OPEN ACCESS

Conference Abstract

Latimer Core: A new data standard for collection descriptions

Matt Woodburn‡, Jutta Buschbom§, Gabriele Droege|, Sharon Grant¶, Quentin Groom#, Janeen Jones ¶, Maarten Trekels#, Sarah Vincent‡, Kate Webbink¶

‡ Natural History Museum, London, United Kingdom
§ Statistical Genetics, Ahrensburg, Germany
| Botanic Garden and Botanical Museum Berlin, Berlin, Germany
¶ Field Museum of Natural History, Chicago, United States of America
# Meise Botanic Garden, Meise, Belgium

Abstract

The Latimer Core (LtC) schema, named after Marjorie Courtenay-Latimer, is a standard designed to support the representation and discovery of natural science collections by structuring data about the groups of objects that those collections and their subcomponents encompass. Individual items within those groups are represented through other emerging or current standards (e.g., Darwin Core, ABCD). The LtC classes and properties aim to represent information that describes these groupings in enough detail to inform deeper discovery of the resources contained within them.
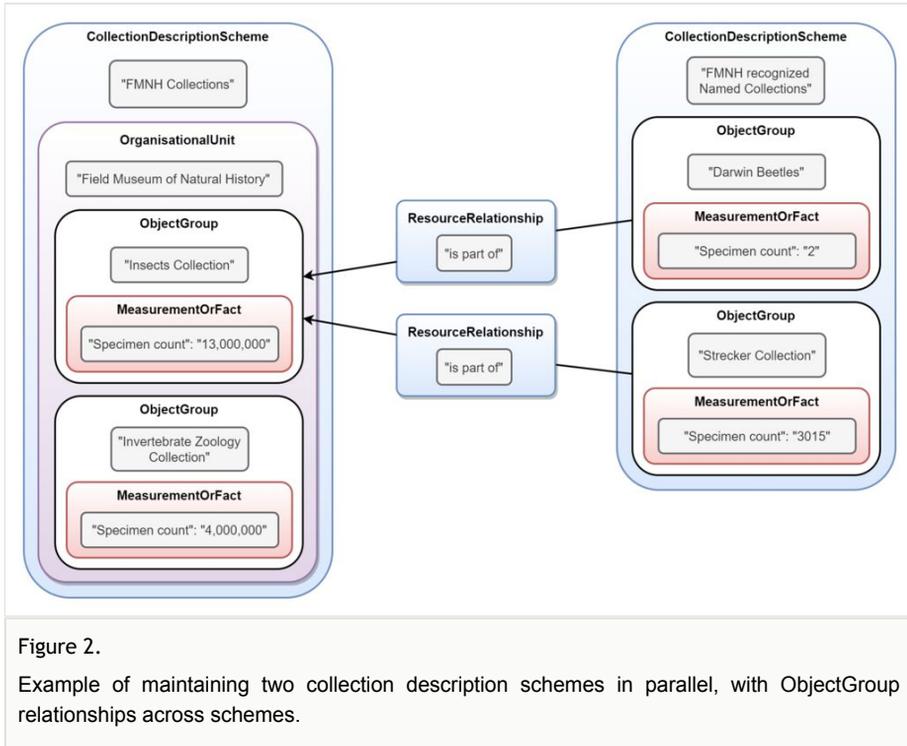
The standard has been developed under the Biodiversity Information Standards (TDWG) Collection Descriptions (CD) Interest Group, and evolved from the earlier work of the Natural Collection Descriptions (NCD) group. Version 1 of the standard includes 23 classes, each with two or more properties (Fig. 1 and Suppl. material 1).

Figure 1.

A visual summary of Latimer Core classes.
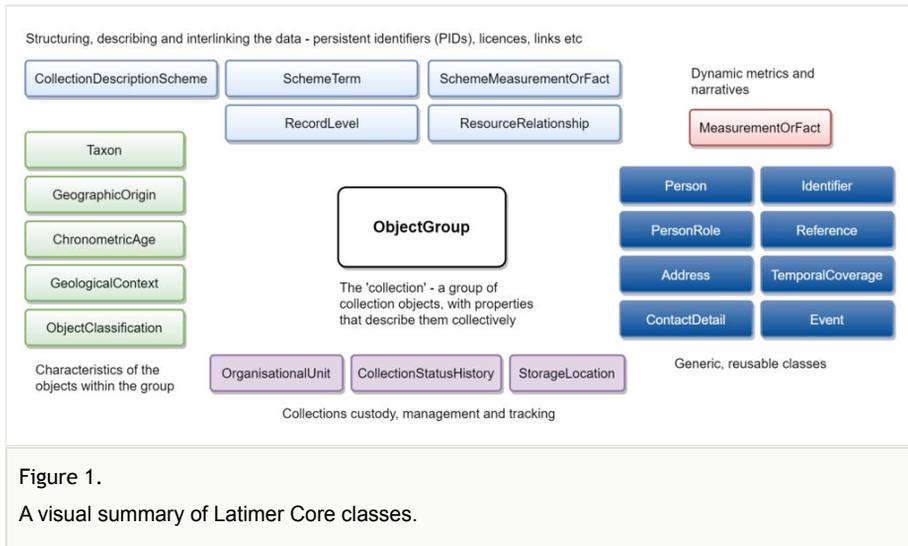
The central concept of the standard is the ObjectGroup class, which represents 'an intentionally grouped set of objects with one or more common characteristics'. Arranged around the ObjectGroup are a set of classes that are commonly used to describe and classify the objects within the ObjectGroup, classes covering aspects of the custodianship, management and tracking of the collections, a generic class (MeasurementOrFact) for storing qualitative or quantitative measures within the standard, and a set of classes that are used to describe the structure and description of the dataset.

Latimer Core is intended to be sufficiently flexible and scalable to apply to a wide range of collection description use cases, from describing the overall collections holdings of an institution to the contents of a single drawer of material. Various approaches are used to support this flexibility, including the use of generic classes to represent organisations, people, roles and identifiers, and enabling flexible relationships for constructing data models that meet different use cases. The collection description scheme concept is introduced to enable adopters to specify rules in the use of LtC within each specific implementation, demonstrated in Fig. 2. Guidance and reference examples for different modelling approaches to suit different use cases are provided in the LtC guidance documentation.

The LtC standard has significant overlap with existing data standards (Suppl. material 2) that represent, for example, individual objects and occurrences, organisations, people and activities. Where possible, LtC has either borrowed terms directly from these standards or less formally aligned with them. Achieving a balance between offering a standard that is sufficiently comprehensive to stand alone and maintains a low technical barrier to adoption whilst minimalising duplication of effort in the context of the wider standards landscape is a notable challenge in the standard development process.

Figure 2.

Example of maintaining two collection description schemes in parallel, with ObjectGroup relationships across schemes.

The draft standard was submitted to the TDWG Executive in June 2022 to begin the process of formal review and ratification. This includes a list of standard terms and a GitHub wiki of guidance on the concepts behind and use of the standard. In the meantime, the Task Group will continue working on reference examples and serialisations, and working with infrastructures such as the Distributed System of Scientific Collections (DiSSCo) consortium, the GBIF (Global Biodiversity Information Facility) Registry of Scientific Collections, the CETAF (Consortium of European Taxonomic Facilities) Registry of Collections and the Global Genome Biodiversity Network (GGBN) on potential roadmaps towards adoption.

In this presentation, we will introduce the key Latimer Core deliverables, highlight some of the challenges faced in the development process, and discuss the potential for community adoption.

Keywords

natural science, TDWG

Presenting author

Matt Woodburn

Presented at

TDWG 2022

Funding program

Supplementary materials

Suppl. material 1: A summary of the classes in the Latimer Core standard. `doi`

Authors: Jutta Bushbom, Gabriele Droege, Sharon Grant, Quentin Groom, Janeen Jones, Maarten Trekels, Sarah Vincent, Kate Webbink, Matt Woodburn
Data type: standard term definitions
Brief description: A summary of the classes in the Latimer Core standard, with links to the normative definitions and related GitHub issues.
Download file (6.01 kb)

Suppl. material 2: Standards with LtC alignments `doi`

Authors: Jutta Bushbom, Sharon Grant, Quentin Groom, Janeen Jones, Maarten Trekels, Sarah Vincent, Kate Webbink, Matt Woodburn
Data type: standards list
Brief description: A list of standards from which Latimer Core borrows terms, or with which aspects of the standard are aligned.
Download file (472.00 bytes)