Data Management and Curation Working Group (DMCWG)¹
Monday, March 25, 2019 11:00 am-12:00 pm, Marston Science Library, room L136

Attended: Plato Smith, Matt Gitzendanner, David Van Kleeck, Ryan Davidson, Dan Maxwell

**Agenda and Notes**

1. UF Data Guide from Integrated Risk Management (thanks Todd and Dan) - [https://irm.ufl.edu/uf-data-guide/](https://irm.ufl.edu/uf-data-guide/)
   1. Briefly reviewed the UF Data Guide “How to Safely Work with My Data at UF” developed by UF Integrated Risk Management. The resource which is currently under construction will develop over time and become iterative. The resource include a list of (1) Data Types, (2) Open/Sensitive Data Solutions, and (3) Restricted Data Solutions. The resource includes UFIT supported solutions for potential data types developed during research. For data types not listed and/or solutions not UFIT approved, then there is contact information to contact the Integrated Risk Management team for support.
   2. It was discussed if the resource would eventually include other solutions not on the list such as third-party solutions currently not UFIT supported.

2. Data Repository
   1. UFL Dryad conversation (contacted by Dryad)
      i. Briefly discussed Dryad’s new model for a consortium membership as a business model to support capacity, development, and infrastructure. Dryad is a digital repository to provide access to data, share data, and store data. See: [https://datadryad.org/](https://datadryad.org/)
      ii. A Dryad use case to store evolution biology data by a UF researcher/scientist a few years was discussed. There were issues with access, discovery, sharing, and storing of multiple semantic and taxonomy rich data sets a few years ago. It would good to see how the features and functionalities have developed.
      iii. It was discussed that invenio ([https://invenio-software.org/](https://invenio-software.org/)), Open Source framework for large-scale digital repositories, and ckan ([https://ckan.org/](https://ckan.org/)), Open Source data portal platform could be candidates for further evaluation by the DMCWG. Invenio is the underlying platform for Zenodo and can view invenio’s showcases via [https://invenio-](https://invenio-)

---

¹ DMCWG resources: [http://ufdc.ufl.edu/AA00014835](http://ufdc.ufl.edu/AA00014835) and Data Resources [http://www.uflib.ufl.edu/datamgmt](http://www.uflib.ufl.edu/datamgmt)
However, rather than develop data repository from in-house, it was briefly suggested to identify existing solutions for comparisons, evaluations, and recommendations that provide consortium/membership service/support.

iv. It was proposed that the DMCWG have a data repository solution by end of 2019. This will require input from the DMCWG, RCAC, and others.

2. More training requests from funded projects for sharing data outputs (UF STRIDE and UF/IFAS NCBS)
   
i. Briefly discussed the recent request for a Zenodo data deposit training video by the data coordinator for STRIDE (https://stride.ce.ufl.edu/). The data coordinator was trained on how to develop a community within Zenodo and is the admin responsible for curation of the data for the community which is linked from the STRIDE’s Deliverables & Reporting page (https://stride.ce.ufl.edu/research-2/deliverables/) along with the DMP in which the DMCWG assisted with development. The first draft of the Zenodo training video was submitted and reviewed by the research coordinator last week. Feedback and recommendations for improvements are pending. Once the final draft is completed and vetted by the STRIDE research coordinator, then that video will be adapted, adopted, and modeled for a general training video for a data repository tab on the research data management libguide (http://guides.uflib.ufl.edu/datamanagement/home).

   ii. The data coordinator for UF/IFAS Nature Coast Biological Station (NCBS) (https://ncbs.ifas.ufl.edu/) has requested a data repository training similar to STRIDE. A meeting has been scheduled at NCBS for 4/4/19. The Zenodo training video developed for STRIDE will be modified and developed for UF/IFAS NCBS and possibly the FACETS project as well.

3. GitHub for Education (versus Enterprise)
   
1. Proposed several questions before next conversation (contacted by GitHub)
   
i. Briefly discussed the outreach by GitHub to resume previous discussions for a university-wide license. However, this is not a priority for UF Research Computing. Nevertheless, the UF Libraries can gather information to contribute to future discussion with UF Research Computing support. The Associate Dean for UF HSCL has reviewed and approved the following proposed questions for future GitHub discussion.

   1. What are the GitHub R1 use cases?
      a. What are the University-wide implementations?
      b. What security protocols, standards, and IT compliances were involved in the University-wide implementations?
2. What are the GitHub options based on current UF implementations/uses (e.g. University-level, college, school, unit, lab levels?
   a. What is the current GitHub usage at UF?
   b. What are the current costs associated with GitHub usage at UF?

3. What are the costs for Enterprise GitHub (unlimited repos) for UF and UF-affiliates?

4. Can current repos be migrated to a University-wide Enterprise GitHub license? If so, how? If not, why?

4. NSF CAREER workshop (4/8/19 – 4/9/19)
   1. Applied and was selected to attend workshop
      https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=503214
   2. Briefly discussed how the NSF CAREER (NSF 17-537) proposal could potentially address (1) funded research projects need for data management/data science capacity, (2) compliment and support NSF CyberTraining (NSF 19-524) proposal, and (3) leverage informatics and R users groups (graduates/doctoral) and Data Science & Informatics (DSI) (undergraduates) to develop a short-term project framework project sprints) whereby students get paid, experiential learning, and course credit for solving and working on projects for funded projects.

5. SCORS meeting (3/26/19)
   1. Five minutes for DMCWG updates
      i. Briefly discussed how the DMCWG is addressing (1) funded projects data management support, (2) data repositories, and (3) training.

6. Miscellaneous
   1. ORCID API target for June 2019 and is being led by UF Libraries IT Chair
   2. The introductory data management course from spring 2019 will be evaluated and enhanced during summer 2019 for future spring 2020 consideration.
   3. A DMCG rep is attending the iConference 2019 @ UMD from 3/31/19 – 4/3/19.