Use case - CLARIN ERIC/SSHOC

Automation of FAIRification and the SSH cluster: how diversity can underline the power of service federation

Franciska de Jong
f.m.g.dejong@uu.nl

CLARIN

CLARIN ERIC - www.clarin.eu



- European RI that gives access to language resources for users from the SSH domain and beyond
- a federation of data repositories since 2012; FAIR *avant la lettre*
- common metadata standard infrastructure (CMDI) for enhanced discoverability of resources via the Virtual Language Observatory (VLO)
- data that is diverse along multiple dimensions: region, modality, period, context

SSHOC –

SSHOpenCluster
Social Sciences & Humanities Resources

- supporting research in the social science and humanities
- model of collaboration among > 10 RIs in SSH to enhance basis for comparative research

Automation of FAIRification

- Interoperable metadata can be generated through
  - human curation
  - automated metadata extraction (due to AI the proportion will increase)
    - language identification
    - named entity recognition
    - stylometry (authorship detection, period classification)
    - …
- FAIR assessment tools (FATs) can help enhance the level of FAIRness
- FAIR metadata can be offered to machines to generate SKGs (outcome of the SSHOC pilot in HE project OSTrails)

Diversity and FAIR metrics

- FAIR assessment tools and FAIR metrics can differentiate between types of resources. No need for a one-size-fits-all approach.
- Reporting the FAIRness of datasets can be coupled to guidance how FAIRness can be improved.
- More thematic communities can benefit from FATs if the tools meet the requirements of the community and their place in the diversity spectrum.
- Well designed DMPs can support projects in anticipating and monitoring compliance /compatibility with FAIR standards.

Diversity and the Power of Federation

- SSHOC is a highly distributed cluster and the interoperability of the services is dependent on a strong focus on FAIRness
- The SSHOC cluster is highly diverse, but repository operators are ready for adopting tools for the automating of FAIRification as long as community-specific implementation profiles are offered and the various frameworks to register FAIR resources, e.g. FAIR Digital Objects, are taken into account.
- SSHOC maintains a wealth of highly diverse legacy data that can be used for AI training and evaluation, which in turn may lead to enhanced support for SSH.
- The inherent diversity of data types in SSH both underlines *and* enhances the added value of service federation.