

Workshop Report

Harmonizing plot data with collection data

Mareike Petersen[‡], Falko Glöckler[‡], Jana Hoffmann[‡][‡] Museum für Naturkunde Berlin, Leibniz Institute for Evolution and Biodiversity Science, Berlin, GermanyCorresponding author: Mareike Petersen (mareike.petersen@mfn.berlin)

Reviewable

v1

Received: 30 Jan 2019 | Published: 31 Jan 2019

Citation: Petersen M, Glöckler F, Hoffmann J (2019) Harmonizing plot data with collection data. Research Ideas and Outcomes 5: e33509. <https://doi.org/10.3897/rio.5.e33509>

Abstract

Although plot or monitoring data are quite often associated with objects collected in the plot and stored in specific collections, controlled vocabularies currently available do not cover both disciplines. This situation limits the possibility to publish common data sets and consequently brings a loss of significant information by combining plot-based research with collection object associated data. To facilitate the exchange and publication of these important data sets, experts in natural history collection data, ecological research, and environmental science met for a one-day workshop in Berlin. The participants discussed data standards and ontologies relevant for each discipline and collected requirements for a first application schema covering terms important for both, collection object related data and plot-based research.

Keywords

application schema, biodiversity, data standard, ecology, monitoring, natural history collection

Date and place

30 May 2018, Museum für Naturkunde Berlin

List of participants

- Brian Baltruschat, Museum für Naturkunde Berlin
- Frederik Berger, Museum für Naturkunde Berlin
- Gabi Dröge, Botanischer Garten und Botanisches Museum Berlin
- Jonas Geschke, Museum für Naturkunde Berlin
- Maren Gleisberg, Botanischer Garten und Botanisches Museum Berlin
- Falko Glöckler, Museum für Naturkunde Berlin
- Sebastian Kirchof, Museum für Naturkunde Berlin
- Rudolf May, Bundesamt für Naturschutz
- Anke Penzlin, Senckenberg Gesellschaft für Naturforschung
- Mareike Petersen, Museum für Naturkunde Berlin
- Fabian Reimeier, Botanischer Garten und Botanisches Museum Berlin
- Martin Stricker, Humboldt Universität Berlin

Introduction

Plot sampling is a widely used method in ecology and biodiversity research. These inventories are commonly accompanied with the collection of voucher specimen for e.g. further identification or analysis. Whereas gathered information of plot observations are usually assigned to a plot or a plot observation to a particular time, specimen in natural history collections are managed on unit level with all information and measurements attached to a specimen. Currently there is a variety of controlled vocabularies (thesaurus) available for biological and related disciplines. A set of defined descriptive terms are either arranged in structured data standards or put in relation to each other in an ontology. However, controlled vocabularies designed for a standardized exchange of collection data, e.g. *Access to Biological Collection Data (ABCD, Berendsohn 2007)*, might not be efficiently used for plot data. On the other hand ecological vocabularies, e.g. *Ecological Metadata Language (EML, ecoinformatics.org 2011)* might not sufficiently cover information about the storage and preparation of physical collection objects. A combination of both disciplines might be rudimentarily expressed using the *Event Core* (https://terms.tdwg.org/wiki/Darwin_Core_Event) of the taxa based standard *Darwin Core* (Darwin Core Task Group 2015) or particular *ABCD* terms (*MeasurementOrFact*, <https://terms.tdwg.org/wiki/abcd2:Biotope-MeasurementOrFactAtomised>), but whether these options allow for an adequate representation of the plot observation event has not been evaluated. Controlled vocabularies for the natural history as well as ecological and environmental science have been developed in separated domains, despite their huge amount of overlapping terms. Both disciplines need to describe properties such as locality, (taxonomic) determinations, species traits, project metadata, involved persons and their affiliations etc. In practice there are examples in which the disconnection of the vocabularies might cause issues regarding data management and interoperability, and in which none of the data standards can be applied adequately. Hence, researchers need a comprehensive and flexible standard schema that also defines the relations between the standards of both domains.

Within the scope of the research and service project “ABCD 3.0 – A community platform for the development and documentation of the ABCD standard for natural history collections” *1 (<https://abcd.biowikifarm.net/>) all ABCD terms were imported into the TDWG Terms Wiki (https://terms.tdwg.org/wiki/ABCD_2), a developmental platform which allows collaborative work of the terminology and the schema itself. Here, also relationships including their specifications (“is part of”, “perfect match”) with other vocabularies or direct translations of terms can be added. This enables the application of ABCD, a direct review, and further development by scientists of different disciplines. Currently the XML-based structure of ABCD is being changed into a semantic form embedding missing terms derived from external, already existing ontologies. In close cooperation with the scientific community, application schemata for particular use cases are formed (compare Petersen et al. 2019). In addition to (technical) mandatory elements and elements of general importance, application schemata comprise parts of the ABCD schema relevant for specific purposes; i.e. discipline, collection, or for the publication in a particular data portal. Thus it is a defined subset of concepts available in the whole ABCD schema and if necessary supplemented with concepts from other standards.

The workshop was carried out in the framework of the ABCD 3.0 project, a collaboration of the Museum für Naturkunde Berlin and the Botanical Garden and Botanical Museum Berlin Dahlem. During the workshop experts of natural history collections, biodiversity standards and ecological/environmental science met in order to share their experience and collect their requirements for the publication of plot-based data. The participants examined different domain-specific vocabularies and discussed terms necessary to describe plot-based research data including e.g. habitat characterization, time series, monitoring, and the collection of sample specimen. The workshop’s results are presented in this report. Furthermore necessary tasks towards an application schema for plot data were discussed and are documented here.

Aims of the workshop

The workshop intended to evaluate whether the standard ABCD fulfills all demands and whether other domain-specific controlled vocabularies contain appropriate, supplementary terms for the publication of plot-based data. The aim for the one-day meeting was a first version of an application schema linking collection objects with plot-based research.

Workshop program

The workshop program included a short informative and an extensive interactive part (see Suppl. material 1). After a general introduction, all participants were asked to introduce themselves and describe their experience with controlled vocabularies. The organizers gave an overview on existing vocabularies, including different data standards and ontologies, associated with plot-based research and collection data prior to the working session. In small groups the participants delved deeper into *Extensible Observation*

Ontology (*OBOE*, Madin et al. 2007), *Observation and Measurements* (Cox 2013), *Humboldt Core* (Guralnick et al. 2018), *Veg-X* (exchange standard for vegetation-plot data, Wisner et al. 2011), and *Ecological Metadata Language* (*EML*, ecoinformatics.org 2011, Fegraus et al. 2005). The participants were asked to check the standards and ontologies for adequate terms in order to model plot-based data and, if possible, to map information associated with objects gathered on the plot and subsequently stored in a scientific collection. The respective findings were presented and discussed with all participants. In the last section, the workshop focused on the first steps towards an application schema for plot-based data and considered essential information, their relation to each other, and whether they are used once or multiple times (cardinality).

Key outcomes and discussions

It was shown, that plot-like data, e.g. DNA samples, can be expressed with *ABCD* and its extension *GGBN* (Droege et al. 2016). The concepts to describe the accruing information are available, but the hierarchical XML structure is limiting the mapping possibilities. This difficulty became even more clear when considering the different record types (lots vs. single specimen) in research projects investigating biodiversity (e.g. <http://www.indobiosys.org/>). The workshop participants collected further use cases common in plot-based research and which therefore should be considered preparing a discipline specific application schema, such as environmental sample (incl. chemical properties), plot properties and vegetation characterization (incl. vouchers), habitat / biotope mapping, or monitoring of plots (time series). Keeping these use cases in mind, the participants analyzed existing controlled vocabularies with respect to adequate terms covering the required information.

- The ontologies *OBOE* and *Observation and Measurement* are more generic and allow the representation of manifold data types derived from plot-based research. Although both ontologies should be taken into consideration when establishing an ontology on plot-based data, some terms might need a more precise definition for the particular use cases discussed during the workshop.
- *Humboldt Core* represents a list of terms for ecological inventories but is not yet a ratified standard. Data related to the sampling event itself (locality, time), the procedure, and the general scope of an inventory can be perfectly expressed with terms described therein. In case any collection object centered standard (e.g. *ABCD*) needs to be extended for plot-based research, one should make use of and refer to the well-defined *Humboldt Core*.
- *Veg-X* is an XML based standard mainly produced for vegetation-plot data. It is structured into several data components such as fixed information about plot (e.g. altitude, slope), plot observation, and observed organism, etc. The most innovative part of the *Veg-X* standard and potentially re-usable for the discussed use cases in our workshop is the plot observation. Other parts of *Veg-X* are already adopted from other standards including *EML* (for protocols and projects definition), *Darwin Core* (for geo-data), and *Taxon Concept Schema* (taxon names).

- *EML* describes the essential aspects of ecological data covering e.g. the general dataset, geographic and temporal aspects, and methods. For the purposes discussed during the workshop, *EML* terms could be valuable for the general project description and for the specification of methods and should definitely be taken into consideration for plot-data application schemata.

Following this, participants were asked to compile thematic use cases for plot-based research. Eventually, it was agreed to collaboratively work on a single, more general use case: the mapping of a habitat or biotope including multiple visits (time series). In a closing discussion session all relevant concepts were collected, their relationships to each other were considered, and the cardinality of each concept was reasoned. Fig. 1 and Table 1 give details on the terms incorporated in a first version of an application schema for plot-based data.

Table 1.

Concepts of a first version of an application schema for plot-based data in combination with collection objects. Given are concepts discussed during the workshop and which should be incorporated in an application schema (compare Fig. 1). In addition to a short description we assigned the cardinality to each concept. Various concepts are of importance for the plot research as well as for the description of the collection object (e.g. measurements, associated multimediaobjects, people conducted research etc.).

Concept	Description	Cardinality
plot research		
project metadata	details on the framework of the plot research (project, institution, scope, etc.)	n
spatial concept	describing and related to the location of the plot	1
temporal concept	describing an observation or measurement at the plot in time	n
measurement	any measurement conducted during plot observations (vegetation, soil, temperature, etc.)	n
person	people conducting the plot research	n
multimedia object	any multimedia objects associated with the plot / a plot observation	n
publication	any publication associated with the plot / a plot observation	n
collection object		
specimen	specimen observed / gathered during a plot visit	n
taxonomy	determination / taxonomic identity of the specimen	1
measurement	individual measurement of the specimen	n
multimedia object	any multimedia objects associated with the plot / a plot observation	n
publication	any publication associated with the plot / a plot observation	n
identifier	persistent identifier for the collected specimen	1
storing collection	collection holding the specimen	1

centralized form. This will facilitate the link to other appropriate vocabularies and to draw application schema for different plot-based research questions. On the other hand, the maintainers of other data standards should prefer to re-use appropriate *ABCD* terms over creating new terms as soon as they are going to extend their schema towards collection objects. Thus the development of a real application schema for plot-based data should be done in collaboration. The list of terms collected in this workshop will however serve as a guideline for the publication of plot data in the meantime and later on in consensus with experience in plot-like examples (e.g. sample with environmental DNA) and knowledge using *ABCD* over the last years (Holetschek 2015, Holetschek 2016, Petersen et al. 2018) be incorporated in the development of application schemata linking collection objects to plot research.

Acknowledgements

The workshop was supported by ABCD 3.0, a DFG project funded under the LIS infrastructure platform. We thank G. Dröge for her inspiring talk and all participants for their valuable contribution and constructive comments on an earlier version of the report. The publication of this article was funded by the Open Access Fund of the Leibniz Association.

References

- Berendsohn W (2007) Access to biological collection data. ABCD Schema 2.06 – Ratified TDWG Standard. <http://www.bgbm.org/TDWG/CODATA/Schema/default.htm>
- Cox S (2013) An explicit OWL representation of ISO/OGC Observations and Measurements. http://ontology.cim3.net/file/work/OntologyBasedStandards/2013-10-17_Ontologies-for-Geospatial-Standards/Observations-n-Measurements-OWL--SimonCox_20131017.pdf. Accessed on: 2018-8-15.
- Darwin Core Task Group (2015) Darwin Core Terms: a quick reference guide. <http://rs.tdwg.org/dwc/terms/>. Accessed on: 2018-9-19.
- Droege G, Barker K, Seberg O, Coddington J, Benson E, Berendsohn WG, Bunk B, Butler C, Cawsey EM, Deck J, Döring M, Flemons P, Gemeinholzer B, Güntsch A, Hollowell T, Kelbert P, Kostadinov I, Kottmann R, Lawlor RT, Lyal C, Mackenzie-Dodds J, Meyer C, Mulcahy D, Nussbeck SY, O'Tuama É, Orrell T, Petersen G, Robertson T, Söhngen C, Whitacre J, Wieczorek J, Yilmaz P, Zetsche H, Zhang Y, Zhou X (2016) The Global Genome Biodiversity Network (GGBN) data standard specification. Database 2016: baw125. <https://doi.org/10.1093/database/baw125>
- ecoinformatics.org (2011) Ecological Metadata Language. <https://kn.b.ecoinformatics.org/#tools/eml>. Accessed on: 2018-8-16.
- Fegraus EH, Andelman S, Jones MB, Schildhauer M (2005) Maximizing the value of ecological data with structured metadata: An introduction to Ecological Metadata Language (EML) and principles for metadata creation. Bulletin of the Ecological Society of America 86 (3): 158-168. [https://doi.org/10.1890/0012-9623\(2005\)86\[158:MTVOED\]2.0.CO;2](https://doi.org/10.1890/0012-9623(2005)86[158:MTVOED]2.0.CO;2)

- Guralnick R, Walls R, Walter Jetz (2018) Humboldt Core—toward a standardized capture of biological inventories for biodiversity monitoring, modeling and assessment. *Ecography* 41 (5): 713-725. <https://doi.org/10.1111/ecog.02942>
- Holetschek J (2015) BioCASE Concept survey, Biological CollectionAccess Service for Europe. www.biocase.org/whats_biocase/concept_survey.cgi. Accessed on: 2018-8-21.
- Holetschek J (2016) Commonly used ABCD 2.06 concepts, Documentation on Wiki of the BioCASE Provider Software. <http://wiki.bgbm.org/bps/index.php/CommonABCD2Concepts>. Accessed on: 2018-8-21.
- Madin J, Bowers S, Schildhauer M, Krivov S, Pennington D, Villa F (2007) An ontology for describing and synthesizing ecological observation data. *Ecological Informatics* 2 (3): 279-296. <https://doi.org/10.1016/j.ecoinf.2007.05.004>
- Petersen M, Glöckler F, Kiessling W, Döring M, Fichtmüller D, Laphakorn L, Baltruschat B, Hoffmann J (2018) History and development of ABCDEFG: a data standard for geosciences. *Fossil Record* 21 (1): 47-53. <https://doi.org/10.5194/fr-21-47-2018>
- Petersen M, Hoffmann J, Glöckler F (2019) Access to Geosciences – Ways and Means to share and publish collection data. *Research Ideas and Outcomes* 5: e32987. <https://doi.org/10.3897/rio.5.e32987>
- Wiser SK, Spencer N, De Caceres M, Kleikamp M, Boyle B, Peet RK (2011) Veg-X—an exchange standard for plot-based vegetation data. *Journal of Vegetation Science* 22 (4): 598-609. <https://doi.org/10.1111/j.1654-1103.2010.01245.x>

Supplementary material

Suppl. material 1: Workshop_Program_Plot_Dat [doi](#)

Authors: M. Petersen et. al.

Data type: Workshop Program

Filename: Workshop Program.pdf - [Download file](#) (195.03 kb)

Endnotes

- *1 ABCD 3.0: Funded by the German Research Foundation (Deutsche Forschungsgemeinschaft), Scientific Library Services and Information Systems; partners: Museum für Naturkunde Berlin (MfN) and Botanical Garden and Botanical Museum Berlin Dahlem (BGBM).