

Implementation Experience Report for the Developing Latimer Core Standard: The DiSSCo Flanders use-case

Lissa Breugelmans[‡], Maarten Trekels[‡]

[‡] Meise Botanic Garden, Meise, Belgium

Corresponding author: Lissa Breugelmans (lissa.breugelmans@plantentuinmeise.be)

Academic editor: Gail Kampmeier

Keywords

TDWG, collection descriptions, dashboard, natural science collections

Introduction and background

Natural science collections are a primary resource for mapping out the world's biodiversity through the long-term preservation of collected specimens (Buschbom et al. 2022). Although significant efforts are ongoing in digitising this important contribution to our knowledge, many collections are still not or far from being digitally available to science. In order to ensure that valuable collections are findable to the community, there is a clear need for a standardised approach to describing the content of collections (Johnson and Owens 2023). Especially, the smaller collections often remain unknown and risk neglect or even disappearance.

To accomplish this goal, it is important to facilitate interoperability between major registries holding information on the collections and institutions, for example, the Global Registry of Scientific Collections ([GRSciColl](#)), [Index Herbariorum](#), the registry of the Consortium of European Taxonomic Facilities ([CETAF registry](#)) and the Distributed System of Scientific Collections ([DiSSCo](#)). The development of the Latimer Core standard is aimed at increasing the FAIRness ([Findable, Accessible, Interoperable and Reusable](#)) of data on collections (Woodburn et al. 2022).

This implementation experience report is initiated from the DiSSCo Flanders*¹ use case. The DiSSCo Flanders project is preparing the Flemish collections for the European DiSSCo research infrastructure (Trekels et al. 2022). DiSSCo Flanders will address biological, anthropological and geological collections, comprising preserved, living, tissues and molecular collections at the regional level. The consortium is comprised of the Flemish universities, research institutions and an association of botanical gardens and arboreta.

The Federal Belgian collections are associated with the project to ensure aligned policies and procedures (Fig. 1). The goal is to increase the digital visibility of the collections, ranging from the institutional level down to the specimen level. At the specimen level, the consortium already makes digitised specimens available through the Global Biodiversity Information Facility ([GBIF](#)) as soon as possible. However, there was a clear need to be able to describe them at a higher organisational level.

Development of the Latimer Core standard

Based on the earlier work of the Natural Collection Descriptions ([NCD](#)) group, the Biodiversity Information Standards ([TDWG](#)) [Collection Descriptions Interest Group](#) is developing the Latimer Core standard (Woodburn et al. 2022). Over a period of more than four years, weekly virtual meetings were held to develop the standard. On top of this, sessions and workshops were organised at relevant conferences (including [BiodiversityNE XT](#) and the TDWG working group sessions) in order to collect use cases for the standard.

In order to facilitate the development of the standard, the approach was taken to create a GitHub issue*² for each of the classes and terms within the standard (Norton et al. 2023). This allowed the group to track all discussions and changes that were taking place during the development.

During the development phase, it was clear early on in the process that a need existed to implement real-world examples using the standard. Wikibase*³ was used as an experimental tool to describe collections using the current terms that were available in the standard (Trekels et al. 2020).

Implementation in DiSSCo Flanders

As stated above, the DiSSCo Flanders project aims at obtaining high-level information on the natural science collections held in the institutions that participate in the DiSSCo Flanders project. This information consists of quantitative data on the overall size of the collections, as well as size by taxonomic groups, preservation types, stratigraphic age, geographic region and level of digitisation (Van Baelen et al. 2022). Based on previous work done in the Synthesis of Systematic Resources (SYNTHESYS+) project (Smith et al. 2019), a survey was designed to retrieve relevant information about the collections (Van Baelen et al. 2022). Although the survey served as one of the use cases of the standard during the development phase, the survey remained static over time while the development of the Latimer Core standard was undergoing major changes. In order to ensure the interoperability of the collected data, a mapping exercise was performed using the current terms and concepts of the proposed standard.

The data were extracted from the original survey spreadsheets*⁴ and pivoted into a vertical format using Microsoft Power Query. A data model for a MySQL database was developed (Fig. 2), taking into account the hierarchical nature of the data and using Latimer Core

terms for table names and attributes (Breugelmans and Trekels 2023). The database was subsequently populated with the survey data*⁴ and used as input for a Microsoft PowerBi dashboard*⁵ which features a graphical overview of the content and digitisation level of the Flemish collections (Fig. 3).

Survey format

Despite the survey being developed with the (preliminary) standard in mind, some characteristics of the design made data extraction, import into the SQL database and analysis more time-consuming than needed. Data on the collections were gathered on two levels. Collections were subdivided, based on their biogeographical origin and the following metrics were recorded: number of objects digitised, number of objects not digitised (documented), number of objects not digitised (not documented) and total number of objects. On a higher level, collections were grouped over all geographic origins and the same measurements were recorded, in addition to: number of objects with images, number of type specimens and number of specimens per MIDS level (Minimum Information on a Digital Specimen, Haston and Chapman (2022)). Levels range from MIDS-0 (minimum level of information that makes a connection between a physical specimen with its identifier and an entry in a database) to MIDS-3 (rich specimen information available). It would have been easier for analysis and visualisation purposes and less error-prone if the total number of objects could have been calculated from the other metrics instead of recorded as a separate metric and if the metrics could be aggregated over the subcollections to yield the higher-level metrics (instead of recording them separately). In addition, specimen counts by stratigraphic period were surveyed as stand-alone data, leading to redundancy in the database.

Latimer Core terms

In general, most of the data could be relatively easily mapped to the Latimer Core terms*⁶. In order to build up the MySQL database, tables were named after LtC classes*⁷ and fields after LtC properties*⁷. The smallest distinct collection subdivisions for which we had recorded metrics were determined and entered as instances of the ObjectGroup table. Subsequently, institute, specimen counts, biogeographical origin, as well as discipline and taxonomic group were split off into separate tables (respectively, the OrganisationalUnit, MeasurementOrFact, EcologicalContext and ObjectClassification tables). Specimen counts by stratigraphic period were stored by entering additional instances of the ObjectGroup and MeasurementOrFact table and their stratigraphic periods in the GeologicalContext table. Finally, additional instances of the ObjectGroup table for collection departments were linked with total specimen counts in the MeasurementOrFact table, curator information in the PersonRole and Person tables and time period of specimen collection in the Event and TemporalCoverage tables.

While implementing the standard for the first time, it was unclear where to map the terrestrial-freshwater-marine origins of the specimens, as well as the geographical

concepts that were used to describe units smaller than continents, but larger than countries or regions. In the meantime, however, an additional class, EcologicalContext (properties biomeType and biogeographicRealm), has been added to address this gap.

Several terms were defined as potentially multi-value (JSON array) fields. However, for the purpose of building the PowerBI dashboard, we were not able to find a way to join tables that used multi-value fields to extract the necessary data (PowerBI queries are constructed through its graphical user interface (GUI), using its own query language). Therefore, we introduced additional fields in the referenced tables in order to create a single value field that refers back to the parent table (e.g. a new field ofObjectGroup in the MeasurementOrFact table replaces the hasMeasurementOrFact field in the ObjectGroup table). We are unsure if the decision to work with multi-value fields was made for specific reasons (performance-related or other), but allowing for the relationship field in the other table might increase flexibility.

For the temporalCoverage class, the Latimer Core documentation suggests leaving the property EndDate blank when the collecting period is still currently running. There is, however, no term defined to use when the period is unknown, which might lead to confusion.

Finally, it would be useful to define controlled vocabularies for the classes and properties that are newly defined for the Latimer Core standard, in order to further enhance interoperability of the data. For certain properties, the use of the controlled vocabulary might be recommended but not mandatory, in order to allow for flexibility.

Conclusions

The DiSSCo Flanders use case surveyed the content of regional Flemish collections. The smaller research collections and living plant collections typically had only limited or no online representation of their content. Even a rough inventory of many collections was lacking. The standardised survey ensured that the content of the collections can be evaluated against each other. This also made it possible to have a graphical representation of the collections through a PowerBI dashboard, which is instrumental in increasing the visibility of the collections for scientists and policy-makers.

Although the survey design proved to be suboptimal with respect to the current version of the Latimer Core standard^{*6}, in general, it was manageable to map the survey results to the data standard. For data fields where this was not possible, we discussed them within the TDWG Collection Description Interest Group, which led to the proposed addition of the EcologicalContext class^{*8}.

From the DiSSCo Flanders use case, four recommendations can be formulated. First, the suboptimal design of the survey shows that there is a clear need to create guidance on performing this kind of exercise. Future surveys in other consortia and institutions could clearly benefit from having a design blueprint for the survey. This is, however, an endeavour that should be performed at a larger scale with many problems and pitfalls.

Large scale infrastructures, such as the future DiSSCo infrastructure in Europe or the [iDigBio](#) (Integrated Digitised Biocollections) initiative in the United States, have to play a key role in providing tests at a larger scale. The tools and training material that are created with this effort should be disseminated and maintained by these infrastructures. Secondly, it is advisable to further develop controlled vocabularies for the newly-adopted Classes and Properties in order to maximise the interoperability of the data. In order to make the data available on a worldwide scale, the third recommendation is that the LatimerCore standard is implemented in the main collection registries (e.g. GRSciColl, CETAF registry). Finally, making it easy for institutions to publish a Latimer Core record once in a registry, would reduce the redundancy for collections to fill out and modify their records in several places.

Acknowledgements

The authors are grateful for the discussions around the implementation with the TDWG Collections Descriptions Interest Group. The authors also sincerely appreciate the time and effort invested by the reviewers, Barbara Thiers and Thomas McElrath and the technical editor, Gail Kampmeier, whose insightful remarks and thoughtful suggestions significantly improved the quality of the manuscript.

Funding program

The work presented in this report was funded by the Research Foundation – Flanders (FWO) as part of the Flemish contribution to the DiSSCo Research Infrastructure under grant n° I001721N (DiSSCo Flanders project).

Conflicts of interest

The authors have declared that no competing interests exist.

References

- Breugelmans L, Trekels M (2023) Latimer Core Implementation Experience report: the DiSSCo Flanders use-case. Zenodo <https://doi.org/10.5281/ZENODO.8214927>
- Buschbom J, Acosta-Buenaño N, Chocho V, Irazábal JR, Macklin J, Ojeda Cabrera LM, Plos A, Tancoff S, Teta P, Valle C, Freire-Fierro A (2022) Campaigning for a Global Collections Network: Improving the digital representation and visibility of natural science collections from Latin America and the Caribbean. *Biodiversity Information Science and Standards* 6: e91356. <https://doi.org/10.3897/biss.6.91356>
- Haston EM, Chapman C (2022) MIDS: The digitisation standard for Natural Science collections. *Biodiversity Information Science and Standards* 6: e94604. <https://doi.org/10.3897/biss.6.94604>
- Johnson K, Owens IP (2023) A global approach for natural history museum collections. *Science* 379 (6638): 1192-1194. <https://doi.org/10.1126/science.adf6434>

- Norton B, Engelbrecht I, Jones J, Buschbom J, Webbink K, Trekels M, Woodburn M, Sanderson R, Vincent S, Grant S, Baskauf SJ, Biodiversity Information Standards (TDWG) (2023) Latimer Core (LtC) GitHub Repository. URL: <https://github.com/tdwg/ltc>
- Smith V, Gorman K, Addink W, Arvanitidis C, Casino A, Dixey K, Dröge G, Groom Q, Haston E, Hobern D, Knapp S, Koureas D, Livermore L, Seberg O (2019) SYNTHESIS+ Abridged Grant Proposal. Research Ideas and Outcomes 5: e46404 . <https://doi.org/10.3897/rio.5.e46404>
- Trekels M, Woodburn M, Paul DL, Grant S, Webbink K, Jones J, Groom Q (2020) How do you Develop a Data Standard? Wikibase might be the Solution.... Biodiversity Information Science and Standards 4: e59211. <https://doi.org/10.3897/biss.4.59211>
- Trekels M, Beirinckx L, Claerhout T, Dugardin C, Leliaert F, Pereboom Z, Slos D, Van Baelen A, Vandepitte L, Veltjen E, Verstockt S, Mergen P (2022) DiSSCo Flanders: A regional natural science collections management infrastructure in an international context. Biodiversity Information Science and Standards 6: e94350. <https://doi.org/10.3897/biss.6.94350>
- Van Baelen A, Poot N, Beirinckx L, Bogaerts A, Bellefroid E, Claerhout T, De Smedt S, Dugardin C, Engledow H, Leliaert F, Ossaer J, Pereboom Z, Semal P, Slos D, Smirnova L, Vandepitte L, Veltjen E, Trekels M (2022) DiSSCo-Flanders WP2 - task 2.1, Detailed inventory of the collections: Report. Zenodo. Version Number: Version 1, May 2022 DOI: 10.5281/ZENODO.6511351. URL: <https://zenodo.org/record/6511351>
- Woodburn M, Buschbom J, Droegge G, Grant S, Groom Q, Jones J, Trekels M, Vincent S, Webbink K (2022) Latimer Core: A new data standard for collection descriptions. Biodiversity Information Science and Standards 6: e91159. <https://doi.org/10.3897/biss.6.91159>

Endnotes

*1 <https://diSSCo-flanders.be>

*2 <https://github.com/tdwg/cd/issues>

*3 The Wikibase cloud environment was updated through time, going from an experimental set-up to a service provided by Wikimedia Germany. This resulted in updated URLs for the wikibases. Currently the sandbox wikibase is hosted at <https://tdwg-cd.wikibase.cloud/>. A more up-to-date version of the standard is implemented at <https://latimer-core.wikibase.cloud/>

*4 The original survey is located at <https://zenodo.org/records/6511351>

The populated MySQL database can be found at <https://doi.org/10.5281/ZENODO.8214927>

*5 This interactive dashboard provides an overview of the nature and size of the collections that each institute houses. In addition, it also provides information on the geographic origin of the specimens in the collections and on the degree to which the collections are digitised.

This dashboard will be integrated in the website of DiSSCo Flanders*¹ and will enhance the visibility of lesser-known collections for scientists, policy-makers and the general public.

The dashboard can be accessed directly [through the following link](#).

*6

<https://github.com/tdwg/lc/>

*7 <https://github.com/tdwg/lc/wiki/2.-Glossary>

*8 <https://github.com/tdwg/lc/wiki/4.-Classes>



Figure 1.

Overview of the DiSSCo Flanders consortium. Participating partner institutions: Flanders Marine Institute ([VLIZ](#)), Ghent University ([UGhent](#)), Flanders Research Institute for Agriculture, Fisheries and Food ([ILVO](#)), University of Antwerp ([UAntwerp](#)), Royal Zoological Society of Antwerp ([KMDA](#)), Botanic Garden Meise ([MeiseBG](#)), Katholieke Universiteit Leuven ([KULeuven](#)), Research Institute for Nature and Forest ([INBO](#)), Vrije Universiteit Brussel ([VUB](#)), The Belgian Association of Botanic Gardens and Arboreta ([V.B.T.A.](#)), University of Namur ([UNamur](#)), Université Libre de Bruxelles ([ULB](#)), Royal Museum for Central Africa ([RMCA](#)) and Royal Belgian Institute of Natural Sciences ([RBINS](#)). Figure by Frederik Leliaert under [CC BY 4.0](#).

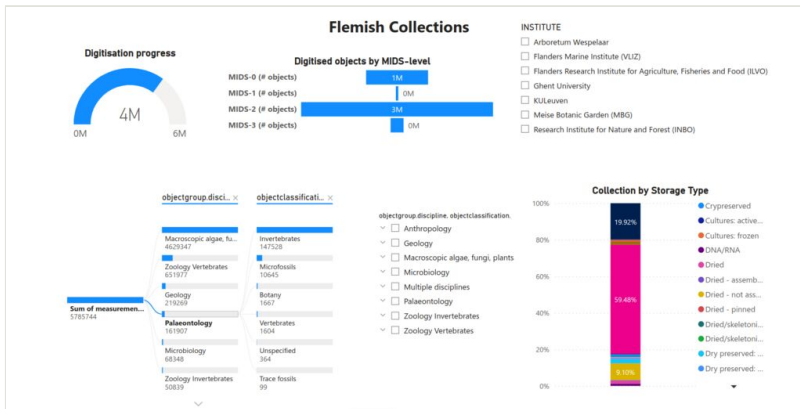


Figure 3.

Screenshot of the landing page of the DISSCo Flanders PowerBI dashboard. Figure by Lissa Breugelmans under [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/).