Apps and Tools Call 20180524

Date	2018-05-24
Topic	OpenHarvester

Call Info

1:00 PM US eastern Time

Join from PC, Mac, Linux, iOS or Android: https://duraspace.zoom.us/j/952326581

Or iPhone one-tap :
    US: +16468769923,,952326581# or +16699006833,,952326581#
Or Telephone:
    Dial(for higher quality, dial a number based on your current location):
        US: +1 646 876 9923 or +1 669 900 6833 or +1 408 638 0968
    Meeting ID: 952 326 581
    International numbers available: https://duraspace.zoom.us/zoomconference?m=UwwKqz4RbGAsBAZgCE9XMorMuL0CeV4Q

Attendees

- Muhammad Javed
- Christian Hauschke
- Ralph O'Flinn
- DJ Lee
- Don Elsborg
- Pierre Roberge
- Rachid Belkouch
- Steve McCauley
- Mike Conlon

Agenda

1. Muhammad Javed will demonstrate OpenHarvester, a tool for importing publication data from open APIs.

   Institutions that do not have subscription to large abstract and citation indexing sources, it is challenging for them to manage their publications and building faculty profiles from open data (such as CVs, websites etc.). There are a number of open APIs available to search, harvest and download citation metadata. Few of these citation APIs are CrossRef, PubMed and DBLP APIs.

   Javed from Cornell University will present OpenHarvester - an interactive tool that process result sets, harvested using above mentioned open APIs, and uses a simple algorithm that refines the result set using a recursive approach.

   This is a preliminary work. The prototype works in two (separate) steps i.e., first downloading potential publications for a person from a database and second processing the result set and claiming the precise publications. Claimed publications can then be saved in RDF and pushed to a VIVO instance.

Notes

See Slides here: OpenHarvester.pdf

Java prototype for claiming publications

Inspired by Elements, high overlap with ReCiter (Weill)

Finding publications for researchers

Need: Elements does not search in CrossRef and EPubMed Central dbs.

Limitations: Works from APIs to download, then processes downloaded data.

Creates VIVO triples from claimed publications

Sources: CrossRef, DBLP, Pubmed, (possibly in future ..SCOPUS, Clarivate)

Fetch and process pubs, use a claiming case

Perhaps 100 hours of work

Claiming interface might have three buttons

      Accept

      Reject

      Ignore

Persistent storage important to remember what user told you

Harvard has an open API for discovering publications in PubMed. Uses co-author patterns to improve ranking of pubs