

2021-10-19 Dataverse Community Call

[Dataverse Community Call 1 \(2 AM UTC\)](#)

[Dataverse Community Call 1 Agenda/Notes](#)

[Dataverse Community Call 1 Attendees](#)

[Dataverse Community Call 2 \(3 PM UTC\)](#)

[Dataverse Community Call 2 Agenda/Notes](#)

[Dataverse Community Call 2 Attendees](#)

Dataverse Community Call 1 (2 AM UTC)

Dataverse Community Call 1 Agenda/Notes

- IQSS Updates
 - 5.7 released - <https://github.com/IQSS/dataverse/releases/tag/v5.7>
 - (Steve) Does this include the ARDC integration?
 - (Jim) Yes, in the GDCC repo. The 5.7 release contains the integration from the Dataverse side.
- GDCC Updates
 - Embargo in QA, MultiLicense in review
 - (Steve) Any idea what release these will go out in?
 - (Danny) I said as soon as Embargo was in, I'd cut a release because it's a long desired feature and DANS paid for this work so it would be good to get it out
 - (Jim) It sure would be nice to get Multiple Licenses in the same release but Embargo is definitely closer, but it would be good to group those.
 - (Steve) Yes, they do go well together
 - (Danny) We'll see where we are once Embargo gets merged in, but expect both of these soon
 - Previewers: newer versions and post v.5.2 instructions in the develop branch. Back to planning merge into master and migration to gdcc repo.
 - (Jim) We'll give a couple weeks notice on this
- Community Questions
 - (Eunice) Any suggestions on approach for dealing with collection of bots hitting backend (e.g. rate limiting, throttling)? [on behalf of Yuyun, NTU]
 - (Danny) We've done some IP blocking of scripts, and we are investigating other options for rate limiting (near the top of our list), first pass will be using the AWS suite and we'll also a
 - (Eunice) For purpose of preventing DDoS, what are the options (server or application level) of disabling /api/access/datafile/{id}, without parameter referrer?

- Follow-up to previous question: Are there any downsides of disabling /api/access/datafiles/{id} without parameter referrer? (e.g. pyDataverse python scripts, wget would not work)
- (Jim) I've talked with Don at Odum about this and it's possible with Apache, the main downside would be that file downloads themselves go through this API. If it's blocked it would block UI downloads as well.
- (Danny) Is this the case if redirect is enabled?
- (Jamie) Any update on TRSA?
 - (Danny) We're starting on the Harvard Data Commons work which will include integration with Globus and will allow us to devote some resources to files stored outside of Dataverse storage.
 - (Jim) The overlay work is out there - now with momentum on the Harvard side and general interest from UNC we may get it merged

Dataverse Community Call 1 Attendees

- Danny Brooke (IQSS)
- Jamie Jamison (UCLA)
- Eunice Soh (NTU)
- Janet McDougall (ADA)
- Jim Myers (GDCC, QDR)
- Mingjing Peng (ADA)
- Steve McEachern (ADA)

Dataverse Community Call 2 (3 PM UTC)

Dataverse Community Call 2 Agenda/Notes

- IQSS Updates
 - 5.7 released - <https://github.com/IQSS/dataverse/releases/tag/v5.7>
 - (Philipp) Are you going to enable controlled vocabularies on the demo instance?
 - (Danny) I will look into it. Not sure if we want to do it on demo, but it would be good to provide some place the community can test
- GDCC Updates
 - Embargo in QA, MultiLicense in review
 - Previewers: newer versions and post v.5.2 instructions in the develop branch. Back to planning merge into master and migration to gdcc repo. We'll give a couple weeks notice on this
- Community Questions

- (Phil) More of a comment, we might see an update to <https://github.com/IQSS/dataverse-ojs-plugin> because Sjaeful from LIPI reached out to me.
 - The current plugin does not work with OJS 3
 - Possible testing opportunity coming up! No code on GitHub yet.
 - (Philipp) We are interested in this, and this is included in a work package for a funding proposal. No capacity to test right now, but maybe in the new year
 - (Phil) Great - I thought of you and Obi
- (Phil) Does anyone have experience monitoring RDS ("database as a service" from AWS)? I'm working on <https://github.com/IQSS/dataverse.harvard.edu/issues/126>
 - (Phil) Slow queries, Cloudwatch are of interest
 - (Phil) Please reach out with any questions!
 - (Philipp) We are moving to AzureCloud - we are trying to make the decision to run it as a service or on a VM. Would be interested in hearing more about this. Are you evaluating a move or do you already use this?
 - (Phil) We use it now, for several years. We are investigating this tooling for help diagnosing performance issues
 - (Deirdre) We are investigating this
 - (Phil) We'll get in touch!
- (Laura) Activity statistics in Harvard Dataverse (<https://dataverse.harvard.edu/>) where are these pulled from? Display?
 - https://github.com/IQSS/dataverse.harvard.edu/blob/master/customization/Harvard_Dataverse_Customization.README.txt
 - (Jim) https://data.qdr.syr.edu/metrics_5ef2ae2be4b/ - using <https://github.com/IQSS/dataverse-metrics/pull/52>
 - <https://github.com/IQSS/dataverse-client-javascript>
 - (Phil) You can find other tools at <https://guides.dataverse.org/en/5.7/admin/reporting-tools-and-queries.html>
- (Philipp) Do the metrics tools use APIs? Could they pull the same information as DB Queries?
 - Yes, we've tried to cover as many use cases in the APIs as possible
- (Kris) I noticed in Call 1 there was a discussion about Bots, DDos. Would it be good to get some knowledge base on this?
 - (Danny) I'm in favor of this! Rate limiting is next up for the IQSS team on this
 - (Phil) Securing your installation could be a spot for this. What do others think? Add "preventing DDOS" to <https://guides.dataverse.org/en/5.7/installation/config.html#securing-your-installation> ?

- (Philipp) This would be useful! We are running a risk analysis right now. This will be an important part of a risk analysis.
- (Danny) I will create an issue.
- (Philipp) There's been a discussion in the Google Group about editing user accounts. When we migrate DataverseNO to the cloud we are planning to move from Shibboleth to OAuth. We manage the current solution through DB triggers to get access, based on the email address. If we move to OAuth how will we do this?
 - (Phil) You said you were moving away from SAML/Shib but this is what we use in the Harvard Dataverse for licensed or other privileged content.
 - (Jim) General opinion is that you shouldn't allow people to change their info especially in cases where it would update permissions.
 - (Gustavo) We need to continue work on email verification.
 - (Philipp) How much would it take to change it?
 - (Phil) A week or two
 - (Phil) Grouper is one tool that's used to manage groups (used at Harvard), may be an option here. More research needed.
 - (Don) I associate Grouper with LDAP, but I'm glad to hear you say it's used with SAML as well.
 - (Phil) If we were to ask your identity people how they know which people are from which institution, would you say email or something else?
 - (Philipp) There are a lot of attributes available, both in the current system and the new system
 - (Kris) We are piloting, and using Shibboleth. We run a nightly process based on ORCID to provision access to specific groups.
 - (Phil) I'll dig up the old issue where we discussed this. Shibboleth Groups: arbitrary attributes and regex support #1515 - <https://github.com/IQSS/dataverse/issues/1515>
- (Philipp) Regarding the different protocols - is there any reason to use one over the other?
 - (Danny) I think our plan from the project perspective is to support as many as reasonably possible.
 - (Phil) We made the pick of SAML over CAS and we've been generally pleased. But we may switch in the future!
- (Philipp) We have a group at the university which requires a video thesis and wants to make them available. Is there anyone in the Dataverse Community that's used a front end?
 - (Danny) We have use cases around images being hosted in Dataverse and being used by other systems, but not videos as far as I know
 - (Laura) Have you checked out <https://www.avalonmediasystem.org/project> ?
 - (Kris) IIIF may be another option
 - (Laura) We'd be interested in continuing conversations around this!

Dataverse Community Call 2 Attendees

- Danny Brooke (IQSS)
- Gustavo Durand (IQSS)
- Phil Durbin (IQSS)
- Jim Myers (GDCC, QDR)
- Kris Dekeyser (KU Leuven)
- Dieuwertje Bloemen (KU Leuven)
- Laura Waugh (TDR, TXST)
- Philipp Konzett (UiT/DataverseNO)