Scholars Portal Dataverse Contacts Meeting

Meeting Notes, 2020-01-23

Attendees (please enter your name and institution):
Jasmine Hoover and Amanda McNeil (Cape Breton University), Chris Hurst, Brandon University, Ahmed Mahdy (Dalhousie), Erin MacPherson (Dalhousie), Alex Cooper (Queen's), Laure Perrier (UToronto), Michael Steeleworthy (Laurier), Isabelle Lorrain (Université de Sherbrooke), Irina Gherasim (Polytechnique), Shahira Khair (UVIC), Alison Farrell (MUN), Ève Paquette-Bigras (Université de Montréal), Moira Davidson (Lakehead), Mathieu Ayoub (Université du Québec à Trois-Rivières), Marie Stewart (École de technologie supérieure), Kathryn Ruddock (UCalgary), Sandra Keys (Waterloo), Martine Gagnon (Université Laval), Chris Shoniker (Carleton), Elizabeth Kreiter (Trinity Western), Sacha Jerabek (UQAM), Louis Michaud (UQAR), Heather Whipple (Brock), Yayo Umetsubo (UofT Mississauga), Jonathan Dorey (INRS ENAP TELUQ), Kathy Szigeti (Waterloo)

31 Participants

Agenda

- Housekeeping;
  - Slides and agenda/notes on DV contacts resources page: http://bit.ly/2lZfWaP
  - Notetaker: Shahira
- Announcements
- CANARIE project updates
- Other updates
- Community Q&A

Notes

Announcements
- Agreements signed with all 4 regional library consortia
- 51 subscribers to SP Dataverse!
- Updates to demo DV - new member will be added on ongoing basis. If you want to be added as an admin to you demo DV, contact SP

Canarie Updates
- Working on improvements to Data Curation Tool
- Working on large file upload/download
  - Large file capture module
  - Integrating with Globus Connect has been successful
- Will also be presenting at CANARIE RDM Summit in Feb

Data Capture Module
- Large file upload utility
- First create dataset in dataverse, below data upload will be a data capture module section where you upload to rsync
  - Requires downloading bash script that you use to upload your files
  - Can be more confusing for those unfamiliar with command line and running shell scripts
  - In script must specify folder path, the script then lists out folder contents, you then confirm before proceeding with file transfer
  - Zips folder to upload
- Issue with Data Capture Module
  - Technical expertise required
  - Bash scripts and Windows issues (especially pre windows 10)
  - Due to security issues would only be able to allow certain IP ranges to use
  - No file tree view
  - Can only upload one file, not even with versioning
- Conclusion from SP is that unlikely to move to production

Globus
- Proof of concept integration is Phase 1 due spring 2020
  - To use, first need to create a dataset and save it.
  - On file upload tab, below usual Upload space will appear a globus upload space
    - Before using Globus need your own globus account and a globus endpoint installed on your end (pc or server)
  - Globus Connect will connect local end point to destination end point, and can run in the background
  - Transfer status is monitored within Globus Connect
  - When transfer finished, need to click in Dataverse file upload page that transfer has finished
  - Supports versioning and adding additional files
- Phase 2 would include SP-designed UI
- SP will make sure dev work will be brought into dataverse core code
- Key Considerations
  - Need to internationalize Globus Connect UI
  - Dataverse permissions cannot be applied at the file level with Globus
  - Syncing and checking files not automated, yet
  - Getting users comfortable with Globus
  - Need to negotiate a Globus license
- Questions/Comments:
  - how do file hierarchies work? Are they changed in DV or Globus?
    - More testing to be done
  - FRDR working group is making screencasts on Globus, which could be repurposed for SP

Sharing Dataverse Resources
- Spotdocs space created for this purpose
- All members need to submit a quick form to be added to the group to edit/access page (https://forms.office.com/Pages/ResponsePage.aspx?id=O8se1rE41UKCxO-yg4uSXNU-VMKkwzxBu3klh0tr0lBUNklHTDJVSlBNT1I0UkdGWEhYOUpBVDYyQi4u)

Upcoming Webinars (open to non-OCUL group members too)
- Needs assessment on RDM infrastructure for OCUL institutions - Jan 27
  - https://ocul.zoom.us/meeting/register/vJMvd-2hrDgv-9_qGs20R0ZW7m3zkj_vcg
- Data Curation Tool - Feb 14
  - announcement and connection details to come

New documentation created for login workflows

DOIs being minted through transition in Datacite Canada

Roadmap for SP Dataverse
- Plan to increase number of upgrades SP does per year. Issue with current approach of 1-2 is all the new versions that need to be updated at once.
- Hearing from community better admin interface is needed, will be working on updating admin interface
- Working on upgrade to cloud storage
- Large file transfer capability
- Other suggested projects from community is capability of double blind peer review of datasets, and enhancements to disciplinary metadata

Next meeting Feb 27 @ 1PM ET

Community Q/A

When is the deadline for usability test?
- Usability test participation is targeting new users (University of Toronto faculty and students who have not used Dataverse) sign up deadline late Jan?

Do people from OCUL schools need to sign up using form to access spotdocs?
- Yes!

How are institutions promoting Dataverse
- Targeting faculty council meetings, in discussions about TA policy, promoting Dataverse as a response to pillars
- (Chat) - From Ève - Université de Montréal to Everyone: Promoting Dataverse through liaison librarians.Also offering full user support.

- uOttawa OVPR mandated library to promote Dataverse as the repository of choice, and do outreach and teaching. Dataverse will have larger research infrastructure support around it. How are people dealing with teaching DV alongside other repositories?
  - Staffing/capacity challenges - if we promote it too much, can be a lot more work on our end, but not everyone guarantees the quality (no institutional directive to do so)
- Portage document "Repository options in Canada: A Portage Guide": https://portagenetwork.ca/wp-content/uploads/2019/06/Repository-Options-Guide_EN.pdf
- Shahira working on minimum standards for deposit document (e.g. file formats and other things that make curation process more clear) - will share
  - Erin noted Portage DV North metadata recommendations
  - Would like to see policies for user deposit, collection development
  - (Chat) From msteeleworthy to Everyone: We have a "selection criteria for inclusion" policy doc in the works, focusing on min standards for deposit (copyright, scientific value, open, etc, uniqueness, etc). but the important nitty-gritty stuff is going to be left to an operational document... hoping to one day have it in a shareable state!
  - More concrete guidelines forthcoming
- Queen's hosting train the trainer events for library staff to demystify platforms and processes around RDM
  - Slides will be shared, also hoping to see pedagogical models
- Chat: Is there material on good readme files?
  - Cornell: https://data.research.cornell.edu/content/readme

For questions like how many files can Dataverse handle? Would be useful to have guide to turn to rather than needing to email SP listserv.