

DARIAH Collection Description Data Model (DCDDM)



Table of Contents

- [XML schema and more documentation](#)
- [Introduction and design principles](#)

XML schema and more documentation

<https://github.com/DARIAH-DE/DCDDM>

Introduction and design principles

The **DARIAH Collection Description Data Model (DCDDM)** specifies a fixed number of classes, elements, attributes, and values used for describing collections. Following the definition of the Dublin Core a collection is “an **aggregation of resources**”, which can be **digital objects** like digitized paintings, books, manuscripts, as well as “**born-digital**” objects like records of metadata, database-entries, or transcripts) or physical ones (paintings, books, statues, stamps, coins, etc.).

It is, however, not always so simple to identify what precisely the collection's resources are. Let's take for instance a collection of early printed books where books' metadata are collected and made public by an “Online Public Access Catalogue” (OPAC). By describing this OPAC as I describing a collection of (physical) books or the collection of respective metadata (author, year of publication, number of pages, signature, ...)? On the one hand we have the common user of an OPAC interested mainly, if not exclusively, in the physical object. He would search for the dates when the books were published (created) rather than look up the date of creation of the metadata records. On the other hand other users might want to harvest the books' metadata and would require therefore access to information about the metadata (for example the date when the metadata records have been created). So the interest of the users who are targeted by the OPAC influences the information the collection's description contains.

But how can we tailor the user's access without knowing who is going to use our collection description and for which purposes? One of the main goals of the DCDDM is to address this issue by providing ways to pack as much information into one collection description as possible and reasonable.

One of the challenges one encounters almost immediately is the most often unclear or not uniform understanding of the term 'collection' and collection's items. Furthermore there is no strict differentiation between the collection's items (e.g. books), the item's metadata (e.g. the information about the author and title) and the actual manifestations of these information in form of, for instance, bitstreams saved in database. Another considerable aspect of providing such a broad range information is that the process of gathering all the relevant information can be very time consuming. Following the DCDDM requirements one can create a valid description by adding only a handful of basic information. This means, for example, that a given collection can have descriptions which vary strongly in their level of detail. A major factor in this respect is the descriptor's vision about the description and his or her willingness to invest time and effort to transform a valid but very basic collection description into a top notch collection description.

On the other hand, the level of detail or number of providable information depends also on the collection itself, the collection's items and the infrastructure which maintains the collection. Institutions, which have at its disposal sufficient personnel and funds, have better resources to describe in detail their collections, than would an independent scholar with his or her collection of notes made while writing a small article.



To ease the workload of describing collections the DCDDM is meant to be used as a **data model for applications dealing with the creation, publication and administration of collection descriptions**. Likewise the DCDDM should provide sufficient **guidelines to create collection descriptions** independent of any tool or applications and their unforeseeable lifetime.

Given the broad definition of collection and the consequently many possible varieties of collections, the DCDDM has to achieve the following objectives:

1. The data model should allow descriptions of digital AND physical collections.
2. The description should offer general descriptive information concerning the content of the collection or the collection's items, such as:
 - a. topics/subjects which are covered by the collection;
 - b. locations and regions the content of the collection could be associated with;
 - c. dates and periods the collection may be related to.
3. The data model should allow for an unambiguous identification of the described collection.
4. The data model should allow to contextualize a collection and mark relationships:
 - a. between collections (one collection being part of another collection);
 - b. between collections and agents who interact with the collection (such as owners or curators);
5. Collection descriptions based on the DCDDM should encourage the exploration and usage of the described collections. Therefore the description should provide administrative information about:
 - a. how to access the collection or its items;
 - b. possible legal restrictions concerning the possibilities to access the collection and to (re)use its items.
6. If collections are made accessible online, technical metadata should be provided so that other services may gain access to the collection's objects or the object's metadata.
7. To avoid ambiguity the data model encourages the use of controlled vocabulary and normative data, unique identifiers, and syntax encoding schemes. To ensure human readability of the collection description the data model provides label-elements.
 - a. To facilitate mappings and transformations of already existing collection descriptions to collection descriptions based on the DCDDM, DCDDM supports several controlled vocabularies and encoding schemes (in full awareness of the downsides of such a policy).

8. The data model should enable multilingual descriptions in a flexible but still practical way. This means that the describing person can choose the language for a given information. To achieve this requirement, the elements providing information in form of natural language and free text may be reused in as many languages as wished.
9. The data model has to be flexible enough to describe well curated collections held in museums, libraries, archives, etc. as well as material collected by individual researcher stored on personal hard drives. To archive this requirement the number of mandatory elements is kept as low as possible.
10. To achieve a high and widespread acceptance the data model should allow at least one easy to produce, read and understand serialization (like a csv-table) or a hierarchically as flat as possible (such as an XML-file, validated by a matching XML-Schema).
11. To ease interoperability and to reveal semantic closeness to well-known and extensively used metadata records (e.g. Dublin Core) the names and definitions of many classes, elements and attributes should be taken from well-established ones. Nonetheless to avoid any (technical) dependencies these names, definitions and specification will not be just referenced by simply providing quotes or links to the referenced elements, but will also be written down in this document.
12. The DCDDM wishes to be an easy to use, but nevertheless functional and therefore widely accepted data model for collection descriptions. To archive this goal it has to be extensively tested and evaluated 'in the wild'. At the moment of writing this specification, the DCDDM is mainly the result of a more theoretical discussion. This means that the model described in the paper has to be understood as an invitation to the interested community to use it, test it but most and for all criticize it.