

IT WG Call 2018-02-14

CLDI IT WG Conference Call

February 14, 2018

3pm EDT / noon PDT

AGENDA

1. Introductions & Welcome (temporarily at least) to Dan Scott!
2. Plans for continuing to integrate with other working groups
3. Status of technology sandbox at WestGrid
4. Ideas for data sets we want to play with, research projects we want to undertake
 1. Bilal - best practices for integrating distinct data sets
 2. Request from Digital Projects for URI minting - how would this work? Anyone want to take this on?
 3. Other ideas?

NOTES

Present: Jenn, Bilal, Peter, Paul, Russell (Ottawa - Canadiana), Gagandeep (McGill), Mutugi (McGill), Dan Scott

- Dan Scott is on the call - joining for a couple months - would like to help bring momentum - worked on schema.org bib extensions - did proof-of-concept client-side sparql fetch of Wikidata info on bands/musicians, display card

2. Integration with other groups

- See Calendar: <https://connect.library.utoronto.ca/display/U5LD/calendars>
- continue the liaison process with other groups: Mutugi: Education; me: Digital Projects; Bilal: Metadata - other groups are not very active

3. Sandbox - not ready yet - will document the process - N.B. on basic Canada Computes infrastructure there will be a 100,000 statement limit

4. Datasets

- thesis dataset - want to show ability to add value via linking
 - Jenn: metadata unit has idea for a project (student newspaper full text, named entity extraction), has asked for help
 - Dan: small project, if can get student help: grey literature collection on mining reclamation - will be doing metadata from scratch - c.140 reports
 - Bilal: serendipitous browse environment built around Banting and Best, migrated from MySQL db
 - Paul:
 - coop student did metadata extraction proof-of-concept with focus on geo (working in OpenRefine)
 - head tax db - c.100k entries of Chinese immigrants - lots of data re place of birth etc. - would like to develop ontology for immigration - also interested in platform for developing and maintaining an ontology - would be cool to link the head tax data to the [Real Face of White Australia](#) transcription project - Bilal mentioned VitroLib for supporting ontologies - Dan suggested WikiBase instance
 - classic conversion of MARC dataset to linked data - pick up on Ian Bigelow's BIBFRAME work
 - problem of displaying this kind of data on the web (will be a component of Bilal's project)
- URI minting - requested by Digital Projects group - need to work out what their needs are: resolvable? content negotiation? etc.