

Considerations for the Vocabulary Platforms

CLARIN: main requirements and best practices

Dieter Van Uytvanck

SSHOC Considerations for Vocabulary Platforms, 6 November 2020



Current concept registries in use by CLARIN ERIC



CLARIN

- Earlier on, ISOcat was used (2008..2015) - <http://datcatinfo.net/>
- Afterwards, migration to OpenSKOS for:
 - CLARIN concept registry – <https://www.clarin.eu/ccr> (since 2015)
 - CLAVAS (vocabulary service pilot) - <https://vocabularies.clarin.eu/clavas/> (since 2017)
- At this time, we are looking for a follow-up platform.

In search of a new vocabulary/concept platform – requirements



CLARIN

- Open Source
- In line with common standards (SKOS)
- Reasonable migration path from OpenSKOS
- Comes with a GUI and editor
- Mature solution
- Fast lookup via API to power all kind of editors
- Not requiring an exotic software stack
- Support for persistent identifiers
- Support for multilinguality

In search of a new vocabulary/concept platform – wishes/dreams



CLARIN

- Know-how and experience available through
 - a CLARIN centre
 - or a nearby organisation (e.g. national library)
- Solution that comes with major vocabularies preloaded and curated (eg ISO-639-3, IANA mime types, ...)
- Allow clustering/management of entries for specific sub-communities
- Broad developers base with long-term perspectives

How does CLARIN see the potential alignment between different vocabulary platforms?



- Multidirectional **synchronisation** to avoid double curation work
- **Linking** to other/external authority files
- Ideally: **compatible access APIs** for different systems

How will you address the interoperability challenges between the vocabularies?



- Reusing existing vocabularies wherever possible
- Relying on well-curated vocabularies with lots of sameAs links
- Stand on the shoulders of the giants and avoid reinvention of the wheel:
 - Realize that libraries have done enormous amounts of curation
 - Link to authority files (eg VIAF) wherever possible
 - Try to actively contribute the outcomes of community-curated entries