Report:
Public Access to Research Data at Virginia Tech

Prepared by: Andrea L. Ogier (alop@vt.edu), Virginia Pannabecker (vpannabe@vt.edu), Jonathan Petters (jpetters@vt.edu) on behalf of the Public Access to Research Data Committee.

Presented to the Virginia Tech Commission on Research on November 7, 2019.

Summary

Recent reports and recommendations from AAU/APLU and National Academies of Science charge public universities, especially those with a land grant mission, to provide needed infrastructure and guidance that will enable researchers to more easily share the data supporting their research. In brief, the AAU/APLU Public Access Working Group states: “In light of governmental mandates and the scientific benefits of making data accessible to the public, **universities will need to adopt new institutional policies, procedures, and approaches that actively support and promote research data sharing**, while at the same time ensuring rigor in the research process and the veracity of its intellectual outputs.” Virginia Tech has resources to support public access to research data in the Office of Vice President of Research and Innovation, University Libraries, and Division of Information Technology; however, there are gaps in communication, policy, training, and implementation that make data sharing difficult and confusing for busy researchers. In response to a few of these identified gaps, the Public Access to Data Committee makes the following recommendations:

1. Policy 130015 should be revised to provide more pertinent information to researchers on how and where to obtain support for both restricting access to data and/or publicly sharing data,
2. The University Libraries should work with the Office of Sponsored Programs to create guidance for PIs who are working on grant applications and processes for ensuring that Data Management Plans are updated and followed over the lifetime of the grant,
3. The University Libraries, Division of IT, and Research Compliance Office should work together to create guidance helping researchers know when they can and should publicly share their data, and when they can and should keep the data secure.

Ultimately, we recommend the creation of a standing Data Security and Sharing Committee reporting to the Commission on Research to consider, create, and maintain this guidance to better support researchers at Virginia Tech.

Background

Virginia Tech takes seriously its history as a land-grant public university, emphasizing its role in creating innovative and positive change in the local community, throughout the Commonwealth of Virginia, and across the country and globe. The University's living motto, *Ut Prosim* (That I May Serve) embodies this spirit of service to the world. However, Virginia Tech is also an R1 research institution that supports thousands of researchers spread through several research institutes and nine colleges, and has half a billion dollars in private contracts and federal grants. While providing public access to research data is vitally important to its land-grant mission of using research for the public good it is also a controversial and costly subject for researchers.To support researchers who need to comply with federal mandates on public data the Commission on Research established a Public Access to Data Committee charged with exploring the current landscape of public data and making recommendations to improve support of public data at Virginia Tech. This faculty-led committee includes representation from administrative, teaching, and research units across campus. Our approach is driven by the idea that preserving and providing access to data directly supports our university's stated goals of bridging the divide between research and practice. Further, we emphasize the researcher-centric view that the university has developed (and should continue to develop) services that lessen the burden of the researcher. This document reports on our semester-long exploration of university mechanisms that support public access to data.

Methods & Deliverables

In order to ensure that researchers at Virginia Tech have efficient access to information, policies, and processes that help them know when and how to make their data publicly accessible, the committee explored three areas:

1. Current services and information provided by administrative units on data security and sharing
2. The relevance of current VT Policies on data security and sharing
3. The effectiveness of Data Management Plans submitted by VT researchers as part of awarded federal grant applications

Findings from each of these three explorations are detailed in this report.

Findings

Current Services on Data Security and Sharing

Research services and support can come from a variety of sources: the institutes themselves, colleges or departments, the Office of Research and Innovation, the University Libraries, and IT/Advanced Research Computing. Each of the units named above has their own areas of concern and expertise: many have developed resources for specific subsets of the research population. Bringing together a group of representatives from each unit on campus allowed us to discuss what services and support exist, and what would be helpful to create.

Clearly articulated needs include:

- Custom image sharing services (i.e. a way to create a custom digital library to share images either publicly or with specific collaborators)
- De-identification support for human subjects data or data containing personally identifiable information (PII)
- Making data more discoverable for a particular research community
- Guidance around what types of data can and cannot be made publicly accessible or shared, including data at different steps in a research workflow (raw data vs analyzed data vs aggregate data)
- Methods for sharing data that are not repository-based (i.e. emailing data to another collaborator)
- Assistance with creating data use agreements for sharing datasets that should be shared but cannot be made publicly accessible
- Guidance on how and when to use a secure research environment to work with and provide access to healthcare data and electronic medical records

Committee discussion centered on kinds of data that can and cannot be made public. Health care and medical researchers, for example, cannot make public any data that can be linked back to a specific person or easily identifiable group without their consent. However, there may be aggregate data (that if properly de-identified) could be shared or made public. Institutional Review Boards (IRBs) can provide some guidance for human subjects data, though there was some question about whether human subjects data that is IRB-exempt could be made public. The cost of providing public access to data was also a point of discussion, though given the prevalence of both institutional repositories and “free” repositories like figshare, the concerning cost was in labor and time rather than in monetary resources.

One identified challenge is the diverse nature of the datasets, another is that the landscape of policy and funding agency mandates is continually changing. The Common Rule, for instance, now requires that IRB approved consent forms be posted publicly. How should these forms be

posted and for how long? Should they be preserved alongside the peer-reviewed article and the dataset in an institutional repository? The committee agreed that it is too much to ask researchers to stay up-to-date on these policies and practices; instead, this is an area where the research institution should step in and provide current, just-in-time guidance and support.

While the committee was able to identify a number of research-related, researcher-centered services provided by the University, only the repository services and guidance provided by the Library actively supported public access to research and research data. Services and guidance provided by both Research Compliance and the security office within the Division of IT are concerned with keeping data safe and secure, while the Library's guidance covers how and why data should be made available.

⇒ In the end, the committee recommends that the Library, IT, and Research Compliance work together on creating guidance to help researchers know when data should or could be shared, and when it must remain secure.

Sharing, Security, and VT Policies

For the second deliverable, the committee reviewed relevant policies governing research data at Virginia Tech to determine whether they provided enough guidance to help researchers know when and how to share their data. The committee was able to identify six official VT policies that have some bearing on this topic (13000, 13015, 7010, 1060, 7100, and 7000), though only 13000 and 13015 address intellectual property and research data.

The Policy on Intellectual Property (Policy 13000, revised 2015) is designed to establish "ownership criteria" and resolve questions surrounding ownership of intellectual property, and applies to "all employees, students, and all other persons or entities using University resources." Although this policy does not clearly consider research data to be a form of Intellectual Property, it does explicitly cover all other research products (research papers, books, software, inventions, articles, etc.). While this policy does explicitly state that "many IPs are best disseminated by publication and placed in the public domain" it also asserts that "a significant number" should instead be protected by IP law "with attendant financial considerations." The policy further distinguishes between "traditional results" which the author owns, and "novel results" of research which are created as a condition of employment, and which hold some significant benefit (i.e. monetization or technology transfer) to the University. In the latter case, the University asserts ownership. Interestingly, sections 2.3.A.3-4 cover sponsor rights, which are interpreted as private sector sponsors (not federal agencies) and federal agency rights, which are interpreted as statutory IP rights to patents. Neither of these specific cases, nor any other in the document, clearly addresses the federal requirements for public access to research results which may be considered to be IP by the University and fall under the purview federal mandates. The question of whether research or data funded by a federal agency, generated by

a private university, and considered to be significant intellectual property is subject to a FOIA request was asked and remains unanswered. Policy 13000 states that the federal agencies have a statutory IP rights to any patents generated, but does not clearly provide guidance for research data or related materials. We also note that intellectual property rights that apply to data may depend on several factors, such as: location of creation, and different types, levels, forms, and uses of data, as briefly described in this 2015 PLOS Biology perspective article.

Differently from Policy 13000, Policy 13015, Ownership and Control of Research Results (revised 2015), does explicitly apply to research data and asserts that the university has ownership of "research results and material (this includes all data)" generated with university resources. However, this policy does not differentiate between research materials that are wholly funded by the university (through researcher salary and facilities space) and those that are partially funded by federal agencies through direct grants and indirect rates. Although data funded by federal money granted to the University are "owned" by the university, per this policy, there is no guidance as to whether those data are also subject to federal access policies and provisions.

⇒ We recommend that Policy 13015 be reviewed and revised to include guidance for researchers on both data sharing and data security.

Effectiveness of Data Management Plans

As public access to data is generally tied closely to federally funded research, reviewing data management plans from funded applications written by VT researchers helped the committee get a better sense of existing gaps in researcher and institutional knowledge around data management best practices. More thorough studies of data management plans submitted to NSF have been performed at other universities, but for the purposes of this committee, a subset of DMPs from recently awarded grants by a variety of federal agencies were anonymized and made available to the committee for review. Rather than considering whether the data management plans were "good" or "bad," the committee instead discussed whether the author had been able to find appropriate guidance about local or disciplinary services and whether the university and researcher would be able to enact the data management plan as written. From a legal standpoint, OSP considers federal grants as awarded to Virginia Tech, not the individual PI, thus, any change to the documents submitted in the application must be approved through the university. A funded application is considered a legally binding document, including the data management plan. Thus any modification or change to the plan needs to be approved through OSP.

The award lifespan for a federal grant can be anywhere from 1 to 5 years, even without adding on the additional 1 or 2 years between submitting the application and awarding the grant. In the intervening 2 to 7 years technology, services, support, and infrastructure can change rapidly and without warning. In the reviewed DMPs, the committee found a number of dead websites

referenced as preservation and access mechanisms, including one or two that had, at one time or another, been maintained by the university. Similarly, outdated technologies were referenced, and quite a few researchers were either not aware of the federal requirements for public access to federally funded data, or considered making the data available on a case-by-case basis "upon request" to be sufficient. While there certainly are instances where "upon request" is warranted, the burden of making such a justification in the DMP falls on the researcher. The DMPs the committee investigated did not provide this justification. Often researchers start from a position of secure or private data, many times for good reason, either because of privacy concerns, requirements from IRB, or because of NDAs from the private sector. However, often data that do not contain IP, PII, or restricted assets are not made publicly available because of a fear of "scooping," because of changes in federal mandates, or simply because researchers are not informed about the difference between data that can be made public and data that cannot be made public.

Over half of the reviewed DMPs could be fully enacted either as written or with minor revisions given current services and technologies in existence at Virginia Tech and elsewhere. The remainder, however, are in need of clarification or major revision. Beyond the comments discussed above, common issues include confusion over whether data can be disseminated via traditional journal and conference publication methods (in most cases it cannot), lack of clarity around data types (raw vs analyzed vs publication-ready), and lack of knowledge of standard data publication and preservation practices (i.e. stating that the High Performance Computing servers have long-term preservation and access mechanisms, which they do not). Although the committee agreed that these issues are problematic from a compliance standpoint, they also agreed that researchers should not bear the burden of these revisions on their own.

⇒ We recommend that the University Libraries should work with the Office of Sponsored Programs to create guidance for PIs who are working on grant applications and processes for ensuring that Data Management Plans are updated and followed over the lifetime of the grant.

Recommendations

Stronger partnerships between units engaged in research support will lead to faster and more efficient solutions for common problems, will ensure that researchers have the information they need when they need it, and will lead to stronger grant proposals. The difference between managing research data poorly and well is largely that of time and the availability of easily accessible guidance and support. The more guidance and support available to researchers, and their graduate students, the more time/cost effective data sharing will be. The cost of making data publicly accessible is significant; if enacted strategically, this burden could be divided across the researcher, sponsored programs, research compliance, IT, and the library. However, creating these partnerships takes time and resources, both of which are usually in short supply. How to incentivize, assess, and value these partnerships at a high level is still an open question for the Committee (and possibly beyond its purview), but we recommend that the Commission

on Research take an interest in and consider ways to better encourage the formation of these partnerships through a standing Committee dedicated to coordinating these research-data related services and relationships between the Research Office, the Library, and the Division of IT. Potential work for this Committee could include:

4. revising Policy 130015 to provide more pertinent information to researchers on how and where to obtain support for both restricting access to data and/or publicly sharing data,
5. working with the Office of Sponsored Programs to create guidance for PIs who are working on grant applications and processes for ensuring that Data Management Plans are updated and followed over the lifetime of the grant,
6. and working with the Research Compliance Office to create guidance helping researchers know when they can and should publicly share their data, and when they can and should keep the data secure.

These recommendations could also apply to general research support across the university. It was interesting for the Committee to note that several members were involved in a parallel discussion on restricted and secure data. While public access and restricted access are normally opposing endeavors, we found that researchers had the same types of questions about restricting data as they did about making data public.

⇒ Thus, another, more specific, recommendation is that the Library, IT, and Research Compliance work together on creating a guidance system to help researchers know when data should or could be shared, and when it must remain secure.

Supporting researchers who want or need to make their data publicly accessible inarguably contributes to Virginia Tech's land-grant mission. Publicly sharing data and other research products can lead to improvements in the public good. However, knowing when, what, and how to make things publicly accessible is much less clear. Researchers need their institutions to provide this guidance and these services. We know that Virginia Tech will rise to meet these challenges in the spirit of its living motto, Ut Prosim (That I may serve).

Committee Members:

Andrea Ogier	University Libraries
Jonathan Petters	University Libraries
Lauren Magruder	Office of Sponsored Programs
Randy Marchany	Division of Information Technology
Elizabeth Grant	Commission on Research (Architecture)
Andy Volker	Grant Specialists (College of Science)

Dale Winling	Faculty (History)
Sarah Stamps	Faculty (Geosciences)
Samantha Harden	Faculty (School of Medicine)
Bill Huckle	Graduate School
Steve Capaldo	University Legal Counsel
Bob Settledge	Advanced Research Computing
Sarah Parker	Institutes (FBRI)
Rajaram Bagavathula	Research Faculty
Mary Potter	Research Compliance