

Joint Meeting
Senate Computing and Information Technology Committee
Senate Research Committee

(online)

26 October 2020

10:30 AM

In Attendance:

Research Committee: A. Crunk, J. Das, D. DeAlmeida, M. Hravnak, M. McCall, P. Morel, M. Oberbarnscheidt, E. Oyler, R. Rutenbar, S. Sant, P. Smolinski, V. So, A. Stephany, W. Yates,

Computing and Information Technology Research: Michael Spring, Adam Hobaugh, Stacie Gabriel, Gosia Fort, Ralph Roskies Marty Levine, Kenny Doty, Marty Levine, Alexandros Labrinidis, Irene Frieze, Arif Abdullah Jamal, Chris Bonneau, Chad Burton, Kimberly Barlow, Dmitriy Babichenko, Stephen Wisniewski

Data management Discussion

M. Spring opened the joint meeting with an introduction to the issue of research data management. He pointed out that there are a number of important issues to consider such as access to data storage, accessibility of stored data and privacy concerns among others. He invited R. Rutenbar to provide an overview of what Pitt is planning in this regard.

R. Rutenbar made a short presentation about the ideas that his office in conjunction with the Office of Research Computing (Ralph Roskies) and IT (Adam Hobaugh) are developing. We need to separate the fact that box may go away and find better, more varied ways for people to store information. He introduced Laurel Gift who is working on issues related to privacy of data and the issues such as GDPR that may be involved.

The discussions have started and will encompass a set of principles known as the Fair Standard. These ensure that any research data to be stored is: 1) findable; 2) accessible; 3) interoperable; and 4) reusable. It is likely that the libraries will be part of this endeavor as they are already moving towards storing many of their collections in a digital format. For example, the Health Sciences library is almost entirely digital at this point and they can provide expertise in the area of data management.

In terms of data storage there are considerations concerning how available the data needs to be, the so-called "iceberg model". In this model data can be stored at low cost, but every time the data is accessed a fee has to be paid – this option may be viable for those who do not plan to make frequent use of the data. In addition, the policy landscape from NIH and NSF is changing and some datasets have privacy issues such as medical data.

Rutenbar indicated that it is important to separate the issue of data storage from the issue of data governance. The universe of data management is very complex. The universe ranges from notebooks to vast data sources, to images, or maps, etc. There are very different modes of use. We need to think about the system in layers – technology, governance, policy issues such as privacy.

P. Morel introduced two researchers, Martin Oberbarnscheidt and Jishnu Das, to talk about their data needs.

M. Oberbarnscheidt: My work generates about 50gb of data regularly and we've accumulated over 100 terabytes of data over time. We've ran into issues with cost when trying to figure out the best data storage methods which led us to managing our own on five different servers. We never felt that Pitt was supporting our data storage policies which is what led us to taking care of it ourselves

J. Das: Our work involves not only data storage, but also high-performance computing (CPU intensive and Ram intensive). In addition to the storage access of box, the CRC's resources seem to be overloaded and the access for the research community is relatively slow.

R. Rutenbar responded that it would be lovely if there was more of an institutional model where we can buy in on deals that can include multiple users. We need to build a set of easy to use on ramps for various data storage cases. The office is undertaking benchmarking of other Universities such as University of Illinois, Champaign-Urbana, and the University of Utah which have good systems in this regard.

S. Sant: The immediate question everyone has; I have tons of data on my box drive. If it goes away, what is the plan? What happens to the data? Also, some of our colleagues have tried OneDrive and haven't had good experiences with it.

M. Spring stated that the alternative to Box, One Drive, is very sluggish and that the some of the data need to be HIPAA compliant. In addition, some IRB protocols specifically state that data will be stored in Box. R. Rutenbar noted that the IT is working with individual faculty members on these issues, and that it is possible to remain with Box if needed.

A, Hobaugh, Pitt IT manager, indicated IT has been working with Rob's office for quite some time. We've been working with Microsoft Azure, AWS, and other Microsoft applications and are in the process of generating a cloud roadmap. We want to share this broadly and gather feedback from everyone. There have been discussions around a research portal but it's never been implemented and is something we'll need to continue looking into. This would include a service catalogue. In terms of the Box transition he noted that IT was contacting the heaviest Box users to discuss their needs and how to transition from, or stay with, Box.

M. Spring said that these plans need to take into account two requirements: 1) mandates from funding agencies such as NIH and NSF which often require data management plans; and 2) the desires and needs of the faculty. There is a need to communicate to the faculty concerning the changes in mandates as well as the development of this new web research portal. He indicated that it would be useful to have several short documents that detail: a simplified version of regulatory concerns, making access controls understandable, how to write RDM plans. In addition, related to people who have an investment in box, how people who want to migrate can do so, and how we will manage and support exceptions. The fact of the matter is that there is a lot of anxiety introduced by the change in technology that is on the horizon.

P. Morel asked about the timeline of these proposals. R Rutenbar responded that discussions with cloud storage services are ongoing, and that he encourages users to get involved. The message to faculty is that they should contact IT or CRC with their computational needs and these contacts are welcomed.

M. Oberbarnscheidt asked about the access to Google Drive as he had been told that Pitt faculty had access to this storage service. A. Hobaugh responded that this was indeed the case, but that his office was working on developing support for the transition. It is not clear what the longevity or stability of this arrangement will be.

A. Labrinidis noted that there is a communication problem and that it might be useful to create a list of typical cases that could be useful to faculty. R. Rutenbar acknowledged that communication is not yet ideal. Faculty need to be made aware that they can connect with folks at CRC or IT with their problems and concerns. They are hoping to create a service catalog for data and computational issues on the research portal.

The joint meeting adjourned at 12:00 pm

The research committee convened briefly

The minutes from the 25 September 2020 meeting stand as approved.

Concerning the Covid 19 situation:

R. Rutenbar noted that since Pitt moved to the “guarded posture” he has received enquiries from research faculty concerning the need follow or create research plans. He emphasized that all research faculty need to develop a plan that describes how they will handle the different posture levels and that this has to be approved by Department Chairs and the relevant Dean. He also noted that travel is less restricted and now only requires approval from the Department Chair.

Systemic Racism, Inequity and Justice initiative

P Morel noted that there was now a website that contains a lot of data concerning gender and ethnic diversity across the University. The web site is <https://www.diversity.pitt.edu/social-justice>. She asked that the committee look at this website and be prepared for the next meeting to discuss how we might use this information.

The next Research Committee meeting will be on 20 November at 1:00 pm.

The meeting was adjourned at 12:15 pm.

Minutes submitted by P. Morel, M. Spring, and S. Sant