

DiSSCo related output

This template collects the required metadata to reference the official Deliverables and Milestones of DiSSCo-related projects. More information on the mandatory and conditionally mandatory fields can be found in the supporting document 'Metadata for DiSSCo Knowledge base' that is shared among work package leads, and in Teamwork > Files. A short explanatory text is given for all metadata fields, thus allowing easy entry of the required information. If there are any questions, please contact us at info@dissco.eu.

Title

Milestone Report MS 5.7 "Compilation of data standards forming the basis for the initial version of the DiSSCo Digital Specimen Object Specification"

Author(s)

David Fichtmueller

Identifier of the author(s)

<https://orcid.org/0000-0002-0829-5849>

Affiliation

Botanic Garden and Botanical Museum (BGBM)
Berlin, Freie Universität Berlin

Contributors

Kessy Abarenkov (<https://orcid.org/0000-0001-5526-4845>)
Wouter Addink (<https://orcid.org/0000-0002-3090-1761>)
Mathias Dillen (<https://orcid.org/0000-0002-3973-1252>)
Falko Glöckler (<https://orcid.org/0000-0002-7127-2738>)
Jonas Grieb (<https://orcid.org/0000-0002-8876-1722>)
Anton Güntsch (<https://orcid.org/0000-0002-4325-4030>)
Elspeth Haston (<https://orcid.org/0000-0001-9144-2848>)
Sharif Islam (<https://orcid.org/0000-0001-8050-0299>)
Sam Leeflang (<https://orcid.org/0000-0002-5669-2769>)
Sabine von Mering (<https://orcid.org/0000-0003-2982-7792>)
Claus Weiland (<https://orcid.org/0000-0003-0351-6523>)

Publisher

DiSSCo Prepare

Identifier of the publisher

Resource ID

<https://doi.org/10.34960/jemd-0y02>

Publication year

2022

Related identifiers

Is it the first time you submit this outcome?

Yes

Creation date

31/03/2022

Version

Citation

Fichtmüller D. (2022) Milestone Report MS 5.7 "Compilation of data standards forming the basis for the initial version of the DiSSCo Digital Specimen Object Specification". <https://doi.org/10.34960/jemd-0y02>

Abstract

This DiSSCo Prepare Milestone Report MS 5.7 "Compilation of data standards forming the basis for the initial version of the DiSSCo Digital Specimen Object Specification" describes what Open Digital Specimen (openDS) is and how it is situation with the DiSSCo Prepare project. We highlight design decisions and modelling approaches that were done for openDS in general and for the first version in particular.

Primarily this report is highlighting the most important data standards that will be used as a basis for the first version of the openDS data model, either be reusing specific terms or by aligning the data model for better compatibility where direct reuse of terms is either not practical or feasible.

In the end we take a look on how the modelling work of openDS will continue.

Content keywords

scientific

Project reference

DiSSCo Prepare (GA-871043)

WP number

WP5

Project output

Milestone report

Deliverable/milestone number

MS 5.7

Dissemination level

Public

Rights**License**

Attribution 4.0 International (CC BY 4.0)

Resource type

Text

Format

PDF

Funding Programme

H2020-INFRADEV-2019-2

Contact email

d.fichtmueller@bgbm.org



Compilation of data standards forming the basis for the initial version of the DiSSCo Digital Specimen Object Specification

DiSSCo Prepare WP 5 – Milestone 5.7

David Fichtmueller

Contributors: Kessy Abarenkov (Tartu), Wouter Addink (Naturalis), Mathias Dillen (MeiseBG), Falko Glöckler (MfN), Jonas Grieb (SGN), Anton Güntsch (BGBM), Elspeth Haston (RBGE), Sharif Islam (Naturalis), Sam Leeflang (Naturalis), Sabine von Mering (MfN), Claus Weiland (SGN)

WP Lead: Mareike Petersen (MfN)

Task Lead: Anton Güntsch (BGBM)





Abstract

This DiSSCo Prepare Milestone Report MS 5.7 “Compilation of data standards forming the basis for the initial version of the DiSSCo Digital Specimen Object Specification” describes what Open Digital Specimen (openDS) is and how it is situation with the DiSSCo Prepare project. We highlight design decisions and modelling approaches that were done for openDS in general and for the first version in particular.

Primarily this report is highlighting the most important data standards that will be used as a basis for the first version of the openDS data model, either be reusing specific terms or by aligning the data model for better compatibility where direct reuse of terms is either not practical or feasible.

In the end we take a look on how the modelling work of openDS will continue.

Note: This milestone report was originally due on project month 24 (M24), but was postponed to project month 26.

Keywords

openDS, Data Standards, DiSSCo Modelling Framework, DiSSCo Architecture

Index

Abstract	2
Keywords	2
Index	3
01 Introduction	4
Situating the Open Digital Specimen (openDS) within DiSSCo	4
Scope of this Milestone Report	4
Modelling approaches for the initial version of openDS	4
02 Core Data Standards	6
Specimen Focused Standards	6
Generic Data Standards	6
Thematic Standards	6
Approaches not considered for now	7
03 Outlook	7

01 Introduction

Situating the Open Digital Specimen (openDS) within DiSSCo

The Open Digital Specimen (openDS) is an evolving data standard being developed within the DiSSCo Prepare project to describe the digital records of specimen data that are processed and made available through the DiSSCo Infrastructure. The development of openDS is still ongoing but early test versions are available.

The development of openDS is done in digital meetings that take place every second week and span the two tasks of *5.2 DiSSCo Modelling Framework and Data Model* and *6.2 Evaluation of the DiSSCo Architecture*. During the meetings the discussions are usually done on an abstract level using diagram software to display the concepts and their relations. The changes are later fed back into the DiSSCo Modelling Framework¹ from where new schema definitions can be generated to test the new data models in the Digital Specimen Repository².

Scope of this Milestone Report

All the standards discussed in this report were described in the *DiSSCo Prepare Milestone Report for MS 5.5 Compilation of relevant data standards*³. That report includes detailed descriptions of the standards with links, citations and further reading suggestions, that will not be repeated in this report.

For each concept introduced into openDS, an evaluation of existing concepts takes place to determine whether a new concept is required. The aim is to reuse as many existing data elements and their definitions as possible.

This report will focus on standards that can provide significant bases for the modelling of openDS, either by reusing their concepts or aligning openDS concepts with them in terms of definition and structure. There are many other standards from which concepts might be included in openDS, even if only for limited individual concepts. Mentioning all of these standards is beyond the scope of this report. This report will focus on the standards that will be used for the first version of openDS, accepting that additional standards will be included in later versions.

Modelling approaches for the initial version of openDS

The initial version of openDS will be the first step towards a complete open Digital Specimen model and will not yet express all of the information that will potentially be included in later versions. The goal is to have a first version that covers many of the most common use cases for

¹ Fichtmueller D. & Güntsch A. (2022) DiSSCo Prepare Deliverable 5.2 "DiSSCo Modelling Framework". <https://doi.org/10.34960/e3nv-zh69>

² <https://nsidr.org>

³ Petersen, M., von Mering, S. & Glöckler, F. 2021: DiSSCo Prepare Milestone Report MS5.5 "Compilation of relevant data standards". <https://doi.org/10.34960/3mg1-7n14>

sharing specimen data. The standard can then be extended further in the future by subsequently adding more concepts. Later versions should be backward compatible, so that any existing data can be automatically upgraded to the newer version, just by changing the reference to the version that is used.

Even in its production stable later versions, openDS will not be able to express all the details that are included in all of the various data standards coming from all of the various data providers. It will not be an all-encompassing standard. Instead it will focus on the needs for DiSSCo by reusing or aligning to many common standards in the relevant domains such as those of the “TDWG Biodiversity Information Standards” and beyond.

OpenDS will also offer pathways to the original data in its original format, so that interested parties can still retrieve information that is relevant to them, even if it is not part of the openDS specification. This could either be done by providing a reference to the original source where data can be retrieved or by having it included within the data of the digital specimen in its verbatim format, as “payload data”.

The data expressed in openDS have to be treated using the Open World Assumption, in that when a statement is not expressed in the data of a digital specimen, it doesn't mean that said statement is not true, unless it is explicitly stated as such in the form of a negative assertion. For instance, when there are no images associated with the digital object of a specimen, this does not mean that there are no images of the specimen. They may not be exported or converted into openDS. However, if a historic specimen is known to be lost (e.g. due to a fire) and it is known that there are no remaining images that could be digitised, then this information needs to be explicitly stated in the digital specimen for other users to know that there are no images available. Negative assertions, however, will most likely not be part of the initial version of openDS.

The decisions on which concepts to include into openDS (both for the initial and subsequent versions) will be guided by the use cases and user stories gathered in *Task 1.1 Analyse life sciences use cases and user stories*⁴ and *Task 1.2 Analyse Earth sciences use cases and user stories*⁵, which are also available in the DiSSCo GitHub repository⁶.

⁴ Fitzgerald, H., Juslén, A., von Mering, S., Petersen, M., Raes, N., Islam, S., Berger, F., von Bonsdorff, T., Figueira, R., Haston, E., Häffner, E., Livermore, L., Runnel, V., De Smedt, S., Vincent, S., Weiland, C. (2021). DiSSCo Prepare Deliverable D1.1 Report on Life sciences use cases and user stories. <https://doi.org/10.34960/xhwx-cb79>

⁵ von Mering, S., Petersen, M., Fitzgerald, H., Juslén, A., Raes, N., Islam, S., Berger, F., von Bonsdorff, T., Figueira, R., Haston, E., Häffner, E., Livermore, L., Runnel, V., De Smedt, S., Vincent, S., Weiland, C. (2021). DiSSCo Prepare Deliverable D1.2 Report on Earth sciences use cases and user stories. <https://doi.org/10.34960/n3dk-ds60>

⁶ <https://github.com/DiSSCo/user-stories>

02 Core Data Standards

Specimen Focused Standards

openDS will be a standard to describe specimen information, so its core will be closely related to other specimen focused standards, namely **ABCD**⁷ and **Darwin Core (DWC)**⁸ for biological specimens and **ABCD-EFG**⁹ for Earth science specimens. These standards are widely used within the community, though they differ in their structure and extent. Many of the general concepts are comparable or compatible to each other.

Though it is not directly a data standard, **MIDS**¹⁰ (Minimal Information about a Digital Specimen) describes various levels of information completeness that digital specimens can offer. This standard therefore influences what kind of information can be expressed within the concepts of the openDS, so that the openDS records in themselves are compliant with the various MIDS levels.

Generic Data Standards

Many of the concepts required for openDS are independent of biodiversity or geological domains, therefore many of the general concepts can be reused or based on existing and widely known data standards. One of the oldest and most widely used is Dublin Core (DC), which offers a wide range of generic terms. However, many of the concepts were created for the library sector, which is reflected in their definitions, making their reuse more challenging without adjusting their definitions for other contexts. A more recent, generic approach is represented by **schema.org**¹¹ and the **Datacite Metadata Standard**¹². The Datacite Metadata Standard, however, is an XML Schema, which makes the reuse of these concepts in a semantic concept difficult. Aligning certain modelling approaches would still be beneficial.

Thematic Standards

For many of the specific subsections of openDS, corresponding thematic standards can be used as bases or reference models for the modelling process. The primary areas for such thematic

⁷ Access to Biological Collection Data task group. 2007. Access to Biological Collection Data (ABCD), Version 2.06. Biodiversity Information Standards (TDWG) <http://www.tdwg.org/standards/115>; or <https://abcd.tdwg.org>

⁸ Wieczorek J, Bloom D, Guralnick R, Blum S, Döring M, et al. (2012) Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. PLoS ONE 7(1): e29715. <https://doi.org/10.1371/journal.pone.0029715>; or <https://dwc.tdwg.org/>

⁹ 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>; or <https://github.com/tdwg/efg>

¹⁰ Hardisty, A., Addink, W., Dillen, M., Groom, Q., Haston, E., et al. (Draft) Minimum Information about a Digital Specimen (MIDS) v0.14, 29 Mar 2021. <https://github.com/tdwg/mids/>

¹¹ <https://schema.org/>

¹² DataCite Metadata Working Group. (2021). DataCite Metadata Schema Documentation for the Publication and Citation of Research Data and Other Research Outputs. Version 4.4. DataCite e.V. <https://doi.org/10.14454/3w3z-sa82>

standards are currently **Audubon Core**¹³ for multimedia data and the **PROV**¹⁴ for provenance information.

Approaches not considered for now

Although there are already approaches to situate the Ontology for open Digital Specimen (ODS, the ontological model of the openDS data model) within the ontologies of the OBO (Open Biological and Biomedical Ontology) Foundry¹⁵, namely the *Ontology for Biomedical Investigations (OBI)*, *Biological Collections Ontology (BCO)* and the *Information Artifact Ontology (IAO)*, it was decided not to focus on this approach for the initial version of openDS.

03 Outlook

The meetings of the openDS working group will continue on a regular basis. The data model will be further expanded based on the standards outlined in this report. A first full release of the openDS data model will be launched towards the end of the DiSSCo Prepare project phase, as part of the Deliverable D5.3 DiSSCo Digital Specimen Object Specification (due in project month 36).

¹³ GBIF/TDWG Multimedia Resources Task Group. 2013. Audubon Core Multimedia Resources Metadata Schema. Biodiversity Information Standards (TDWG) <http://www.tdwg.org/standards/638> ; or <https://github.com/tdwg/ac>

¹⁴ W3C PROV Working Group. 2013. PROV Model Primer. <https://www.w3.org/TR/prov-primer/>

¹⁵ <https://github.com/DiSSCo/openDS/blob/master/ods-ontology/ods-ont-intro.md>