Several communities have expressed interest in making use of aggregations of objects with particular focus on building such aggregations through persistent identifiers and providing PIDs for aggregation objects.

Building collections within diverse domains and then sharing or expanding them across disciplines should enable common tools for end-users and e-infrastructure providers.

Individual disciplinary communities can directly benefit if such tools are made widely available, and cross-community data sharing can benefit from increased unification between collection models and implementations. PID providers may benefit from marketing additional services on collections.

The WG is working in close collaboration with scientific communities to come to develop a data model together with an accompanying API that provides a pragmatic and usable solution.

What is a collection?

Imagine you have a number of objects that belong together. The type of object is a bit flexible, as long as it is in some digital form; this can include digital documents or scientific articles, individual data files, a zip of several files, digital images, audio or video recordings.

The reasons why these objects belong together are just as diverse. There may for example be a number of files that came out of a scientific model calculation, or a number of recordings from a study session or a very distinctively selected choice of files grouped together for a particular analysis.

What can we do with it?

A familiar concept from computer programming are abstract data types such as lists, arrays and sets, andwe know how to add, insert, replace or delete objects from such constructs.

While we often talk about research objects that are less defined and more complex than computational data structures, we still want to apply the same operations: Put objects in the collection, take them out again, learn about the number of objects and their total size, look at all objects in the collection in an orderly manner and so on.

Depending on the domain we may want to apply constraints on the collection, such as whether its objects are ordered or unordered, or whether there are further hierarchies inside it.

Definition

A collection is a digital object which is identified by a PID and consists of a set or a list of PIDs/Ids and a set of additional pointers/links and metadata together with each PID/Id. A collection can be given explicitly by naming each PIDs/Id directly as well as implicitly by a generating rule. A collection is called finite, if the set of PIDs/Ids, generated by iteratively resolving its "sub-"collections, is finite.