Determining metadata quality in Edurep

Wim Muskee
Kennisnet Foundation
EDRENE 2011
Edurep in Context

- harvesting more than 50 repositories
- > 800,000 records
- queried by more than 15 production portals
- more than 1,500,000 queries each month
Application Profiles

• We're getting 2 different application profiles of LOM, each with its own business rules.

• CZP
  • At least one classification field is needed.

• NL-LOM
  • A classification with purpose “educational level” is obligated when aggregation level is 3 or 4.
Metadatascheme

- The compliance to the application profile is listed in the field `/lom/metametadata/metadatascheme`

  `<metadatascheme>LOMv1.0</metadatascheme>`
  `<metadatascheme>nl_lom_v1p0</metadatascheme>`
No Validation

- In Edurep, only LOM xsd validation.
- Validation is needed, because the metadatascheme value often does not comply with the record.

- For example; a record says it's CZP compliant although there is no classification field.
Proposal

- Each record is validated against business rules of the two application profiles.
- Depending on outcome, the appropriate value is filled in metadatascheme of that record.
- When not compliant to an application profile, only the default “LOMv1.0” is filled in.
Consequences for …

- Search portals: filter results on quality metadata, for instance:
  
  `metadataschema exact "czp_povobve_v1p3"`

- Content providers: get an overview of the quality of their provided metadata

- Edurep: easily find content providers who provide low quality metadata and leverage to increase quality
More needed

- Vocabulary check, determine if provided vocabulary identifier value in LOM is actually present in the listed source.
- Dead link checker.