White paper - “New approach for the Dutch digital heritage infrastructure”

In order to improve the visibility of digital heritage information institutions should work on aligning their metadata with the reference sources provided on (at least) a national level. By building on previous work (OpenSKOS, erfgeo.nl, CultuurLink) and emerging tools in the market (like PoolParty) we started working on building a knowledge graph for digital heritage domain in the Netherlands. We hope to profit from similar work being done for the Digital Humanities/Linguistics research infrastructure (CLARIAH) and Europeana.

In the end we should have a reasonable set of definitions for common entities for 'who', 'where', 'what' and 'when' firmly connected to the 'world outside' using available linked data sources like DBpedia and others. At the same time we urge the suppliers of the registration systems to implement functionality that enables easy linking to these central entities but also support for proposing new or changed entities. We will organize a support infrastructure at central level for maintaining the contents of the knowledge graph.

We strongly believe that there is a need to build a national (possibly distributed) registry of datasets in order to give a more complete overview of the available heritage information. Currently there are a few incomplete registries that do not comply or interact with each other. We like to follow the standard Open Data approach (and technology like CKAN) and build a central registry that shows all the available datasets together with some administrative information like collection type, license model, technical information for accessing them, and maybe even some hints about the contents (by stating some links to one or more relevant terms).

Build on the network with terminology sources and the collection registry, and assuming we have unique (preferably persistent) identifiers for the resources that describe cultural heritage objects, we think it will be realistic to build a (distributed) object registry by only recording triples stating the relation between the resources URI and the related term URI. We consider using a synchronization protocol like Linked Data Notifications in order to achieve this. In our view the object registry should be completely agnostic of a semantic datamodel. This object registry will be the entry point for any service or portal that needs to aggregate or query data for a delivering a specific service. But even local systems could query the registry in order find similar objects while annotating a new object or make links to related objects from other collections.

Based on the relations described in the registry a portal service selects relevant datasets (or objects) for its purpose and starts a collection process lead by the information in the registry. Currently this process will be mainly based on OAI-PMH but we expect (and would like to stimulate) that information will be accessible as Linked Data, so direct access to the full resource is possible.

We will start building the registry information based on the currently available national and regional aggregation platforms but in our roadmap the local systems should at some point interact with registry directly. So the registry is not a replacement of the aggregation process itself for now but it will be a tool for improving the quality of the aggregation process and it

will force the institutions to align their information with the available references on the national level in order to become visible.

Anticipating the growing Linked Data support in the local systems we expect it should be possible to evolve to a situation where (real time) Linked Data querying will become dominant over the current copying and transformation approach. Our program is currently focussed on describing the specifications for the new infrastructure and evaluating the proof-of-concept in pilot projects.

For more information please contact me through: enno.meijers@kb.nl

Enno Meijers
Dutch Digital Heritage Network
March 1st, 2017