# Comments on the Paris doc

Tobias Weigel, Michael Lautenschlager  
2015/08/11

## Data Management (2.3)

We agree with previous comments that this section needs further refinement.

If a huge number of Digital objects receive identifiers that are persistent for at least the object's lifetime, we need adequate management tools and accepted processes that allow for a significant amount of automation of typical data management tasks. A first suggestion for a principle such tools should support is to move away from treating every object in singular form and work towards repeatable actions on larger numbers of similar objects. The reasoning here is that the increasing number of objects must still be managed with the same limited amount of resources.

The document already proposes to use persistent identifiers as the primary tokens for object access. Establishing identifier services at large scale is however a costly and time-intensive effort; to some extent, this is so because such identifiers must be managed just like the objects themselves. Establishing management for identifiers only therefore seems wasteful: Rather, the management of objects and identifiers should work through the same mechanisms as much as possible. To enable management of objects beyond a view focusing on single items, adequate mechanisms should, for example, be able to select objects by their most important characteristics or aggregate them at multiple levels of granularity and provide basic CRUD operations on such object collections. This should be part of a step-by-step transition that leads from classic file systems to a PID-based data organization approach. Putting forward such a strategy should help with building acceptance for PID-based solutions and addressing scalability concerns.

## Persistent Identifiers (2.4 / 5.1)

The document should include a viewpoint that for PIDs to provide added value over other forms of identification, we will need smarter resolvers that offer additional services beyond getting from an identifier to an object location. Examples for such services include being able to retrieve an object’s metadata or licensing information, to learn about possible processing services or aggregation mechanisms. Registries (section 5.8) for such added-value services at the resolvers’ level are also needed and should be maintained by recognized international organizations.

## Organisational approaches, particularly at the regional/international level (6.2 / 6.3)

As a consequence of the high value that is put in persistent identifiers and their relation to data management as explained above, we also need coherent organizations that support such approaches.

An operational setup for such organizations must find a compromise between two goals:

1. Different communities have varying requirements which cannot be ignored. Policies that are essential to one community (such as long-term availability of identifiers for well-curated objects) may be a hindrance for others (who want to manage large numbers of objects with a limited lifetime).
2. Nonetheless, it is inefficient to establish completely separated organizations for these purposes. The basis for agreement should be that digital objects need identifiers and the provisioning of such identifiers must be well managed. In addition, there should be an open set of operational service providers that offer operational policies and added-value services geared towards specific community needs. There can and should be some competition among these service providers; however, the general principles of object identification and value-added services should remain a common element. Already existing legacy systems must be integrated to reduce the impact on the scientific communities that currently use them and operate services depending on them.

Providing general identifier services for different kinds of entities is a commonality stretching across initiatives such as DONA/EPIC, THOR and EUDAT. These organizations should come to a coherent view and sustainable business model that acknowledges both points 1 and 2.