

Attendees: Peter Portante, Daniel McPherson, Philip Durbin

Proposed agenda (we didn't talk about metadata blocks much):

- Background: Phil and Peter chatted at the 2017 Dataverse Community Meeting
- How Dataverse could help Red Hat create a "custom metadata block" for their data on performance metrics (number of CPUs, etc.).
 - A good starting point may be Johns Hopkins software metadata block at <https://github.com/dilauro/dataverse-ansible-role/blob/master/roles/dataverse/files/dv-mdblocks/jhuSoftware.tsv>
 - Tim DiLauro from Johns Hopkins has documented his process in a doc called "Dataverse 4.x Metadata Blocks syntax/semantics" that's linked from <https://github.com/IQSS/dataverse/issues/3168>
- How Red Hat could help Dataverse get the app running on OpenShift, Red Hat's Platform As A Service offering. The first project is free and might be an easy way for people to kick the tires on Dataverse.
 - Phil was originally saying that a good starting point may be the AWS/Docker work done by Virginia Tech at <https://github.com/IQSS/dataverse-aws> but Dataverse community members in #dataverse on freenode are leaning toward using Docker images built from <https://github.com/nds-org/ndslabs-dataverse> according to http://irclog.iq.harvard.edu/dataverse/2017-06-28#i_53829 . See also discussion at <https://github.com/IQSS/dataverse/issues/3938> .
 - OpenShift would be a much better "kick the tires" story for Dataverse than the Vagrant solution we describe at <http://guides.dataverse.org/en/4.7/installation/prep.html#vagrant-for-testing-only> .
- (Optional) Tools Dataverse could be using on top of Solr that Peter knows about.

Notes

- Since the default license for datasets in Dataverse is CC0, should they be available in <https://search.creativecommons.org> ? Update: It's not clear how to be included in Creative Commons Search.
- OpenShift Dedicated can be run on AWS: <https://www.redhat.com/en/about/press-releases/red-hat-launches-openshift-dedicated-delivering-its-container-application-platform-public-cloud-offering>
- Reference architectures are open access, for example: https://access.redhat.com/documentation/en-us/reference_architectures/2017/html-single/deploying_red_hat_openshift_container_platform_3.4_on_red_hat_openstack_platform_10/
- Interest in deleting intermediary versions of datasets or at least run. Talked about deaccessioning versions and if files should or could be deleted. Could just create many

datasets and only publish some of them. With S3 you can get rid of files older than 90 days.

- Datasets consisting of a tarball with a visualization of that tarball as HTML.

Action Items

- Schedule next meeting in late August to discuss steps toward getting Dataverse running on OpenShift:
https://docs.google.com/document/d/1xvUb98BmVPn4Nr-3i2HgGLgeObs5lz9_WJ8v5u40gGY/edit?usp=sharing
- Phil and other Dataverse developers are welcome to attend Dan's OpenShift for Beginners class.