

Borealis Institutional Contacts Meeting

Meeting Notes, 2022-09-29

Zoom link:

<https://engagedri-ca.zoom.us/meeting/register/tZwud-yvrTMsH9ZZK815QWLyAIA-NyT60-dld>

Attendees (please enter your name and institution):

Alex Cooper (Queen's), Lacey Cain (Carleton); Jane Fry (Carleton); Prasun Lala (ÉTS), Eugene Barsky (UBC); Alisa Rod (McGill); Isabelle Lorrain (UdeSherbrooke), Chris Hurst (Brandon), Irina Gherasim (Polytechnique Montréal), Sarah Giesbrecht (UWinnipeg), Krista Alexander (Concordia University Montreal); Ève P.-Bigras (Uni de ,Montréal), Dana McFarland (VIU), Sacha Jerabek (UQAM), Michael Steeleworthy (Laurier), Sandra Keys (Waterloo), Dan Scott (Laurentian) Kathy Szigeti (Waterloo), Martin Regimbald (UQO)

Hosts: Meghan Goodchild (SP/Queen's), Jonathan Dorey (INRS/SP), Lina Harper (Alliance), John Huck (UofA), Amber Leahey (SP),

Number of participants on call: 36 participants, 4 community facilitators

Agenda & Notes

1. Housekeeping

- Notetaker: Dan Scott, Lina, others (name yourself! :))
- [Slides/Diapos](#) and/et [notes](#); DV contacts resources [page](#)

2. Overview of Borealis

- New branding for what was Scholars Portal Dataverse. Bilingual, multidisciplinary, secure research data repository with 60+ academic institutions
- Supported by librarians, developers, and systems people at Scholars Portal, as well as community facilitators
 - i. Co-chairs: Amber Leahey (Service Director, Borealis) & John Huck (DVN chair)
 - ii. Meaghan Goodchild (Queen's / Scholars Portal) and Jonathan Dorey (INRS & Scholars Portal)
 - iii. Lina Harper (Alliance)
- Community meetings & support mailing list (DATAVERSE-SUPPORT-L@listserv.utoronto.ca)

3. Results from the Community Meeting Survey

- 27 respondents from across Canada
 - i. 7.4% CAAL / CBPA, 11.1% COPPUL, 40.7% OCUL, 41% BCI
- 85% said the current time (last Thursday of the month at 1PM Eastern) was convenient
- 89% attend monthly or every second month
- Majority found the community meetings to be beneficial (12 %, 12 5/5, 2 3/5)
- Most useful parts of meetings (in descending order):
 - i. Platform updates
 - ii. General discussion & Q&A
 - iii. Community questions
 - iv. Did you know / what's new
 - v. Invited presentations
 - vi. Alliance updates
 - vii. Dataverse North updates
 - viii. Other community updates
- Very positive responses about the efforts to make the meetings more bilingual (from both anglophones and francophones), with support for dual approach of bilingual updates and breakout rooms—particularly for asking detailed technical questions in French
- Interest in special topics:
 - i. Sensitive data
 - ii. APIs
 - iii. Metadata
 - iv. Learning from other institutions (curation, metadata management, service models, outreach, training, integrating with REBs)
- Special topics will need to be more in-depth and thus scheduled outside regular community meeting times

4. Borealis Service Updates

- Welcome to Thompson Rivers University (COPPUL), Aurora College (COPPUL), and Durham College (CLO/OCLS)
- Overview of Borealis service
- Ongoing and new projects update
 - i. [New user guide](#): more in-depth coverage of features, step-by-step instructions with updated screenshots
 - ii. CoreTrustSeal Project
 - [Borealis CTS Application Template](#) for 2020-2022 requirements is completed
 - Plan to develop version 2 for 2023-2025 requirements
 - [Dataverse North Policy Templates](#) - translated by Alliance and updated to reflect new name of the service
 - iii. [Admin Guide](#)

- Draft available to support the CTS cohort; currently being translated and feedback will be requested soon
- iv. Sensitive data
 - Developing guide on sensitive data sharing (current features) to be share for community feedback this fall
 - Hope to participate in Alliance FRDR sensitive data project
- v. [Globus](#) integration
 - Experimental (minimum viable product) version building on early work by Scholars Portal / Borealis team is available in 5.11
 - Borealis team is currently testing this
- vi. Software upgrade (late fall 2022)
 - Will offer “click to approve” licensing of datasets
- Community questions
 - i. Institutional storage allocation is based on [CRKN banding/size](#):
 - Small - 1TB
 - Medium - 4TB
 - Large - 8TB
 - Extra-large - 10TB
 - More storage can be purchased at the OLRC rate of \$276/TB/year
 - ii. You can ask Borealis to tell you how much storage you’re using, or calculate your storage through a report and total the file size
 - iii. Borealis have not yet hit a spot where any institution needs more storage. We are currently using 6TB in total.
- What’s new / did you know?
 - i. Large file support
 - 3GB file size limit - same for both user interface and APIs (including DVUploader)
 - Even if you manage to upload a file larger than 3GB, your users will only be able to download a maximum of 5GB
 - Zip files can contain up to 1000 files; limit is still 3GB but it currently does not support uploads of files with filenames containing accented characters
 - Benefit of [DVUploader](#) is that it allows multiple file/bulk uploads, maintaining structure without needing to zip files
 - See also the [Borealis User Guide: ‘Files’](#)
 - Downloading large files / volumes of files
 - wget, which downloads all of the files at once without preserving the folder structure

- Or the Borealis team's [python script](#) which downloads the whole dataset with the hierarchy intact (tested up to 35 GB).
- ii. Eugene Barsky has experimented with a [Harvard-developed tool to upload items from a Github repository](#) (useful for researchers who have stored their data in Github);
- iii. also, UBC's updated their [File_Naming_Guidelines.pdf](#)

5. Dataverse Community Updates

- Dataverse North (DVN), chair, John Huck
 - i. two groups have wrapped up their work (Metadata and Training WGs)
 - ii. Still interested in training activities, open to restarting either group in the future
 - iii. Dataverse 101, Module 2 will be posted soon [here](#) (translated!)
 - iv. Coordination of potential training activities is being discussed among Borealis, CEG (Alliance) and NTEG (Alliance)
 - v. [Survey of Canadian Dataverse Administrators](#) working group
 - Goals are to develop an understanding of who the Dataverse administrators are in Canada, what kind of services we support, and the challenges we face
 - Timeline: working group to begin in October, data collection in early 2023

6. Alliance updates

- Continuing series of sessions to develop an institutional RDM strategy
- Next one is October 21st between noon and 1:00 PM Eastern: "Panel: Working Through the Planning Stage of Developing an Institutional RDM Strategy"
- Resource Allocation Competition is open with applications due November 2, 2022
 - i. Priority access to researchers for storage, compute and cloud resources beyond what can be obtained via the Rapid Access Service
 - ii. Q&A sessions will be held [October 4 \(English\)](#) at 12:00 PM Eastern and [October 5 \(French\)](#) and at 12:00 PM Eastern
 - iii. Please ensure your research office knows about this and has reached out to researchers!

7. Discussion / Q&A

- Next meeting on October 27th conflicts with the Scholars Portal Update Webinar; however, a recording will likely be available from Scholars Portal
- Tabular ingest guidance - researchers have run into problems with some Excel and CSV files and it generates an error message that concerns researchers. Excel in particular is problematic as some researchers use multiple sheets within a single file, or use it "creatively"

not just for rows & columns of data. See [the Tabular Ingest section of the Files guide](#) for best practices. Some Dataverse sites have disabled tabular ingest as a result, but it seems like a user experience problem. Being able to disable it for selected files would help, or even just editing the error message so it's less dramatic would be helpful.

- i. The Borealis team and Jonathan Dorey said he would look into the problem in general.
- Can/should "Guestbook" be implemented/activated by the admin, rather than leaving it as is for the data depositor to use?
 - i. "As a user, I would assume I can get my own guestbooks"
 - 33,000 rows of guestbook entries need to be sent to researcher
 - ii. Could improve the messaging, adding additional Admin controls
- Question: There have been a few moments this year where it would have been nice to share a troubleshooting question in a more searchable format than emails. This would also allow everyone to read the answers that have been responded to Dataverse administrators off-list. Is there any chance we could set up something for the Borealis community like the PID Forum? <https://pidforum.org/>
- Was there resolution to the recent issue/concern of Contributor Plus users being able to add admins?