CMDI 1.2
Header WG
Status Report

Axel Herold & Oliver Schonefeld

BBAW & IDS Mannheim

herold@bbaw.de & schonefeld@ids-mannheim.de

CMDI Taskforce Meeting
Utrecht
2014-02-21

Issues

- Mandatory MdProfile header element
- MdType for indicating collection status in instance

Issue 1: MdProfile header

Problem:

"In CMDI 1.1, all header elements are optional including <MdProfile>. Because of this, and the fact that the xsi:schemaLocation attribute cannot be made mandatory either, it is possible to create schema valid CMDI instances that do not link back to the profile they are based on."

Issue 1: MdProfile header

Solution 1:
Make /CMD/Header/MdProfile a mandatory element path in all CMDI instance.

- Pros
  - With this solution, it is guaranteed that all valid metadata instances have the ID of their profile specified
- Cons
  - upgrading from 1.1 instances without a MdProfile header item may be complicated or in some (theoretical) cases impossible because the information is simply not available.
  - the correctness of the referenced profile (i.e. the value of MdProfile actually identifying the profile the instance was based on) is not enforced (this could be done through the schema but may be overly restrictive)

- Discussion
  - Should we allow "unknown"? No, because that would make MdProfile more or less optional again.

Issue 1: MdProfile header

- Proposed solution
  - Solution 1 (Make MdProfile mandatory)

- Centre Impact
  - Centres need to convert metadata records
  - Centres may need to manually add MdProfile, if profile cannot automatically determine profile by conversion script

- Open Issues/Questions
  - None

Issue 2: MdType header

Problem:

"Because the granularity of the existing CMDI files and collections varies a lot (from 1 CMDI = 1 large corpus to 1 CMDI file = single small textfile) it would be handy to indicate that certain CMDI descriptions are about a collection as a whole"

Issue 2: MdType header

Solution 1:
A new header element that indicates the type.

- Pros
  - Easy

- Cons
  - There can be a tendency to keep on extending the header instead of using CMDI's flexibility.

- Discussion
  - What would be the "fixed" set of types? "collection", „part_of_some_bigger_collection"? Or more then two, e.g. „archive", …?

Issue 2: MdType header

Solution 2:
One collection profile to be used by all.

- Pros
  - Easy

- Cons
  - Inflexible

- Discussion
  - Will there ever be a stable catch-all collection profile? Probably not …

Issue 2: MdType header

Solution 3:

Any collection level profile should contain a specific CLARIN collection component.

- Pros
    - Easy, may have low impact if the component is optional (but can be detected in the schema)
- Cons
    - What should be in the component?

- Discussion
    - Collection Profile might duplicate a lot of what is already needed for a collection and thus will probably bloat profiles.
    - What if this component becomes deprecated? Potentially, a lot of profiles would suddenly depend on a deprecated profile.

Issue 2: MdType header

Solution 4:

The profile root should use one of the data categories (DC) from a specific collection relation set in RELcat.

- Pros
  - Easy, low impact
- Cons
  - Data category might be unintentionally used

- Discussion
  - Backwards compatible with CMDI 1.1 (no new elements in header)
  - No need to change CMDI again, if "new" collection type was needed
  - Using a suitable DC for the profile root could encode the data provider's classification of the resource

Issue 2: MdType header

Solution 5:

Like done for web services collection CMD records can be explicitly requested by harvesting a center specific OAI-PMH set.

- Pros
  - No need to touch CMD profiles or instances
- Cons
  - VLO currently doesn't use endpoint information, i.e., the facet mapping can't select based on the OAI-PMH endpoint or set.

- Discussion
  - Collection information would be „outside of CMDI"
  - Rather a "hack", than a solution …

Issue 2: MdType header

- Proposed solution
  - Solution 4 (DC in root profile)

- Centre Impact
  - Centres may need to change their profile(s) and convert metadata records (if they wish to provide collection type information)

- Open Issues/Questions
  - What would be a set of suitable DCs? Currently, no "container DCs" in ISOcat