonstration of the new IPT and some examples of sample-based data published by EU BON partners. It was followed by practical exercises where participants could publish data (test dataset or own dataset) using the test version of the tool.
- The **3rd training** was part of the training for GBIF Node Managers[19] organized by GBIF and held in Madagascar, Antananarivo on 4-5 October 2015. The event focused on the GBIF/EU BON IPT v.2.3 and publishing of sample-based data, with strong emphasis on hands-on training (both own data sets and use cases, representing a wide variety of database types) and the importance of metadata. Since the 2nd EU BON Training, these new IPT functionalities had been tested by the EU BON and GBIF consortium, particularly using data from the EU BON test sites. The two-day training[20] covered different aspects of data publishing, including (1) a review of the GBIF publishing landscape, (2) the various ways in which it is currently possible to publish biodiversity information and the existing mechanisms to publish complex datasets, (3) recent adaptations of the Darwin Core standard to accommodate data coming from sampling efforts, and (4) application of sample-based data and their importance for modern biology.
- The **4th training** was organized together with Pensoft, Bulgaria and held in Sofia on 22-23 March 2016. The location was chosen to promote EU BON and its products for data holders from East European countries. This event looked at core topics of data publishing to demonstrate various methods, platforms and workflows for managing, sharing and publishing biodiversity data, particularly species occurrences, sample-based and citizen science data. Special attention was paid to data publishing led by specialists from Pensoft Publishers[21]. The training session from Plazi[22] addressed semantic markup and treatments. Topics covered include, (1) the publishing of biodiversity data online using a new version of the IPT enabling publication of sample event datasets; (2) managing data with PlutoF workbench, also data from citizen science projects; (3)

[15] http://digitarium.fi/en/content/eu-bon-and-cetaf-joint-informatics-workshop
[16] http://www.uef.fi/en/etusivu
[17] http://digitarium.fi/en
[18] http://rs.tdwg.org/dwc/
[19] http://gb22.gbif.org/GB22
[20] http://community.gbif.org/pg/pages/view/47903/agenda-for-the-gb22-training-event
[21] http://www.pensoft.net/index.php
[22] http://plazi.org/resources/treatmentbank/

automated creation of data papers via ARPHA Publishing Platform, and its associated journals; (4) semantic markup and treatments via Plazi's TreatmentBank and GoldenGate workflows; (5) registration of data in GEOSS and introduction to EU BON biodiversity portal developments.

5.2 PARTICIPANTS

In total, 104 persons (participants and trainers) from 44 countries were involved in the EU BON training events (**Figure 2**). A list of all participants and trainers involved is given in **Annex 3**. Initially, training events promoted introductory topics targeting consortium members. Later on in the project, when tools had been identified (see MS 231, 232 and D2.2) and made available, the training events targeted a more global audience, including non-consortium members and users from outside Europe, thus increasing data mobilization and promotion of the EU BON network.

The target audience included biologists and other life scientists who are actively involved in monitoring and managing biodiversity data. Environmental agencies, biodiversity monitoring NGOs, volunteer groups etc. are also part of our target audience since EU BON has implemented the Citizen Science gateway and developed the CS oriented tools for data collection and management.

- The **1st training** was attended by 23 participants from 13 countries and from almost all WPs, and was facilitated by 6 trainers - experts from EU BON and from related Networks of Biodiversity Information.
- The **2nd training** was attended by 28 participants from the EU BON consortium and beyond (from 13 countries) and was facilitated by 5 trainers (all EU BON).
- The **3rd training** was attended by 43 participants from all around the world (34 countries) and was facilitated by 5 GBIF trainers and one invited trainer from the EU BON (Larissa Smirnova).
- The **4th training** was attended by 10 participants from across Europe, most coming from Bulgaria, and was facilitated by 10 trainers (all EU BON).

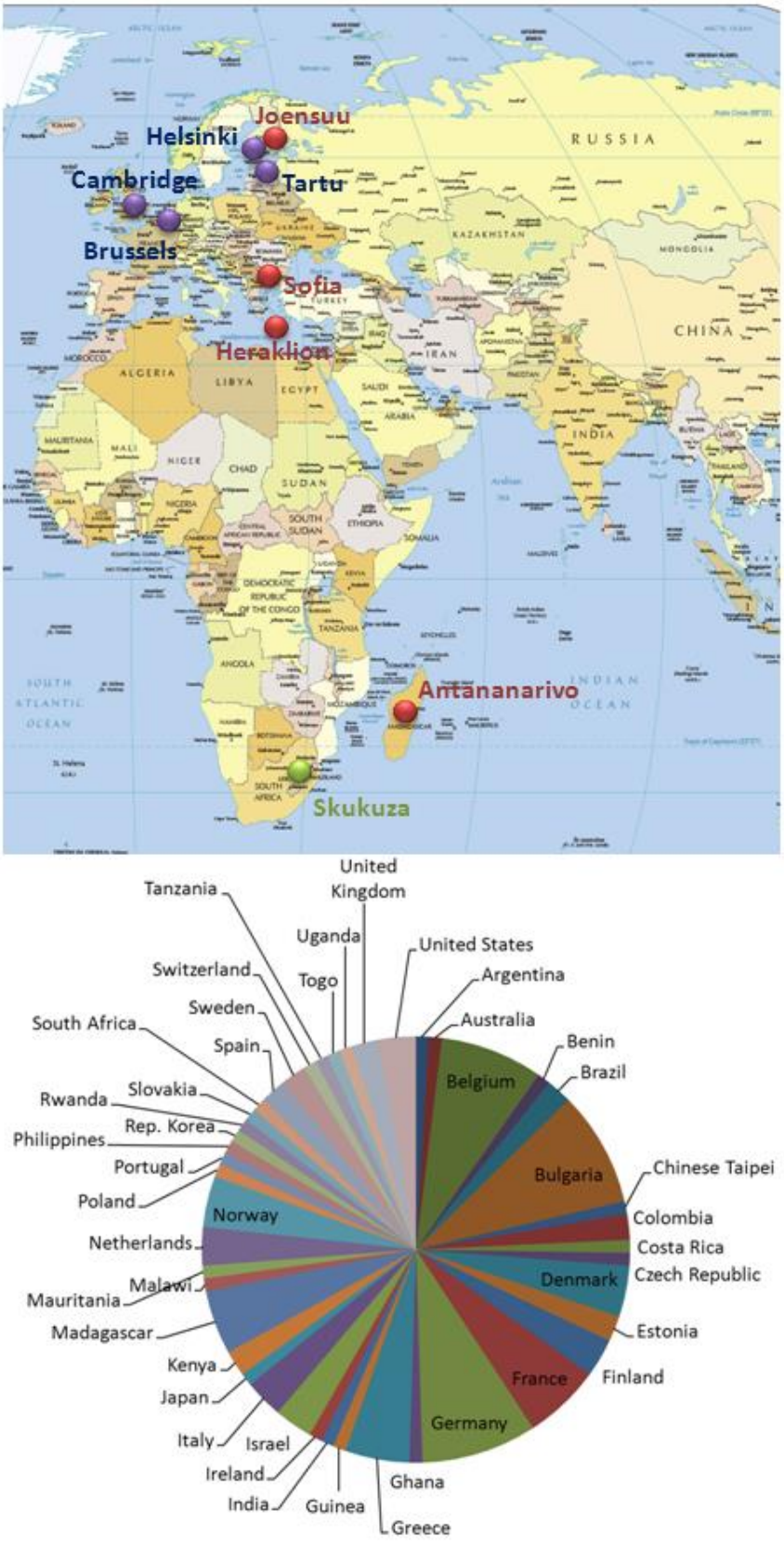

Figure 2: Geography of trainings and participants (Legend: Red – WP2 training events, Violet – related training events, Green – planned training event .

5.3 ADVERTISEMENT OF TRAINING EVENTS AND OUTREACH

On the Helpdesk site[23] a training page[24] was created where all information on past, current and planned training events is advertised and regularly updated. The announcement for each training event was simultaneously published on co-organizers' websites (e.g. GBIF or Digitarium) and registrations were handled using a web form on the Helpdesk platform.

Advertising consisted of:

- EU BON mailing list,
- A mailing list consisting of ~100 email addresses for researchers in Eastern European biological stations, nature reserves and monitoring programs compiled from the EuMon[25] repository,
- A mailing list consisting of the national focal points of EIONET[26] (European Environment Information and Observation Network) in Eastern Europe,
- Invitation through the Institute of Biodiversity and Ecosystem Research and National Museum of Natural History Museum, Bulgarian Academy of Sciences mailing lists to some 150 Bulgarian scientists,
- 20 contacts from European citizen science initiatives,
- EU BON test sites,
- Publication on the EU BON website[27], CORDIS[28], DEST[29], GBIF[30], CETAF, EU BON profiles in Twitter, Facebook, Google+ and LinkedIn
- The 4th workshop was also disseminated via the CORDIS Wire service[31] to ensure outreach to Europe-wide audiences.

Outreach and guidance for participants and the extended public after each training event is facilitated through the online training manuals and links, accessible via the Helpdesk (see also **Annex 1** for the links to the training manuals of selected tools).

5.4 FEEDBACK PROVIDED BY PARTICIPANTS

After the training events, feedback was collected from participants using Survey Monkey. Three trainings were evaluated (see the list of questions in **Annex 4**). The training in Madagascar was evaluated separately by the GBIF coordination team.

Below a summary of the feedback received for each training events is provided. For detailed feedback results the reader is referred to the Milestone reports.

[23] http://eubon.cybertaxonomy.africamuseum.be/
[24] http://eubon.cybertaxonomy.africamuseum.be/trainings
[25] http://eumon.ckff.si/
[26] https://www.eionet.europa.eu/
[27] http://www.eubon.eu/
[28] http://cordis.europa.eu/event/rcn/145116_en.html
[29] http://taxonomytraining.eu/content/biodiversity-data-sharing-and-data-publishing-workshop
[30] http://www.gbif.org/event/82602
[31] http://cordis.europa.eu/event/rcn/146435_en.html

- For the **1st training event** we received 16 responses from participants (roughly 50%) and 6 from the trainers (100%), giving high appreciation of the organisation and content of the courses provided.
- For the **2nd training event**, 14 people responded (50%), rating the courses as good (36%) to very good (43%). The participants stated that the objectives of the training were clearly defined and achieved. Most of the participants found that the training experience will be useful for their work (93%); 75% of participants were new to the material provided during the training course, 25% had used the IPT before. Most participants were familiar with the Darwin Core standard (50% advanced, 25% - intermediate and 25% – beginners).
- The **3rd training event** was evaluated by 37 respondents (86%), who gave a very high score for the courses on sample-based data (av. 4.3 / 5). Respondents underscored the importance of the new skills learnt, particularly with regards to sample-based data publishing. Interest was also high for data paper publishing and data cleaning tools. Certain aspects of data preparation before publication were also cited as being of high interest and an introduction to good data management practices would have been appreciated.
- The **4th training event** received a very positive rating from 6 respondents (60%). The IPT tool, PlutoF and the GoldenGate Imagine tools were singled out as very useful. For most participants, such tools were new and all mentioned its high potential for modern biologists. The suggestion was made to popularize PlutoF more, although it was not entirely clear how the software will be made available to members who are not from the Tartu University. The introduction to the DwC standard showed them the necessity of data standardization for further data integration in global networks. Participants also appreciated the networking possibilities. Other highlights included the EU BON Biodiversity Portal (beta release), demonstration of some tools in development and registration of data in GEO registry[32] system. Several participants struggled with the large number of acronyms. The EU BON/GBIF IPT (core and extension for sample-based data) was deemed user friendly and important by some respondents, although two disagreed. Experience with DwC ranged from advanced to beginners, with one participant never having heard of DwC before. Trainees expressed interest in taking part in the following topics and suggested to be addressed in the next EU BON trainings: (1) novel methods for publishing data; (2) legal concepts in relation to data management, publishing rights etc.; (3) bioinformatics and computational biology; (4) goals and positioning of EU BON in comparison to international Institutions like World Data Systems, planned cooperation with

32 http://geossregistries.info/index.html

local/national initiatives/institutions like GFBio[33], Pangaea[34]; and (5) other tools for data incorporation and data mining.

Overall the training received very positive feedback. The topics were deemed useful, providing a good overview of existing infrastructures. The practical format was also appreciated. Most of the participants found that the training experience will be useful for their work and the possibility of using their own data sets and work in the tools was greatly appreciated. Prior experience varied, with some participants having no previous knowledge at all, with others already possessing more advanced levels of literacy. Particular appraisal was also given to aspects of data preparation and best-practice for data management, as well as copyright issues and legal issues regarding biodiversity information management and publishing. **Figure 3** gives an illustration of participants' appreciation of the training events. In general, responses were overwhelmingly positive, with only 2 participants (2%) responding less positively for the 3rd training event. Average scores were (very) good.

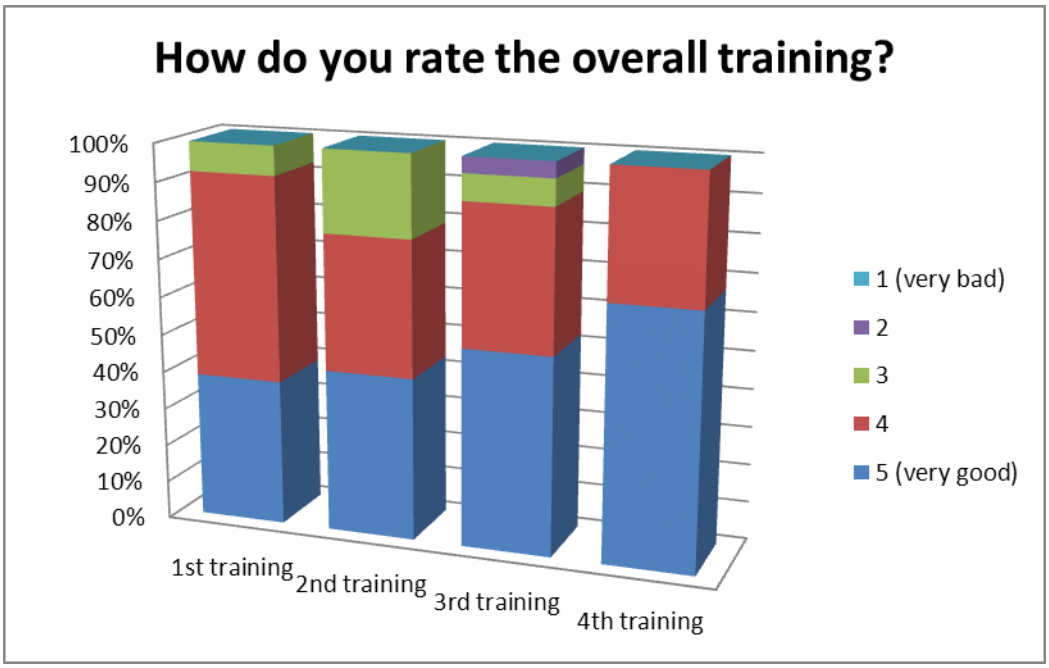

Figure 3: Respondents appreciation of the overall training event (SurveyMonkey feedback).

In general, the new IPT functionalities enabling the publication of sample-based data received a very positive reaction from participants in all training events where it was presented. Both scientists who sporadically share data sets and GBIF node managers who have extensive experience and knowledge of modern biodiversity data challenges, were impressed by the suitability of the tool. All participants were satisfied that with the knowledge gained during the training events on the new functionalities they will be able to reach a wider audience of data providers and publish more complex data than

[33] http://www.gfbio.org
[34] https://www.pangaea.de/

just occurrences. Many have clearly shown an intention to publish or re-publish datasets after the training course, especially the vegetation plots, which were frequently mentioned as a data type suitable for publishing with the new Event core.

6. OTHER RELATED TRAININGS

Due to the large demand from data providers, the training program was expanded by several smaller workshops and training sessions during different biodiversity-oriented events. Demand is explained by:

- The IPT tool is quite a recent tool (developed in 2009, Robertson *et al.* 2014) and there are still data providers using other, mainly distributed software such as BioCase, Digir, Tapir etc. (according to the last statistics: 163 IPT installations versus 290 other software types). Ultimately some training or self-learning is required to be able to use the IPT to publish biodiversity data. GBIF puts a lot of efforts in dissemination of the tool and training the staff and users (GBIF training courses, webinars, BID programme[35]).
- Even data managers from GBIF Nodes are not yet fully familiar with the extended functionalities of the IPT v.2.3 (e.g. enhancements allowing publication of sample-based data).
- The possibility to publish a dataset as a data paper makes the tool more attractive to data providers spurring additional interest for Pensoft developments (ARPHA Publishing Platform), in particular Data Publishing Toolkit.
- Demonstrations and try-outs for other EU BON products, for a wider audience, such as citizen science tools, are also in high demand due to growing CS involvement.
- The EU BON Biodiversity Portal also generates huge interest, but needs promotion and explanation. Even EU BON partners are often not familiar with all developments and products the project is offering, demanding additional in depth explanations and demonstrations.

Therefore, since the first introductory training, EU BON partners are regularly asked to participate in additional events, training workshops, seminars, etc. to give a presentation, demo or lead a hands-on training session. Below we summarize some of the more important related trainings.

- **Cambridge workshop:** Cambridge workshop on data sharing tools was held during the annual EU BON General meeting in Cambridge, UK on 1-4 June 2015. Several sessions were dedicated to data mobilization efforts and data sharing tools recommended by EU BON: (1) **Functionality and development of the data publishing and dissemination toolbox:** The aim of the workshop is to

[35] http://www.gbif.org/programme/bid

identify data publishing formats, metadata standards, data paper templates and workflows that will provide a scholarly data publishing venue for the main data types available via the EU BON's European Biodiversity Portal and serve the general objectives of EU BON to facilitate and increase data mobilization and interoperability; (2) **Demonstration of GBIF/EU BON IPT for monitoring networks:** The next version of GBIF IPT was demonstrated, with features to share sample-based data. The newly ratified Darwin Core terms for sample-based data was explained; (3) **Data Mobilization Examples:** Showing examples for EU BON data mobilization efforts – ranging from tools for specimens and observation data to strategies and approaches at the test sites. Presentation on DINA virtual appliances; (4) **The Citizen Science Gateway and Citizen Science Approaches in EU BON:** demo of PlutoF tool and Mobile App "I saw a butterfly" on training. Including also the (a) CS tools repository on the Helpdesk, (b) Best practice cases: Norwegian/Swedish citizen science system Species Observation System, COMBER and Anymals and Plants, Eastern Europe and Data Mobilisation, and (c) hands-on training; (5) **Developing tools to prepare, extract and mine published legacy & prospective biodiversity literature: status quo:** The workshop provided a recapitulation of the goal of this task, the use cases under discussion, the data types, the planned and implemented workflows and integration into the EU BON data portal as well as the status quo of the tools. The meeting was attended by a total of 85 participants with various organizational background and relation to EU BON. Among these were almost all EU BON alongside representatives of eight associate partners and many guests.

- **Citizen Science workshop:** The workshop on the final version of the EU BON citizen science gateway (MS154) was held on 27-28 June 2016 in Tartu, Estonia. The workshop focused on Eastern and Central European stakeholders, as previous work in EU BON had identified the gaps in biodiversity observation data in this region (D1.1). During the two days, the participants listened to presentations about Estonian citizen science initiatives and European best practice cases, engaged in training to use EU BON tools for citizen science data management and learned the framework of the citizen science gateway. Also there was a world café session to clarify the needs of stakeholders for a pan-European citizen science gateway. There were 34 participants from 19 institutions and organisations and one person not associated with any organisations. The following topics were presented: (1) **Citation list of CS data providers:** contains information about CS-based biodiversity observation data providers. The list is managed via the PlutoF publication module[36] and populated on the Biodiversity Portal[37]; (2) **Directory of citizen science tools:** has information about available CS tools for biodiversity data management, project management, publication etc. and is administrated via the EU BON portal CMS

[36] https://plutof.ut.ee/#/reference/reference-index
[37] http://beta.eubon.ebd.csic.es/web/citizen-science/directory-of-cs-data-providers

(content management system); (3) **Guide for citizen science project management:** step-by-step guide for managing biodiversity monitoring citizen science projects. The guide will lead users to the according PlutoF citizen science module for project management; (4) **Other elements:** news about CS related events, important software releases; description of best practice cases etc.; (5) **PlutoF citizen science module:** this module acts as a separate workbench and users can create and manage their biodiversity observation projects. The database is also used as source for some content of the EU BON portal and is integrated with citizen science mobile tools (butterfly sightings app by GlueCAD and animal sound recording app by University of Tartu; (6) **World café group discussion on recommendations for CS gateway.**

- **Belgian Biodiversity Platform workshop on data quality:** MRAC participated in the Data Quality workshop[38] organized by Belgian Biodiversity platform in Brussels on 9 November 2015. L. Smirnova gave a presentation on the IPT and publishing of sample-based data. The workshop was attended by 20 participants from different Belgian institutions and biodiversity-related initiatives.

- A short **half-day training course** on sample based data was given by P. Mergen (EU BON partner, MRAC) to 6 African colleagues from the Biodiversity Monitoring Center CSB from Kisangani, DRC on 15 December 2015.

- **DataONE Member Node Implementation Workshop: A total of 8 EU BON representatives** participated in this workshop held during the Open Repositories 2014 conference, in Helsinki, Finland, 9-13 June 2014. The attendees engaged in the discussion to find ways how the EU BON project and its data could be linked to the DataONE Infrastructure. The workshop (1) provided an overview of the Data Observation Network for Earth project (DataONE), (2) explained the benefits for groups and institutions of collaborating with DataONE as a Member Nodes, (3) presented different ways to participate as a Member Node, (4) discussed the process of becoming a Member Node, and (5) demonstrated how to use the DataONE web services to access content from client applications. At the end of the workshop, participants understood the design of DataONE, the services that DataONE provides to its Member Nodes and the users of those Member Nodes, and the technical information needed to establish a Member Node at their organization. The target audience was information managers and technical staff at organizations that are interested in becoming DataONE Member Nodes or that have started the process of becoming a DataONE Member Node.

- Several follow-up **training events** have been organized by **GBIF**, using the same structure and resources developed for 3rd training event (GBIF-EU BON

[38] https://etherpad.net/p/DataCleaningPublishingworkshop

workshop in Madagascar) to promote the **IPT** tool and it's new functionalities[39]. For more information see GBIF training page[40].

7. PLANNED EVENTS

The **5th training** has been postponed to October 2016, following an invitation by the International Long-term Ecological Research (ILTER) Network to give a talk at the ILTER 1st Open Science Meeting, which will take place on 9-13 October 2016 in the Kruger National Park, South Africa[41], during the European Best Examples session. This meeting, with more than 500 attendees, provides a unique opportunity to showcase the achievements of the EU BON project. Opportunity will also be created to have a demonstration area and hands-on training for interested participants. The ILTER community has been identified as an important target audience for collaboration and joint data mobilization efforts. Most ILTER members are national or regional networks of scientists engaged in long-term, site-based ecological and socio-ecological research (known as LTER or LTSER). In D2.2 an agreement between EU BON and ILTER was reported to collaborate on sharing metadata among EU BON and LTER tools and sites (**Figure 4**). Integration of DEIMS in the EU BON registry, extraction of metadata from DEIMS are important topics in this collaboration.

A **webinar** may also be held before the end of the EU BON project, provided sufficient interest is generated during the ILTER meeting, to further increase outreach beyond the EU BON consortium.

A final training event will be organized during the Final EU BON meeting in 2017. Updates on the last developments, upgrades, state of art on the portal and tools will be addressed during the training planned to be part of the last EU BON general meeting in Brussels/Meise, Belgium (March 2017).

Other outreach and dissemination paths may include the European tropical ecology conference in Brussels, Belgium in February 2017. Organization of additional training sessions depends strongly on initiative of the involved partners and availability of funds.

[39] http://www.gbif.pt/EuropeanNodesMeeting/workshop
[40] http://www.gbif.org/capacity-enhancement/training
[41] https://www.eiseverywhere.com/ehome/ilter2016/373608/

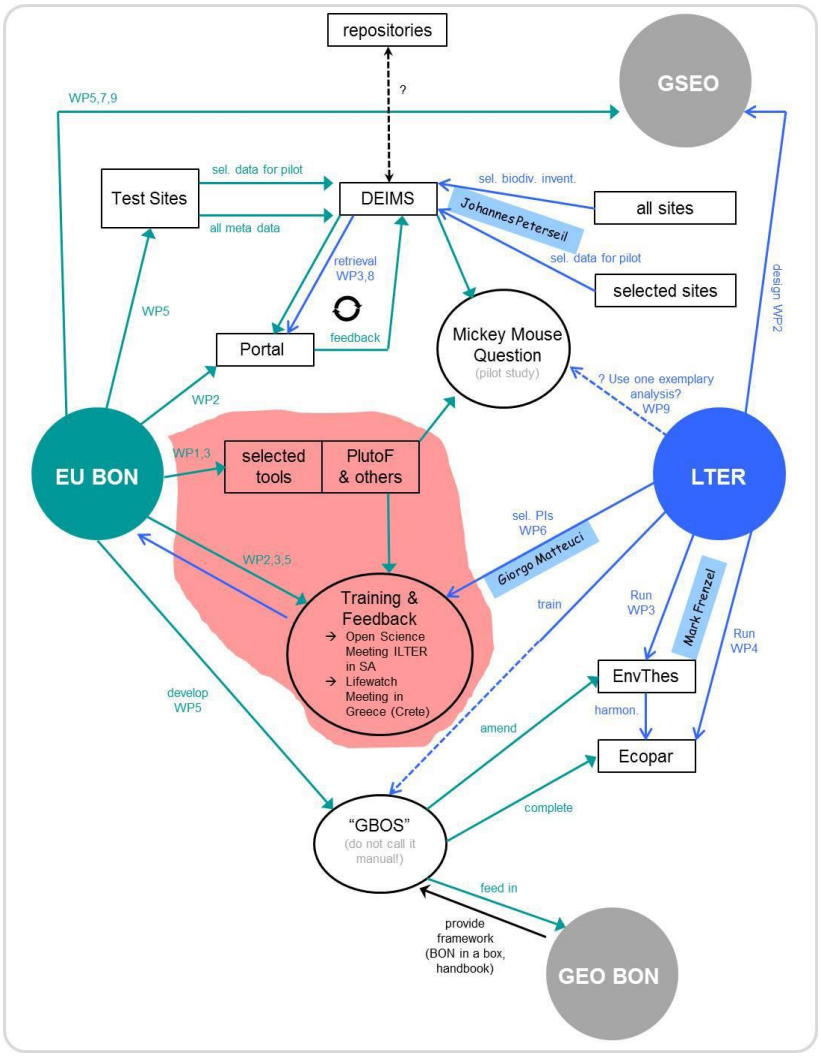

Figure 4: Information flows between EU BON and LTER Europe, as envisaged on the 3rd EU BON Stakeholder Roundtable in Granada (Spain) on 9-11 December 2015.

8. LESSONS LEARNED AND RECOMMENDATIONS

The training courses have shown that:

- There is a demand from content providers,
- The new functionalities and tools introduced during the training events enable publication of more diverse datasets, with an extended range of data providers and improved data quality published to GBIF,
- Technical documentation and support are important prerequisites for successful implementation of these new developments
- Provision of online training materials prior to the event is seldom used by participants to prepare. Afterwards these materials are appreciated, however.

To facilitate the organization of the training events and maximize their impact we would like to make the following recommendations:

1. Advertise widely: Use existing platforms such as GBIF, CETAF and DEST for better outreach and a greater audience. Integrated into the EU BON Biodiversity Portal, the Helpdesk Infrastructure will be another important platform for dissemination of training news.
2. It is easier to attract more applicants when the training is held in conjunction with a conference or international meeting; this reduces costs for participants, which might be a particular constraint for participants from less wealthy countries. Partial support and provision of a list of funding opportunities might also be beneficial for future trainings. Another contributing factor for the sometimes low application rate might be that the objectives of the training were fairly broad and not targeted to a specific participant profile.
3. Reuse, adapt and improve existing resources, to save time and effort, and learn from the experiences of others. The EU BON Helpdesk provides: (1) all training materials used during the EU BON training events, (2) a list of potential trainers, (3) an EU BON trainer mailing list to facilitate the organisation of a training event or to request a presentation. A wide range of training material is available on the GBIF website[42].
4. Provide a detailed curriculum.
5. A pre-course assessment of the participants' background and level (IT, biodiversity information, taxonomic…) can be built into the registration form. This could ensure better involvement of the participants and ease uptake of shared knowledge, by adapting the curriculum to the level of participants. Participants' expectations can also be gauged pre-event and afterwards, through the feedback form, evaluation of participants' satisfaction can be gauged.
6. Providing an easily accessible, multilingual resource point for training materials will be appreciated by participants, who are actually interested in using the tools after the training event.
7. Combine theoretical and practical sessions. Demonstrations and presentations that take too long are less effective than hands-on exercise. But beware that practical sessions demand more personal coaching. More than one facilitator is desirable.
8. Although it is rewarding for participants to use their own data sets during the training, it is more practical to use sample data sets. Raw data poses too many time-consuming practical issues.
9. Consider using e-learning to reach a wider audience. See examples of e-learning in the GBIF community[43].

[42] http://www.gbif.org/resources
[43] http://www.gbif.org/capacity-enhancement/training#elearning

During the training events and the discussions thereafter, it was emphasized that because we are at the beginning of the process and because of the complexity of data, it is not our intention to provide a single solution to map all data to the DwC. Rather, it is our intention to help participants understand the star scheme principle of DwC, the use of Event core, and the choice of appropriate extensions. Over time, and with an increasing amount of published sample data, we will be able to evaluate it and give better recommendations in the form of best practices for different data types.

Future training events might benefit from a focus on DwC mapping, taking as an example the most popular monitoring protocols such as vegetation plots, marine sampling protocols, freshwater surveys, bird surveys etc.

The IPT is a powerful tool for data sharing. So, promoting IPT with hands-on training sessions, providing sample data sets for pre- or post-training exercises, will empower users' insights and skills.

The IPT training could be combined with the introduction and practical session on data paper publishing using Pensoft Publishing Toolkit. Thus enabling a deeper understanding of the importance of metadata, and increasing attractiveness of these tools, particularly for monitoring programs with well-designed protocols and descriptions. Interactions and interoperability between different metadata standards and repositories could also be addressed.

9. SUSTAINABILITY

Activities that will further contribute to the sustainability of the EU BON training package include:

- **GBIF** has recently (2015) launched the **BID programme**[44] (Biodiversity Information for Development). This EU-supported, multi-year funding opportunity for data mobilization is aimed at developing countries in sub-Saharan Africa, the Caribbean and the Pacific (ACP). Different capacity building activities will contribute to data mobilization, especially data that supports ACP policy needs with regards to protected areas, threatened species and invasive alien species. The GBIF BID program includes online and on-site training events, workshops. The program will be co-mentored by staff associated with EU BON partners, including MRAC and NBGB.
- The Distributed European School of Taxonomy (DEST[45]) and the Learning Workgroup of CETAF[46], BeBIF (the Belgian node of GBIF), are already established **partners** with EU BON and will maintain training materials and help continued training outreach for EU BON beyond the end of the EU BON project. These initiatives will be supported by the **Helpdesk infrastructure** at MRAC.

44 http://www.gbif.org/programme/bid
45 http://taxonomytraining.eu/
46 http://cetaf.org/

- **Future collaboration** with initiatives such as the learning components of ESFRI[47], future EU projects in Earth Observation, EEA[48], training programs of LifeWatch[49], ILTER[50] and DARIAH[51] can also be envisioned.

10. REFERENCES, FURTHER READING AND TUTORIALS

- Giuliani G, Lacroix P, Guigoz Y, Roncella R, Bigagli L, Santoro M, Mazzetti P, Nativi S, Ray N, Lehmann A (2016). Bringing GEOSS Services into Practice: A Capacity Building Resource on Spatial Data Infrastructures (SDI). Trans. in GIS. doi:10.1111/tgis.12209
- Robertson T, Döring M, Guralnick R, Bloom D, Wieczorek J, Braak K, et al. (2014). The GBIF Integrated Publishing Toolkit: Facilitating the Efficient Publishing of Biodiversity Data on the Internet. PLoS ONE 9(8): e102623. doi:10.1371/journal.pone.0102623
- Smirnova L, Mergen P, Groom Q, De Wever A, Penev L, Stoev P, Pe'er I, Runnel V, Camacho A, Vincent T, Agosti D, Arvanitidis C, Bonet F, Saarenmaa H (2016). Data sharing tools adopted by the European Biodiversity Observation Network Project. Research Ideas and Outcomes 2: e9390. doi: 10.3897/rio.2.e9390
- Wetzel FT, Saarenmaa H, Regan E, Martin CS, Mergen P, Smirnova L, Ó Tuama É, García Camacho FA, Hoffmann A, Vohland K, Häuser CL (2015) The roles and contributions of Biodiversity Observation Networks (BONs) in better tracking progress to 2020 biodiversity targets: a European case study, Biodiversity, DOI: 10.1080/14888386.2015.1075902
- Wieczorek J, Braak K (2015). The GBIF Integrated Publishing Toolkit User Manual, version 2.3. Copenhagen: Global Biodiversity Information Facility.

47 http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri
48 http://www.eea.europa.eu/
49 http://www.servicecentrelifewatch.eu/lw-training
50 https://www.ilternet.edu/
51 http://be.dariah.eu/

11. ANNEXES

ANNEX 1: TRAINING MANUALS FOR SELECTED TOOLS

For a comprehensive overview of data sharing tools and guidelines the reader is referred to the EU BON Biodiversity Portal (http://beta.eubon.ebd.csic.es/tools). Below are the links to training manuals for EU BON selected tools.

Training manuals for EU BON selected tools:

GBIF IPT Tool to publish and share biodiversity data sets and metadata through the GBIF network, with a focus on the new version of the GBIF IPT v.2.3. The IPT was adapted in collaboration with the EU BON to enable a workflow using sample-based data in support of Essential Biodiversity Variables (EBVs). It is also compatible with the DwC-A model.

IPT manual v.2.3

Sample Event Data

Spreadsheet tools:

1) GBIF Spreadsheet processor is a web application that supports publication of biodiversity data to the GBIF network using pre-configured Microsoft Excel spreadsheet templates;

GBIF resources

GBIF Spreadsheet Templates

User Guide

2) DataUp tool is the tool developed by DataOne to help environmental scientists to upload files to a DataOne repository for data management. The DataUp tool has merged with Dash. Dash is a University of California project to create a platform that allows researchers to easily describe, deposit and share their research data publicly.

Best practices database created by DataOne

The ARPHA Publishing Platform: Narrative (text) and data integrated publishing workflow to mobilize, review, publish, store, disseminate, make interoperable, collate and re-use data through the act of scholarly publishing.

Arpha Pensoft Guidelines

TreatmentBank: A platform to store, annotate, access and distribute taxonomic treatments and the data objects within. GoldenGATE is an editor which allows the creation of new XML content from plain text, html or pdf data.

GoldenGate manual

Metacat: A repository that helps scientists store metadata and data, search, analyse and effectively use the data sets they manage or those created by others. A data provider using Metacat can become DataONE member node with a relatively simple configuration.

MetaCat Administrator's Guide

Morpho Metadata Editor: Application designed to facilitate the creation of metadata so that scientist can easily locate and determine the nature of a wide range of data sets. It interfaces with the Knowledge Network for Biocomplexity (KNB) Metacat server.

MorphoUserGuide (pdf)

PlutoF: An online service and repository to create, record, manage, share, analyse and mobilize biodiversity data from a range of sources such as natural history collections and citizen science This tool is developed by the University of Tartu Natural History Museum, dedicated to citizen science.

PlutoF video tutorials

PlutoF manuals

ANNEX 2: LIST OF PRESENTATIONS

The presentations and related resources are available for downloading from the EU BON Helpdesk:
http://eubon.cybertaxonomy.africamuseum.be/training%20materials

1ST TRAINING

Integrating biodiversity networks through Software Oriented Architecture (Francisco Antonio García Camacho)

In this training session we will introduce the concepts of Software Oriented Architecture, Business Process Modelling and Enterprise Application Integration, analysing their relevance with EU BON architectural design (D 2.1) and presenting different ways to accomplish the integration of biodiversity networks and other data sources. We will present a demonstration of a working system that will integrate several data sources through and Enterprise Services Bus.

Data standards: publishing sample-based data using the GBIF Integrated Publishing Toolkit (Eamonn Ó Tuama)

The GBIF Integrated Publishing Toolkit (IPT) is the recommended application for publishing data to the GBIF network. To date, it has been used to publish three types of data: taxon occurrences, checklists and data set level metadata. In this session, we will explore its adaptation for publishing sample based data. First, we will review the essential attributes of sample data that need to be captured. Then we will introduce the Darwin Core Archive data format, explain the constraints imposed by its star schema, relational data model, and address the requirement for additional terminology in the Darwin Core vocabulary to describe the attributes of sample data together with controlled value vocabularies for some attributes. A prototype of the IPT adapted for sample data will be demonstrated and participants are encouraged to test it with their own data sets.

Information architecture – GEOSS perspective (Lorenzo Bigagli)

The session aims at introducing the architecture of the Global Earth Observation System of Systems (GEOSS), the GEOSS Common Infrastructure (GCI) and the GEOSS Brokering Framework. GEOSS has been created by an international voluntary effort that connects geospatial, Earth Observation and information infrastructures, acting as a gateway between producers of environmental data and end users. GEOSS aims at enhancing the relevance of Earth Observation and at offering public access to comprehensive, comprehensive, and sustained near-real time data, information and analyses of the environment. The GCI allows the user of Earth observations to access, search and use the data, information, tools and services available through GEOSS. The GEOSS Brokering

Framework implements multi-disciplinary interoperability and lower entry barriers for both users and data providers, allowing them to continue using their tools and publishing their resources according to their standards.

The session includes a live, interactive demonstration of the GEOSS Discovery & Access Broker, based on material from the "Bringing GEOSS services into practice" workshop[52], which session attendants may practice on their own computer.

Data sharing and repositories in the GBIF network (Tim Robertson)

The GBIF network is diverse, spanning more than 500 institutions and connecting thousands of databases using a variety of protocols and tools. The key components of the network include the data publishing repositories, a central coordinating registry and a sophisticated search index, which supports the GBIF portal[53] - itself consider a data repository in wider networks. During this session a live demonstration of data sharing between repositories will be given, during which the architecture of the network will be described. An installation of the GBIF Integrated Publishing Toolkit (IPT) which acts as a data publishing repository will be used to demonstrate the services of the GBIF registry component, and specifically the management of data profiles (standards) available to data publishers. A dataset will be mapped, and registered with GBIF. Crawling components will be alerted automatically, and the data will be indexed and made available for discovery and access through the GBIF portal and web services API. Some observations about this architecture will be offered, including the opportunity to collaborate with the EU BON partners to improve data security through redundant storage.

This session is targeted for people interested in the GBIF architecture and key components, the data flows within the GBIF network, the GBIF publishing tool and those interested in interfacing with GBIF through the portal web services API. Being a live demo, opportunity will be given to address questions along the way, with the overarching goal that participants leave with a better understanding of the data flows than before the session.

Data flow and modelling in virtual laboratory (Hannu Saarenmaa)

The Biodiversity Virtual e-Laboratory, BioVeL, addresses research challenges by having scientists and computer engineers working together to develop tools for pipelining data and analysis into efficient analytical pipelines, called "workflows." Workflows are complex digital data manipulations and modelling tasks that execute sequences of web services. BioVeL designs and deploys such workflows for a selected number of important areas in systematic, ecological, and conservation research, e.g. for the analysis of data sets with ecological, taxonomic, phylogenetic, and environmental information.

[52] https://www.unige.ch/tigers/en/enseignements/geossinpractice/
[53] http://www.gbif.org/

BioVeL data refinement and ecological niche modelling workflows allow researchers to (i) explore, access, refine, and format large data sets from major data providers; (ii) combine disparate data sets with the researchers' own data; and (iii) run complex and computationally intense analytical cycles. (iv) generate comparative maps of species distribution.

The training workshop will demonstrate use of the informatics tools and services developed by the BioVeL project to address research topics such as historical analyses, invasive species distribution modelling, endangered species distribution modelling, and dynamic modelling of ecologically related species, and Essential Biodiversity Variables. In particular, there will be introduction to the BioVeL e-infrastructure and portal. Examples of taxonomic data cleaning, ecological niche modelling, model testing, statistical analysis of GIS data, invasive and endangered species distribution modelling, and historical comparison biodiversity from museum collections will be shown.

Data sharing and repositories in DataONE (Bruce E. Wilson)

The mission of the Data Observation Network for Earth (DataONE) is to enable new science and knowledge creation through universal access to data about life on earth and the environment that sustains it. Organizations that collect, manage, or distribute data relevant to the Earth and the environmental sciences can collaborate with each other and with DataONE by becoming Member Nodes in DataONE. This collaboration brings broader exposure to the organization’s holdings, tools for end-users to more directly access and use data (the DataONE Investigator Toolkit), and tools to assist the organization with their preservation and curation missions. DataONE also makes available a wide range of educational materials and best practice guides for community use in data management education and has conducted sociocultural studies on the barriers and enablers for improved data sharing. This talk will provide an overview of DataONE, highlight the sociocultural and technical approaches used by DataONE to enable data sharing and data interoperability, and explore ways that DataONE and other projects can collaborate with each other.

2ND TRAINING

Introduction to GEOSS, GEO BON, EU BON (Hannu Saarenmaa):

The presentation covers origins, organisation, and current plans of the Group on Earth Observation (GEO), the goals and information architecture of the Global Earth Observation System of Systems (GEOSS). GEOSS Portal, registry system, data sharing principles, available data, and brokering mechanisms is explained. Participation in the GEOSS Architecture Implementation pilot process is discussed. The GEO Biodiversity Observation System (GEO BON) and its aims for information management were also covered.

Information architecture of EU BON (Francisco Antonio García Camacho):

The presentation covers the current update or the software architecture that will support the EU BON Biodiversity Portal, focusing on the brokering alternatives and data/metadata sharing standards (EML, OGC-CSW, SOAP and REST interfaces...). GI-cat is introduced as a brokering tool to integrate new data sources. This session also includes a demonstration of the first implementation prototype of the EU BON Biodiversity Portal.

Data standards, Darwin Core and extensions for sample-based quantitative data (Éamonn Ó Tuama):

The Darwin Core vocabulary, extended with a small number of additional terms, can be used in a Darwin Core Archive to encode information from sample-based data sets, i.e., data sets associated with environmental, ecological, and natural resource investigations. Such data are usually quantitative, calibrated, and follow certain protocols so that changes and trends in populations can be detected. This session introduces the new terms, the star schema model underlying Darwin Core archives consisting of a core table linked to one or more extension tables, and the associated enhancements to the GBIF Integrated Publishing Toolkit (IPT) to support publishing of sample-based data.

Demonstration of GBIF/EU BON IPT for monitoring networks (Larissa Smirnova and Franck Theeten):

The Integrated Publishing Toolkit (IPT) is a software tool developed by GBIF, aiming to facilitate the sharing and publishing of biodiversity data on the Internet using the GBIF network. It uses the Darwin Core standard to map species occurrence datasets and checklists, and can also handle data from natural sciences collections or observations. Since additional, sample-based terms are added to the DwC vocabulary, the IPT tool can be used by various monitoring networks collecting mainly quantitative data (environmental, ecological, and natural resource investigations). The practical part of the training will present the IPT tool, and explain how to publish your dataset using the IPT tool, via a practical example using occurrence data. Extensions of the IPT will be presented and issues will be discussed after the training.

Practical exercise with sample dataset and with own data (Larissa Smirnova and Franck Theeten)

Hands-on-training

Presentation of Israel Peer on example of sample-based data

No abstract

3RD TRAINING

Introduction to sample-based data publishing (Larissa Smirnova)

In this session we will address the recent adaptation of the Darwin Core standard to accommodate data coming from sampling efforts. We will introduce the changing landscape of natural sciences and the wide variety of the sample-based data produced these days. We will discuss the possible application of sample-based data and their importance for modern biology. Several example use cases illustrating different types of sample data will be presented, including a set of them that we will use later in the practical work.

The session facilitators will present an introduction though a presentation containing these elements:

- The definition and types of sample-based data
- The use of sample-based data
- How to express sample-based data using the Darwin Core model
- Enabling discovery and access to sample-based data via IPT
- Examples of sample-based datasets
- Presentation of the use cases

Data publishing practice in groups (Alberto González-Talaván, Danny Vélez, Larissa Smirnova, Laura Russell, Mélianie Raymond and Nicolas Noé)

Practical session based on real-life use cases in two parts: one exercise on defining strategies to promote data publishing and a second one on publishing data using IPT. This session has a 100% focus on practical work. First, we will have the opportunity to apply and share our experience in the promotion of data publishing by designing a draft strategy on how to approach a new community. Second, we will do some real data publishing using the latest version of the IPT and using extensions. Both exercises will be articulated around realistic use cases designed to match the interests of the different work groups.

Sample-based data publishing practice in groups (Alberto González-Talaván, Danny Vélez, Larissa Smirnova, Laura Russell, Mélianie Raymond and Nicolas Noé)

Practical session focusing on sample data publishing and its promotion among the communities producing this kind of data. This will bring specific challenges both in terms of the different communities to address, and because of the particularities of this kind of data. The work will be based on realistic use cases, that in this case they have been developed around different methodologies of data sampling.

4TH TRAINING

The data publishing landscape, gaps and mobilization efforts (GBIF/EU BON)(Smirnova Larissa)

How to make vegetation sampling data accessible on GBIF.org (Kyle Braak)

Scientists can now make their sampling data accessible on GBIF.org in order to enhance accessibility for other researchers and show a commitment to open access and reproducibility that are integral to scientific inquiry. GBIF.org is the world's largest source for species occurrence data, providing free and open access to more than 640 million occurrences from more than 15,000 datasets published by nearly 800 institutions. Its near real-time infrastructure is widely used, too, currently averaging more than one substantive use in peer-reviewed research per day. Over the past two years, however, the GBIF Secretariat has been working with EU BON partners and the wider biodiversity informatics community to enable sharing of “sample event datasets”. These data derive from environmental, ecological, and natural resource investigations that follow standardized protocols for measuring and observing biodiversity.

GBIF.org could not provide this type of data previously due to the complexity of encoding the underlying protocols in standardized ways. In March 2015, TDWG, an international body responsible for maintaining standards for the exchange of biological data, ratified changes to Darwin Core that enable support for mobilization of sample event based data, in particular species abundance. Then in September 2015, GBIF enhanced a new version of the Integrated Publishing Toolkit, or IPT (its free open-source data publishing software) that enables publication of sample event datasets and updated GBIF.org with enhanced indexing and discovery of these datasets. The purpose of this presentation is to highlight that GBIF now supports sample event datasets, and to explain how scientists can share their datasets freely and openly through GBIF.org using the IPT. As an example, the presentation will demonstrate how vegetation plot surveys exported from TurboVeg gets converted into the new Darwin Core sample event format.

Practical demo on how to prepare and map different types of data (Larissa Smirnova)

You have your observation data in a spreadsheet or database and you want to publish it at GBIF? We will demonstrate it using the IPT and Darwin Core Archive extensions mechanism. All steps between will be demonstrated and discussed.

The Integrated Publishing Toolkit (IPT) is a software tool developed by GBIF, aiming to facilitate the sharing and publishing of biodiversity data on the Internet using the GBIF network. It uses the Darwin Core standard to map species occurrence datasets and checklists, and can also handle data from natural sciences collections or observations. Since additional, sample-based terms are added to the DwC vocabulary, the IPT tool can be used by various monitoring networks collecting mainly quantitative data (environmental, ecological, and natural resource investigations).

Data publishing practice in groups (Larissa Smirnova).

This session will focus on practical work. You will do some real data publishing using the latest version of the IPT and using realistic use cases (or own data).

Registration of data in GEOSS. EU BON registry and biodiversity portal review (Francisco Antonio García Camacho)

The session aims at introducing the user the last architectural updates in GEOSS Common Infrastructure (GEOSS GCI), focusing in particular on the GEO Discovery and Access Broker (GEO DAB) and its relation to earth observation and biodiversity networks.

The session will also introduce GI-cat, the message broker that acts as the main subsystem in the core of the GEO DAB and the EU BON metadata registry as well, harvesting datasets and storing their metadata using standardised models. The last part of the session will introduce the EU BON Biodiversity Portal (beta release), exploring some tools currently in development.

The Data Publishing Toolkit at EU BON: Automated creation of data papers, data and text integrated publishing via the ARPHA Publishing Platform (Teodor Georgiev and Viktor Senderov)

The session aims at introducing the notion of data publishing by some newly developed data publishing workflows via the ARPHA Writing Tool, ARPHA Publishing Platform, and its associated journals: Biodiversity Data Journal (BDJ), Research Ideas and Outcomes (RIO), and One Ecosystem. ARPHA is an innovative publishing solution developed by Pensoft that supports the full life cycle of a manuscript, from authoring and reviewing to publishing and dissemination. The data publishing strategy of ARPHA aims at increasing the proportion of structured text and data within the article content, so as to allow for both human use and machine readability to the maximum possible extent.

Biodiversity Data Journal is one of the most innovative and technologically advanced open access scholarly journals in biodiversity science. Since its launch in September 2013 more than 260 articles have been published in BDJ and the interest of the scientific community is continuously growing. BDJ and the ARPHA Writing Tool associated to it are core elements of EU BON's Data Publishing Toolkit.

A key element of the training will be to demonstrate how data paper manuscripts can be created at the click of a button by importing EML metadata from GBIF IPT, DataONE and LTER, as well as online import of occurrence records into manuscripts for BDJ from GBIF, BOLD, iDigBio and PlutoF, export of published specimen records in Darwin Core Archive and GBIF, and treatments to Plazi.

The session includes a live, interactive demonstration of the functionalities of both, ARPHA and BDJ, which attendants may practice on their own computers.

Managing citizen science projects with PlutoF workbench (Veljo Runnel and Allan Zirk)

You want to start a citizen science project but you are not so sure how to manage mass of observations from your project participants? You want to engage experts to evaluate observations based on photos or sound recordings but lack communication tools? You need to export your project data as Darwin Core compliant spreadsheet?

PlutoF workbench offers tools to design observation data forms, moderate observations, publish project data, etc. In addition, your observation data can be linked to GBIF - a respected biodiversity data hub.

During the demo session you can get the basics for biodiversity related citizen science project management with PlutoF workbench.

From a scientific publication to data in EU BON: GoldenGate Imagine conversion and TreatmentBank as access and dissemination tool of published research data (Donat Agosti and Guido Sautter)

Taxonomic, floristic and faunistic publications and expedition reports include the entire description of all the species of the world's biodiversity. They can be very rich in data and are often the only record known of a given species. There are millions of such taxonomic treatments in the printed record, but unlike the rapidly growing number of online accessible observation and DNA records, this data is still sleeping a Sleeping Beauty's sleep. Treatments are the explicit part of taxonomic name usages, and are the first hand link between a name and the underlying research data. In cyberspace, this constellation allows to provide a link from a taxonomic name to the treatment and from there to a wealth of linked data, from the host article to specimens to multimedia. Finally, access to treatments and data therein allows a novel level of data analyses and visualization.

Plazi's TreatmentBank and GoldenGate provide the tools to convert biodiversity literature into semantically enhanced documents, its storing, worldwide dissemination and analyses. It also allows direct import of articles born with Taxpub Journal Article Tag Suit based biodiversity domain specific mark-up, such as the journals published by Pensoft. The minting of persistent identifiers allows creating a link from a taxonomic name to its treatment.

The trainings session will provide an introduction into the concept of semantic mark-up and treatments. The workflow will be demonstrated from the conversion of an article to the upload to TreatmentBank is visualization and import to the Global Biodiversity Information Facility, where the data is hosted for usage in EU BON. Participants will have to option to convert their own articles. Software can be downloaded from http://plazi.org/resources/treatmentbank/.

ANNEX 3: LIST OF TRAINERS AND PARTICIPANTS

LIST OF TRAINERS

Name	**Organization**	**Country**
Allan Zirk	PlutoF	Estonia
Bruce E. Wilson	DataONE	United States
Donat Agosti	Plazi	Sweden
Eamonn O Tuama	GBIF	Denmark
Francisco Antonio García Camacho	CSIC	Spain
Franck Theeten	MRAC	Belgium
Guido Sautter	Plazi	Sweden
Hannu Saarenmaa	UEF	Finland
Israel Peer	Gluecad	Israel
Larissa Smirnova	MRAC	Belgium
Lorenzo Bigagli	GEOSS	Italy
Lyubomir Penev	Pensoft Publishers	Bulgaria
Pavel Stoev	Pensoft Publishers	Bulgaria
Teodor Georgiev	Pensoft Publishers	Bulgaria
Tim Robertson	GBIF	Denmark
Veljo Runnel	PlutoF	Estonia
Viktor Senderov	Pensoft Publishers	Bulgaria

LIST OF PARTICIPANTS

(*EU BON partner)

	Name	**Organization**	**Country**
1	Aaike De Wever*	Royal Belgian Institute of Natural Sciences	Belgium
2	Aboudou Radji	GBIF Node	Togo
3	Agosti Donat*	Plazi	Switzerland
4	Alan Zirk	University of Tartu Natural History Museum	Estonia
5	Alberto González-Talaván	GBIF	Denmark
6	Alex Asase	GBIF Node	Ghana
7	Ana Casino	CETAF	Belgium
8	Anabela Plos	GBIF Node	Argentina
9	Anders Telenius	GBIF node	Sweden
10	André Heughebaert	GBIF Node	Belgium
11	Andreas Allspach	Senckenberg Research Institute Frankfurt (SGN)	Germany
12	Andry Rakotomanjaka	GBIF Node	Madagascar
13	Anke Hoffmann*	Museum für Naturkunde	Germany
14	Anne-Sophie Archambeau	GBIF Node	France
15	Anton Güntsch*	Freie Universität Berlin	Germany
16	Antoniya Hubancheva	National Museum of Natural History - Sofia	Bulgaria
17	Aurelie Delavaud	French Foundation for Research on Biodiversity	France
18	Björn Quast	Zoologisches Forschungsmuseum Koenig, Bonn	Germany
19	Bruce Wilson	DataONE	United States
20	Carlos Rodriguez*	CSIC (EU BON)	Spain
21	Cedric CHAVERIAT	French Fundation for Biodiversity Research	France
22	Cees Hof	GBIF Node	Netherlands
23	Christian Elloran	ACB (Asean Centre for Biodiversity)	Philippines
24	Christian Langer	GEO BON	Germany
25	Christian Svindseth	GBIF Node	Norway
26	Corinne Martin *	UNEP World Conservation Monitoring Centre	United Kingdom
27	Dag Endresen	GBIF Node	Norway

28	Dairo Escobar	GBIF Node	Colombia
29	Daniel García	GBIF Node	Colombia
30	Danny Vélez	GBIF Node	Brazil
31	Dimitri Brosens	GBIF Node	Belgium
32	Éamonn Ó Tuama*	GBIF	Denmark
33	Elspeth Haston	Royal Botanic Garden Edinburgh	UK
34	Emmanouella Panteri	HCMR	Greece
35	Eric Chenin	GBIF Node	France
36	Eva Chatzinikolaou	Hellenic Centre for Marine Research	Greece
37	Fabricio Baccaro	INPA	Brazil
38	Fatima Parker-Allie	GBIF Node	South Africa
39	Faustin Gashakamba	ARCOS	Rwanda
40	Florian Wetzel*	Museum für Naturkunde	Germany
41	Florian Wolf*	iDIV	Germany
42	Francis Oguya	GBIF Node	Kenya
43	Francisco Antonio García Camacho*	CSIC	Spain
44	Franck Theeten*	RMCA	Belgium
45	Galia Georgieva	Institute of Biodiversity and Ecosystem Research at the Bulgarian Academy of Sciences	Bulgaria
46	Gautam Talukdar	GBIF Node	India
47	Giorgio Brunialti	TerraData environmetrics	Italy
48	Gregor Hagedorn	Museum für Naturkunde	Germany
49	Guido Sautter	Plazi	Sweden
50	Hanna Koivula	Finnish Museum of Natural History - Luomus	Finland
51	Hannu Saarenmaa*	UEF, Digitarium	Finland
52	Henry Engledow	Botanic Garden Meise	Belgium
53	Hulda Gideon	GBIF Node	Tanzania
54	Hyung Park	GBIF Node	Rep. Korea
55	Ilia Gjonov	National Museum of Natural History, Sofia (NMNHS)	Bulgaria
56	Innocent Akampurira	GBIF Node	Uganda
57	Israel Peer*	GlueCAD	Israel
58	Jaona Ranaivo	GBIF Node	Madagascar

59	Jean Ganglo	GBIF Node	Benin
60	Jiri Frank	National Museum	Czech Republic
61	Kari Lahti	Finnish Natural History Museum	Finland
62	Katerina Vasileiadou	HCMR	Greece
63	Larissa Smirnova*	RMCA	Belgium
64	Laura Russel	GBIF Node, VertNet	United States
65	Lawrence Monda	GBIF Node	Kenya
66	Liam Lysaght	GBIF Node	Ireland
67	Lorenzo Bigagli	CNR-IIA	Italy
68	Lyson Kampira	GBIF Node	Malawi
69	Lyubomir Penev *	Pensoft	Bulgaria
70	Manuel Vargas	GBIF Node	Costa Rica
71	Marco Calderisi*	TerraData environmetrics	Italy
72	Marian van der Meij	Naturalis	Netherlands
73	Marie-Elise Lecoq	GBIF France - MNHN	France
74	Martin Kalfatovic	BHL	United States
75	Matúš Kempa*	Institute of Botany, Slovak Academy of Sciences	Slovakia
76	Melianie Raymond	GBIF Node	Denmark
77	Miguel Fernandez*	GEO BON	Germany
78	Moulaye Ainina	GBIF Node	Mauritania
79	Nicolas Noé	GBIF Node	Belgium
80	Nils Valland*	Norwegian Biodiversity Information Centre	Norway
81	Ofer Steinitz	GBIF Node	Israel
82	Patricia Mergen *	RMCA	Belgium
83	Pierrette Ramasiarisoa	GBIF Node	Madagascar
84	Piotr Tykarski	GBIF Node	Poland
85	Radoslav Stanchev	Executive Environmental agency	Bulgaria
86	Razfimpahanana Andriamandimbisoa	GBIF Node	Madagascar
87	Rostislav Bekchiev	National Museum of Natural History - Sofia	Bulgaria
88	Rui Figueira	GBIF Node	Portugal
89	Rumen Tomov	University of forestry-Sofia	Bulgaria

90	Saïdou Doumbouya	GBIF Node	Guinea
91	Sarah Faulwetter	HCMR	Greece
92	Segre Hila	Hamaarag	Israel
93	Sophie Pamerlon	GBIF France - MNHN	France
94	Tahiana Andriaharimalala	GBIF Node	Madagascar
95	Takeshi Osawa	GBIF Node	Japan
96	Teodor Georgiev*	Pensoft Publishers	Bulgaria
97	Thanos Dailianis	Hellenic Centre for Marine Research - IMBBC	Greece
98	Tihomir Stefanov	National Museum of Natural History	Bulgaria
99	Tim Robertson*	GBIF	Denmark
100	Ursula Smith	GBIF Node	Australia
101	Veljo Runnel*	University of Tartu Natural History Museum	Estonia
102	Wouter Addink	Species2000	Netherlands
103	Wouter Koch	Norwegian Biodiversity Information Centre	Norway
104	Yu-Huang Wang	GBIF Node	Chinese Taipei

ANNEX 4: EVALUATION SURVEY STRUCTURE

General questions about the overall training:

	Strongly agree	agree	neutral	disagree	Strongly disagree	Total
The objectives of the training were clearly defined.						
The training objectives were met.						
This training experience will be useful in my work.						
How do you rate the overall training						

Evaluation of separate courses:

	Strongly agree	agree	neutral	disagree	Strongly disagree	Total
The topic was relevant to EU BON						
The content was organized and easy to follow						
The trainer was knowledgeable about the topic.						
The materials distributed were helpful						

General comments: What was particularly helpful about the training? What other subjects for training are you interested in? Would you be willing to teach a course at an EU BON training event? If yes, what topic? Other comments, observations, suggestions.