Milestone MS283

Version: 1.0
Date: 2015-11-26
Author: L. Smirnova, P. Mergen
Document reference: Milestone_MS283

Training workshop on data standards and sharing tools for external users worldwide
(M33)

STATUS: FINAL

EU BON

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

Large scale collaborative project

MS283
Training workshop on data standards and sharing tools for external users worldwide

Milestone number	MS283
Milestone name	Training workshop on standards and prototype data sharing tools for the consortium + external users
WP no.	WP2
Lead Beneficiary (full name and Acronym)	MRAC
Nature	Workshop held
Delivery date from Annex I (proj. month)	2015-08-31 (M33)
Delivered	yes
Actual forecast delivery date	2015-11-26
Comments	

Name of the Authors	Name of the Partner	Logo of the Partner
Smirnova Larissa	MRAC	
Mergen Patricia	MRAC	

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 MS283 summarises the results of the **Third EU BON Training Event** and its evaluation based on the participants' feedback. This training was in fact part of the training for Global Biodiversity Information Facility (GBIF) Node Managers[1] organized by the GBIF and held in Madagascar, Antananarivo on 4-5 of October 2015.

The EU BON representative (Larissa Smirnova, MRAC) was invited to facilitate the sample data publishing session and MRAC has been also entrusted to write a manual summarizing the best practice for publishing sample data (lead by Patricia Mergen).

The training event addressed a core topic for GBIF: the **publishing of biodiversity data** online. Diverse data types were covered and **sample-based data** were introduced. The accent was made on the hands-on sessions using the new version of the GBIF Integrated Publishing Toolkit (IPT v.2.3) which was adapted in collaboration with EU BON to enable the flow of sample-based data in support of Essential Biodiversity Variables (EBVs).

The training was attended by 43 participants from all around the world (34 countries, **Annex 1**).

Introduction

This training event was clearly in line with the objectives set by the EU BON project:

- *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.*
- *Ensure **global coordination of development efforts** through an international data interoperability task force and adoption of the results through helpdesk and a comprehensive training program.*

---

[1] http://gb22.gbif.org/GB22

This activity is also aligned 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".*

GBIF in collaboration with the EU BON project has adapted its Integrated Publishing Toolkit (IPT) with new Event core and associated extensions in order to enable the flow of sample data from numerous ecological, environmental and natural resource monitoring programmes. Such data are usually quantitative, calibrated, and follow certain protocols so that changes and trends of populations can be detected[3] and thus can be used in modelling and analysis to detect changes and trends in populations, aim at supporting the EBVs.

First introduced during the 2nd EU BON Training these new IPT functionalities were tested by EU BON and GBIF consortium, particularly using data from the EU BON test sites. The testing was done to solve the issues before the updated version of IPT was released and the published datasets served as use-cases for the hands-on training sessions.

The two-day training[4] covered different aspects of data publishing (see **Annex 2**):
- review of the GBIF publishing landscape,
- the various ways in which it is currently possible to publish biodiversity information and the existing mechanisms to publish complex datasets,
- recent adaptation of the Darwin Core standard to accommodate data coming from sampling efforts,
- application of sample-based data and their importance for modern biology.

Achievements and current status

Already during pre-course activities several important conclusions were stated by participants (mainly GBIF node managers):
- There is a demand from content providers.
- Some datasets already published in GBIF would better fit sample-based data concept.
- New IPT functionalities will allow publishing of more diverse datasets.
- New type of data will extend range of data providers.
- Technical documentation and support are important prerequisites for successful implementation of this new development.

During the two days hands-on training participants received (or enhanced) knowledge on data publishing using the EU BON IPT v.2.3 and were introduced to the concept of sample data, its definition, what is the difference with opportunistic occurrence data and what is an application area of these data. Students followed step by step the IPT demonstration (for occurrence and sample-based data) and could try out to publish data themselves during two group work sessions. They were encouraged to use a new Event core and extensions (for example Measurement-or-Facts, Relevé) to publish additional environmental and vegetation data. Attention was also paid to the importance of metadata for such type of data, laying the link to data papers. During practical exercises students were

---

[2] www.cbd.int/gbo4
[3] Sample-data publishing primer
[4] http://community.gbif.org/pg/pages/view/47903/agenda-for-the-gb22-training-event

confronted with real but simplified use cases based on sample datasets published through the test EU BON IPT.

For the first group exercise classical types of biodiversity datasets were used: natural history collection data, observational data, occurrence data from literature and taxonomic checklists.

The group exercises of the second day were articulated around example datasets based on different methodologies for samples and inventories: environmental impact assessment in aquatic environments, surveys of *Lepidoptera* by citizen scientists, vegetation studies using the relevé method and effects of climate change in coral reefs.

The program and presentations are available online[5].

After two days of training the participants evaluated the course activities and trainers. Based on 37 answers the two sessions on sample-based data were evaluated very positive with an average of 4.3 (from 5), see **figure 1**.
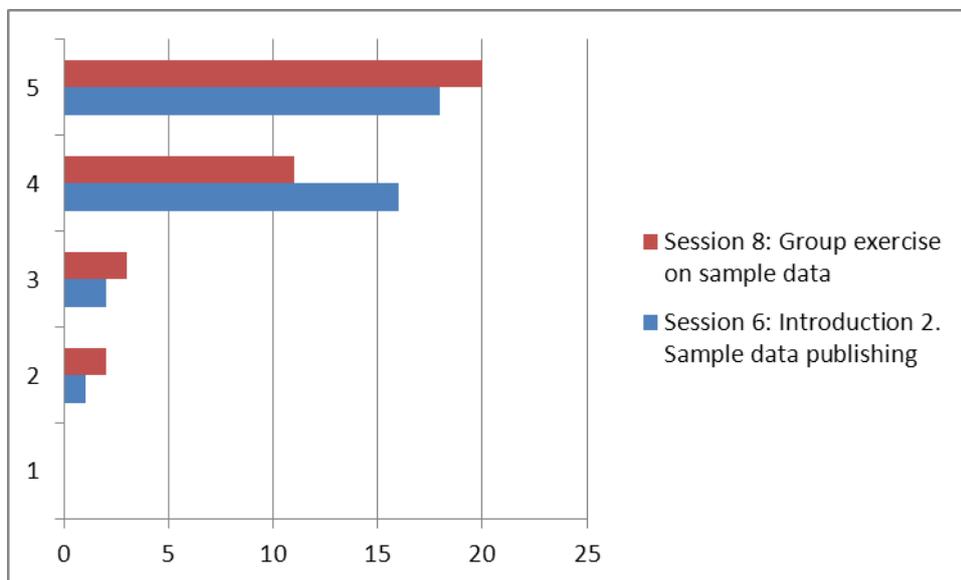
Figure 1: Evaluation of the training sessions

In comments and suggestions it was mentioned that such training have great potential for knowledge increase in general and that almost all participants have learned new things on data publishing and in particular on sample-based data publishing. Many participants showed interest in data paper publishing and data cleaning tools. Also aspects of data preparation before the publishing would be of great interest and introduction to good data management practices would have been very valuable.

Challenges and further/future developments

The new IPT functionalities enabling publish sample data have got a very positive reaction from the participants, which were mainly GBIF node managers working on daily basis with data and knowing needs and concerns of data providers. They were satisfied that with these new functionalities they can reach a wider range of data providers and publish more complex data than just occurrences. Many have clearly shown intention to publish or re-publish datasets after the training workshop, especially the vegetation plots were mentioned as a favourite type of data to be suitable for publishing with Event core.

---

[5] http://community.gbif.org/pg/file/read/47920/gb22-nodes-events-brochure

During the training and the discussion afterwards, 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 for mapping of data to the DwC, but more to help participants to understand the star scheme principle of DwC, the use of Event core and the choice of appropriate extensions. Over time and with increasing amount of published sample data we will be able to evaluate it and give better recommendations in form of best practices for different types of data.

So for the next training it could be interesting to pay more attention to DwC mapping, taking as an example the most popular monitoring protocols such as vegetation plot, marine sampling, fresh water survey, bird survey etc.

The IPT is a power tool for data sharing, so in general promoting IPT, giving the people the possibility to test it with hands-on training sessions, providing them with sample data sets for pre- or post-training exercises would be a good approach.

Also providing them with multi-lingual support (manuals, interface, help functions) would be of great importance. The IPT training could be combined with the introduction and practice on data paper publishing using Pensoft Publishing Toolkit as it will give the participants more clear and deeper view on the importance of the metadata and for the monitoring sites or programs who have quite well-designed protocols and descriptions it would make use of the IPT more attractive. Also interactions and interoperability's between different metadata standards and repositories could be addressed.

Annex 1: list of participants.

	Name	Surname	Country
1	Anders	Telenius	Sweden
2	Andry	Rakotomanjaka	Madagascar
3	Dairo	Escobar	Colombia
4	Innocent	Akampurira	Uganda
5	Jaona	Ranaivo	Madagascar
6	Lawrence	Monda	Kenya
7	Piotr	Tykarski	Poland
8	Takeshi	Osawa	Japan
9	Aboudou	Radji	Togo
10	Eric	Chenin	France
11	Faustin	Gashakamba	ARCOS
12	Moulaye	Ainina	Mauritania
13	Pierrette	Ramasiarisoa	Madagascar
14	Razfimpahanana	Andriamandimbisoa	Madagascar
15	Saïdou	Doumbouya	Guinea
16	Sophie	Pamerlon	France
17	Tahiana	Andriaharimalala	Madagascar
18	Alex	Asase	Ghana
19	André	Heughebaert	Belgium
20	Dag	Endresen	Norway
21	Daniel	García	Colombia
22	Francis	Oguya	Kenya
23	Lyson	Kampira	Malawi
24	Martin	Kalfatovic	BHL
25	Ursula	Smith	Australia
26	Wouter	Addink	Species2000
27	Anne-Sophie	Archambeau	France
28	Dimitri	Brosens	Belgium
29	Hanna	Koivula	Finland
30	Hyung	Park	Rep. Korea
31	Jean	Ganglo	Benin
32	Liam	Lysaght	Ireland
33	Ofer	Steinitz	Israel
34	Rui	Figueira	Portugal
35	Anabela	Plos	Argentina
36	Cees	Hof	Netherlands
37	Christian	Elloran	ACB
38	Christian	Svindseth	Norway
39	Fatima	Parker-Allie	South Africa
40	Gautam	Talukdar	India
41	Hulda	Gideon	Tanzania

| 42 | Manuel | Vargas | Costa Rica |
| 43 | Yu-Huang | Wang | Chinese Taipei |

Annex 2: training program

1. Sunday 4 October

General overview of the 2015 data publishing landscape

08:30   *Registration*

09:00   **Session 01: Welcome, practicalities, introductions**
Led by Alberto González-Talaván
Welcome by organizers. Review of the programme. Practical information. Participants' introduction. Introduction to certification and assessment during the course. Review of pre-course activities.

10:30   *Coffee/tea break - Madagascar DLC terrace*

11:00   **Session 02: The GBIF data publishing landscape in 2015**
Led by Laura Russell
First introductory session where we will review basic concepts that we will use in the rest of the sessions.

12:30   *Lunch break - Madagascar DLC terrace*

13:30   **Session 03: Practical demo on how to prepare and map different types of data**
Led by Nicolas Noé
Step-by-step demo where we will have the opportunity to see in detail (and practice!) how to publish a dataset using IPT and DwC extensions.

14:30   *Coffee/tea break - Madagascar DLC terrace*

15:00   **Session 04: Data publishing practice in groups**
Led by 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.

17:00   *End of the day*

2. Monday 5 October

Sample-based data publishing and social aspects of data publishing

08:30   *Registration*

09:00   **Session 05: Promoting data publishing: a practical exercise in groups**
Led by Mélianie Raymond.
Practical session to develop our skills in data publishing promotion by generating custom promotional materials to be used in a specific scenario.

10:30   *Coffee/tea break - Madagascar DLC terrace*

11:00   **Session 06: Introduction to sample-based data publishing**
Led by Larissa Smirnova
During this session we will work on the particularities of the sample-based data, and the challenges and opportunities that this data type brings to GBIF.

12:30   *Lunch break - Madagascar DLC terrace*

13:30   **Session 07: Sample-based data publishing demo**
Led by Danny Vélez
We will go together through a step-by-step example of publishing of sample-based data, to

make sure that all participants have seen the complete process in detail at least once. We will highlight the differences with other data types.

14:30   *Coffee/tea break - Madagascar DLC terrace*

15:00   **Session 08: Sample-based data publishing practice in groups**
Led by 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.

16:30   **Session 09: Course summary and evaluation**
Led by Alberto González-Talaván
Review of the contents of the course and all follow-ups agreed. An evaluation survey will be circulated among participants.