

Metadata Quality Improvement*

Background and Description

One of the aims of the DASISH project was to analyse and compare the different metadata strategies of CLARIN, DARIAH and CESSDA, and to identify the possible mutual benefits from cross-fertilization of approaches. To support this analysis the context was defined in terms of metadata types and quality criteria and a structure was created which extended common lifecycle models to address metadata issues.

We looked at the metadata policies and strategies of the three infrastructures and evaluated these in terms of metadata quality against the Bruce and Hillmann criteria. Additionally we described in more detail how the individual data repositories within the different infrastructures implemented metadata management.

Findings & Recommendations

The infrastructures of CESSDA, CLARIN and DARIAH differ in visions, strategies and initiatives regarding metadata issues; similarly there is a difference in metadata management among the various repositories. Despite these differences a number of elements of the intensions, plans, and initiatives touch upon the same issues and challenges for metadata. This opens the potential for cross fertilisation.

One recommendation would be that the three infrastructures could agree to define a common list of metadata elements that - crossing the different communities and standards – can be used as compatible between the different communities.

Furthermore, easily accessible definitions of these elements and mappings across the different metadata standards should be

available. Moreover, sharing of knowledge and linking resources would be beneficial for all the three infrastructures.

Especially the sharing of knowledge about linked data initiatives concerning discovering vocabularies, Simple Knowledge Organisation Systems (SKOS) and references to definitions of data categories, currently active or planned in all three infrastructures, would leverage these developments.

Evaluation of the prototype of the joint CESSDA, CLARIN and DARIAH metadata portal endorses the opinion that more coordination would be beneficial for the metadata quality.

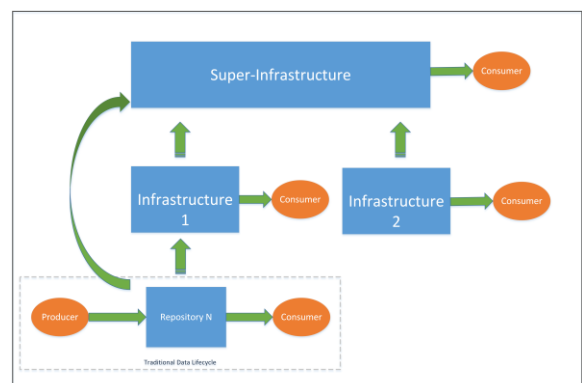
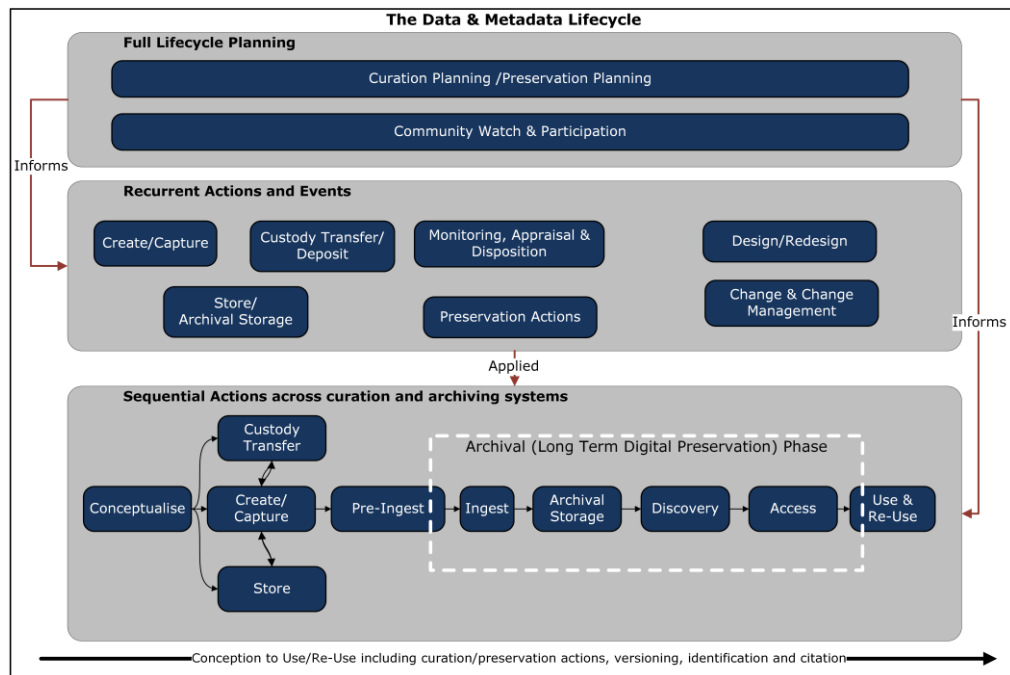


Diagram of cooperating infrastructures

Metadata Lifecycle

Metadata design, redesign, creation and management can continue to be ‘live’ issues for those preserving or providing access to data even when the data itself remains unchanged. Most views of the research data lifecycle tend to treat data as fairly ‘static’ from the point of ingest into an Archive until the next Access/Use/Re-use cycle but repositories must apply new or update existing standards and re-enrich metadata to meet the changing needs of their target community. To support these more dynamic metadata issues we adapted existing research data lifecycles.

Metadata Lifecycle



This metadata lifecycle may serve as a baseline, which, alongside an understanding of metadata quality evaluation and metadata types, can be used to design and benchmark a local approach to describing, delivering and improving metadata quality. The metadata lifecycle aligns with the OAIS model, but places it in a wider context. It consists of three levels of activities:

Full life cycle planning

Communication is a key aspect within the metadata lifecycle. To deliver planning across the lifecycle the outcomes of *Community Watch and Participation* must be integrated into *Curation/Preservation Planning* processes. If your goal is to serve a community then the starting point is to engage with and understand that community. Good planning, communication and practice throughout the lifecycle reduce costs and complexity and contribute to improved quality.

Recurrent actions and events

A number of data/metadata related activities occur numerous times during the lifecycle of a digital object; these benefit from centralised design and planning so they can be implemented coherently, thereby supporting consistency and quality. These activities are defined by *Curation/Preservation Planning*, often influenced by *Community Watch and Participation*.

Sequential Actions across curation and archiving systems

Despite the fact that 'circular' approaches display more of the innate complexities of the process, the 'birth to re-use' sequence is commonly understood and support communication in day-to-day business processes. This follows the traditional research data lifecycle stages.

A more detailed description of the metadata lifecycle can be found in the Deliverable 5.2 A of WP5: Metadata Quality Improvement, available at: <http://dasish.eu/publications/projectreports/>