Enhancing Collaboration Across the Research Ecosystem: Using Libraries as Hubs for Discipline-Specific Data Experts

Andrea L. Ogier
Virginia Tech
5094 Newman Library
Blacksburg, VA
alop@vt.edu

Anne M. Brown
Virginia Tech
3030 Newman Library
Blacksburg, VA
ambrown7@vt.edu

Jonathan L. Petters
Virginia Tech
4064 Newman Library
Blacksburg, VA
jpetters@vt.edu

Amr E. Hilal
Virginia Tech
5090 Newman Library
Blacksburg, VA
ahilal@vt.edu

Nathaniel D. Porter
Virginia Tech
304 Newman Library
Blacksburg, VA
ndporter@vt.edu

ABSTRACT

Computationally-intensive, cross-disciplinary research collaborations are difficult to create and maintain over time, though many yield impressive results. The need for researchers to share, maintain, and manage data is increasing, while also integrating new tools and approaches to make their work more impactful. The University Libraries at Virginia Tech has a team of disciplinary data and informatics consultants working to connect research environments on campus with emerging library services enabling collaboration across disciplines. Partnerships with university-level research service providers, such as high-performance computing (HPC) services and statistical data consulting, have presented interesting use cases and innovative solutions to common problems. While traditional library services may not overlap with high performance computing environments, new library services (such as data management, publishing, curation, archiving, and preservation) provide new avenues for collaboration and situate the libraries in a unique position in relation to research ecosystems. Moving large datasets from HPC environments into research environments present significant barriers to research data sharing between collaborators; working with libraries to make these datasets better organized and documented lowers some of these barriers. Discipline-specific informatics consulting allows researchers to integrate new tools and approaches to solve research questions. Here, we highlight the utilization, need, and scope of informatics and research data management services in and around libraries, while also providing examples of how these services have created new collaborations and adoption of improved research practices surrounding data management and integration of computationally intensive techniques (e.g. bioinformatics, humanistic informatics, etc.). This work lays the foundation for these services in an academic setting and the influence of such on the practice and experience of understanding data.

PEARC '18, July 22–26, 2018, Pittsburgh, PA, USA

ACM ISBN 978-1-4503-6446-1/18/07…$15.00
https://doi.org/10.1145/3219104.3219126

CCS CONCEPTS

• *Applied computing~Collaborative learning*

KEYWORDS

Research collaborations, data management, informatics, training

ACM Reference format:

Brown, A.M, Petters, J.L, Porter, N.D., Hilal, A.E. and A.L. Ogier. 2018. SIG Proceedings Paper in word Format. In *Proceedings of Practice & Experience in Advanced Research Computing conference, Pittsburg, Pennsylvannia USA, July 2018 (PEARC18)*, 6 pages. https://doi.org/10.1145/3219104.3219126

1 INTRODUCTION

As the barriers to computationally-intensive research infrastructures lower, needs grow rapidly for researchers to share, maintain, and manage data, all while integrating new tools and approaches to increase their work's impact. Many computational methods currently used across disciplines and academic sectors depend on access to, and connections between, diverse datasets. However, researchers who create or derive these data are rarely formally trained in methods to make their data interoperable, shareable, discoverable, and usable in the long-term. A variety of scholars, scientists, and publishers have expressed concerns about big data management [1], reproducibility and replication in science [2], and data analysis in high performance computing (HPC) environments [3], prompting development of new tools and services by startups and open source communities that profess to solve some of these challenges [4]. However, keeping track of every new tool, much less using it properly and efficiently in a research environment is a time-consuming task. The continual integration of new tools and protocol for using such tools is a burdensome task for researchers and research groups already strapped for resources. Additionally, integrating a brand new, untested tool or service into the research environment can be a daunting task that some researchers choose not to pursue. In many cases, traditional methodology is more time efficient and effective for researchers who cannot devote needed resources to exploring new tools and protocols designed, ostensibly, to make their processes more replicable and easier to manage and share. The gap between those researchers who can use these new tools and services and those who cannot afford to is widened by funders who prioritize interdisciplinary and collaborative research projects which necessitate ready access to these tools and services.

The University Libraries at Virginia Tech has taken a two-pronged, strategic approach to helping researchers close that gap regarding ease of integrating new technologies and working with data over the past two years: (1) we have built specific, client-centric services around making data more interoperable, shareable, discoverable, and usable as part of a research data management (RDM) team, and (2) formed an Informatics Lab comprised of disciplinary specialists whose mission is to help researchers embed these new computational tools and services into their research workflows. Although located within the University Libraries, the RDM and Informatics Lab do not adhere to the typical “liaison” or “embedded librarian” models within academic libraries, wherein trained librarians with subject expertise serve the information needs of specific departments or colleges. Rather, the Informatics Lab consultants act as partners to researchers and educators in order to bridge the gap between teams and the tools, techniques, and services needed to raise the effectiveness, reach, and impact of research and teaching at the University.

In this paper, we will begin with a brief consideration of informatics and RDM services developed in and around libraries before moving on to identify specific needs of researchers identified within RDM literature and through interactions with researchers at Virginia Tech. We will then provide examples of how situating disciplinary informatics specialists within the Libraries has provided new avenues for collaboration and adoption of improved research practices. Ultimately, we hope this aids in highlighting researcher and training needs in a university setting, as centered around the finding, cleaning, analysis, and sharing of data in various formats created across disciplines.

as in the design of microwave isolators and circulators.

2 ESTABLISHED RDM SERVICES IN LIBRARIES

2.1 Reviewing Current Programs and Services

RDM services in the US and UK have grown largely in parallel with federal or funder mandates; a recent scoping review of RDM literature from academic libraries has identified 301 full text articles on the subject, 256 of which were published after 2010 [6] aligning with the National Science Foundation requirement that all proposals submitted after January 18th, 2011 include a data management plan [7]. In the UK, organizations such as the Digital Curation Center [8] have been in operation since 2004, though the Research Councils UK did not have a RDM policy until April 2011 [9]. Interestingly, neither of these efforts on the federal funding level were specifically tied to libraries or information science; development of RDM services within libraries has been a natural outgrowth of digital curation and preservation programs. One of the first broad studies of data management within libraries was done in 2012-2013 through the Association of Research Libraries SPEC Kit Survey (see Fearon, et al) [11]. In 2014, a landmark study (Tenopir, et al) published in *Library and Information Science Research* found that only 25% of surveyed Libraries in North America offered RDM services, though many had plans to do so [12].

Data management and curation services are in natural alignment with the more well-known library services; data, much like books and journals, are research objects that need to be organized, preserved, and made accessible. Data generated in a completely foreign technological time period may only be a few years old in terms of software and hardware; with the rise of cloud services and ever-increasing transfer speeds, even terabyte-capacity portable drives may soon be a thing of the past. Tabular data created by or saved into a proprietary file format in the 1990’s or 2000’s may no longer be supported by modern software. Library data management and curation services help researchers document, store, and publish data in open, non-proprietary formats that are more likely to withstand technological change and be useful in the future. Additionally, with the constant creation of large datasets from simulation and modeling experiments, as facilitated through HPC, we have a strong need to effectively share and move this data to truly utilize the data in the most impactful ways. Methods must be in-place between HPC hubs and sharing mechanisms to easily facilitate this type of sharing and transfer.

2.2 Framework for Meeting Researcher RDM Needs

Our current research data management service offerings at Virginia Tech [12] are fairly typical of those at a US R1 University. We offer consultative assistance with the creation of data management plans (DMPs) for research proposals and projects, administer VTechData (an institutional research data repository [13]), and assist with curation of research datasets. Our DMP consults review research proposals with an eye toward helping researchers address data management and publication requirements before they are submitted to the funder. Simultaneously, our curation services assist researchers with publishing their data for long-term access and preservation in the local data repository, or wherever the data will be hosted. We assist researchers mostly on the "ends" of the research data lifecycle (e.g. planning and publication/sharing) and are therefore complementary to the services provided by the Informatics Lab, which assists researchers more in the operational phase of their research. Training in good data management practices, both in planning and towards data publication, continue to be critical in engaging researchers. Our research data management services are available to everyone at the university, thus we cannot always provide the deep domain data management expertise our researchers desire. In effort to moderate this limitation, it has been critical to build partnerships between library data management, Advanced Research Computing (ARC), and other university units supporting research data. These collaborations will be described in greater detail below and highlights the need for this kind of service in a university environment and discussion as a community as to the best practices to share large data sets as created by the HPC community.

3 THE RISE OF INFORMATICS IN LIBRARIES

3.1 Informatics Service Needs

While research data management in Libraries has largely been developed around a common set of services in the US and the UK, the term "informatics" has varied meanings across disciplines, institutions, and countries. Generally, in non-healthcare academic settings in the US, "informatics" is seen as the practical application of computer or information technologies to real-world problems; this term exists both in higher education curricula [14] and research centers [15]. However, the term has gained wide acceptance in the medical fields: bioinformatics [16], health informatics [17], and nursing informatics [18] are all applied fields related to medicine. While a number of iSchools now have informatics programs to train graduate students on informatics theories and techniques or informatics research centers, there is little to no evidence of informatics-based service development in Libraries beyond the services described herein. We observed a distinct need of researchers to incorporate discipline-specific computationally intensive tools into their research but lacked formal training. By developing services and hiring experts in these sub-disciplines, we have created a bridge-type architecture model to connect the computing power and tools of the university to the researchers, to increase research impact and use of such resources.

3.2 Framework for Developing Discipline-Specific Informatics Consulting

Development of informatics-based services within the University Libraries at Virginia Tech began in early 2014 with the creation of a Data and Informatics Consultant position co-funded with the College of Engineering, and quickly grew into an Informatics Lab with the hiring of Data and Informatics Consultants for the Sciences, Social Sciences, and Arts. These four positions work both as an interdisciplinary team and as individuals with deep knowledge of specialized research environments at Virginia Tech. They synergistically work with partners on campus (e.g. ARC, statistics computing, etc.) to determine best ways to harness big data of their specific disciplines and high-performance computing in order to analyze large data sets. The need for text-data mining resources is also increasing for many researchers.

Services within the Informatics Lab center around a hierarchical model of "collaboration - consultation - partnership"; each interaction is classified with one of these terms which increase in time and intensity and decrease in number. Consultations are the least time-intensive interactions and most numerous; partnerships are few, time-intensive, and complex but often result in the biggest outcomes (grant funding, publication in more broadly read journals, etc.). Each interaction begins with a question or a request which falls within the technical or disciplinary expertise, and ends with a defined answer, outcome, connection to an applicable library service, or a referral to another service area within the University. Our services have organically developed with minimal solicitation. Recognition of past success through previous partnerships has been a positive indicator of increased utilization. These partnerships have also highlighted the need to include the graduate and undergraduate population in this dialogue.

4 BUILDING SERVICES TO ENHANCE COLLABORATION

4.1 RDM Services and Training

A recent survey of practices surrounding research data at Virginia Tech has provided some insight into faculty attitudes towards the research process; one respondent specifically commented on the administrative burden of research:

> "The main problem is that faculty have to do their jobs (intellectual advancement, scholarship, etc.) and then also act as a full time administrative support staff for themselves. Time is the biggest factor. The physical things that are missing are: easily accessible server space, examples of how to meet the data sharing requirements of granting agencies, and collaborative networks where you can share your data without having to take on ANOTHER project to administer in perpetuity." (171) [19]

These types of statements are common across research survey results and serve to highlight perceived gaps between processes researchers use, they highlight a larger need in academic institutions and partners that can bridge the gap between big data and high-performance utilizing techniques and disciplines.

While RDM services, such as those herein described, cannot shoulder the entire burden of these administrative processes for researchers, our ongoing data services development aims to fill some of these gaps and make research processes more efficient and effective. The model we use for informatics consulting services, which places the informatics consultant within both the library and positions them as a disciplinary expert within the research context, also helps to bridge some of these gaps. Partnerships with other library units and service units across campus is vital to the success of these services for two reasons: 1) there is no one-solution-fits-all answer, and 2) the perception that these services somehow duplicate or overshadow services offered by other units on campus (such as IT or ARC services) must be avoided at all cost. Basing these gap-bridging services within the libraries places them in semi-neutral territory, so that they are not locked into disciplinary silos and allowing them the freedom to advocate for collaborative and transdisciplinary solutions to common problems. This type of consulting, while dependent on disciplinary knowledge, also relies on a non-disciplinary network of partners, services, and systems throughout the university.

Through these partnerships, we supplement our deficiencies with technical referrals. As an example, working with ARC has led to the integration of Globus, a secure high-performance system for transferring large datasets, with our institutional repository, VTechData. This enhancement will streamline the deposit process for researchers using ARC HPC for simulation data by allowing them to directly transfer their research into storage managed by the Libraries' preservation systems. This allows us to create efficient and reliable data publication and sharing processes for researchers and makes these large datasets more widely accessible. We have also partnered with ARC to promote supercomputing resources in order to orient researchers to such services and to be the service that performs some of the simulations that necessitate high-performance computing. This allows more projects to be filtered through and utilize ARC resources, while also having an individual that is well-versed with the system efficiently utilize the resources.

Additionally, we have recognized the need for training and education to the undergraduate and graduate population in this area of research data management. We created the graduate level class, GRAD5024 Data Management Skills, that showcases the importance and benefits of data management. This course guides student in creating a data management plan and is a first step to train students at the start of their research careers to better hone their research data management skills. Other topics that have been integrated into the course, some by trial and suggestions from previous versions of the course, include data organization, policies, documentation, storage, ethics, access, security, analytical tools, grant requirements (proposal and post-award), proper attribution, archiving, other issues and topics related to data in scholarly communication. This is a large amount of material to cover in a 1-credit semester course; however, this speaks to the education needs of researchers from any field in regard to data management and highlights a specific training need in this area. As we enhance our scope of services and while researchers need increase regarding utilization of big data in their disciplines, we expect these courses to grow in enrollment and our approach to be refined to especially deal with large amounts of simulation data (from many disciplines and in many formats).

In order to target an undergraduate population in research data management skills and ideas, we have created a hands-on, experiential learning environment that deals with data management, transformation, and utilization in a low-risk environment. We have done this by creating and working with a space-centered model, wherein we created studios that focus on a piece of the research process that is needed no matter discipline or background. To date, we have two studios that are being piloted – the Data Transformation Studio (DTS) and the Data Visualization Studio (DVS). These studios specifically focus on both the training of undergraduate students in data management, visualization, and usage, while also supporting the research community at the university. Specifically, we see these studios as being an essential piece of supporting the ever-growing needs of the research community by helping with analyzing, understanding, and visualizing data across disciplines.

In DTS, our experts and their team of student workers support the transformation, recovery, curating, and sharing of research data. DTS activities include supporting the transformation of proprietary file types to more accessible formats, providing hardware and software for the recovery of data from outdated and obsolete media, supporting the visualization of data using geospatial information systems (including geocoding, georeferencing, symbolization, and lidar processing), and providing processes and best practices for data publishing and archiving. In DVS, our experts and student workers offer consulting on data transformation and visualization, specifically focusing on working with researchers to conceptually and graphically communicate their findings effectively. In both studios, students work on individual projects and help researchers during walk-in hours in order to share their new knowledge in this field. We feel that this kind of experience is essential to undergraduates as they prepare for their careers given the role of data manipulation in a variety of careers in many disciplines. The synergy between training in these areas and the actual aid it provides researchers has created an interesting approach that we hope to further highlight as piloting stages finish. We present it here to begin the conversation and highlight the need for this type of service to meet the data management, analysis, and visualization needs of a growing population of university researchers.

4.1 Informatics Services and Training

To best meet the needs of an ever-evolving research environment in a university setting, we have recognized that the skills present in libraries are undergoing an organizational change; as Burton et al [20] note that "data-savvy" individuals who offer hybrid "T-shaped" skills offering both breadth and depth are seen as desirable in the data science and analytics communities. The breadth of experience in data management, curation, and stewardship, and depth of knowledge in a specific methodology, discipline, or practice, allow us to forge the partnerships necessary for this effort. We find this approach to be essential in defining a service set and meeting a need at the university, as well as finding ways to bridge often-occurring frustrations of researchers.

We have already successfully hired and developed the services of science informatics at the university. We partner with multiple colleges and departments to integrate bioinformatics, cheminformatics, and computational biology into the research domains. This has been a very in-demand field given the need for researchers to combine both experimental and computational approaches to be more competitive for federal funding and publications. This individual has served in multiple capacities to enhance the bioinformatics, cheminformatics, and general utilization of HPC in the domains. Molecular visualization and computational thinking are core tenants to this domain and have been integrated into several undergraduate and graduate level courses. How and in what ways can we leverage data, tools, and software to enhance scientific impact of research has been a key tenant of this position. Success has been observed through number of collaborations, partnerships, and grants proposed and funded.

Most recently, a need was identified for a specialist in social science data. Relative to other data-heavy disciplines, social scientists tend to be substantial users of secondary data. As a result, social science archives have been early adopters of both multi-institutional [21] and open-access data distribution models [22, 23]. Many social science studies require combining multiple data files with complex longitudinal or network structure and may also involve sensitive human data (such as relationships, health conditions, geodata, and socially proscribed behavior such as drug use). A social science data consultant can assist subject specialists (including digital humanists, who frequently deal with at least equally complex data to social scientists) in navigating this maze of data discovery, access and preparation needs.

In addition to these area-specific needs, a social science consultant is well-positioned to assist with planning both qualitative and survey data collection and analytics, adding value across many fields, including program and service evaluation across the university. Finally, both quantitative and qualitative researchers increasingly recognize the potential wealth of social media and other digital trace data, but lack technical computational social science skills, for example, to interact with online APIs or prepare text data to network and/or machine learning models [24].

Finally, engineering has been at the core of the consultation services in the Data Services Unit since its birth in Virginia Tech libraries in 2014. The scope of engineering data and informatics problems is very broad, which naturally exceeds the boundaries of the core expertise of our engineering data consultant. Therefore, focus on essential skills such as data wrangling and utilization makes up a core piece of this service. Additionally, utilizing scripting that more easily extracts and manipulates file formats that are more difficult to work with (e.g. audio files) is a key hallmark of this service. Making sure the research problem is well understood is also essential to providing the right help. In order to provide the best support, we have found that strategically working to clarify problems and distill questions down to a few key domains allows us to push the limits of our skill sets. Thus, as a group, we are able to assist clients at various points within the research cycle and collaboratively connect researchers with other library services enabling a greater impact of the research conducted. An example of the type of collaboration, we helped a Ph.D. student who had collected data to investigate veteran reintegration into civilian life. The client wanted to create personas based on common attributes of a group of veterans. We were able to help the client by extrapolating this research question as a clustering problem. This type of analysis was commonly solved manually in the student's discipline; however, we used methods from computer science, data analysis, and data visualization to create an algorithm that mined the veteran data for attributes that could be mapped to personas and more easily grouped.

Throughout the discipline-specific consultation services offered by our team, our service model aligns with the vision of enhancing use of emerging technologies in research-based activities. We have two objectives: (1) reduce the struggle in data-related research problems including data creation, finding, processing, analyzing, etc. and (2) replace the struggle with knowledge and practice, so that the client is more capable of dealing with the problem now, and in the future. Towards this second objective, we make sure the client has a technical share in solving the problem, whenever possible, while we provide resources for learning and practicing the new technologies or methods. However, the depth of each consultation depends on the distance between what the client needs and their specific area of expertise. Thus, each consultation may require different skills and depth of involvement. Recognizing the uniqueness of this distinction has allowed us to position ourselves as partners rather than demonstrators, which has significant advantages for all stakeholders in this process, including the client, the consultant, collaborating teams, and referring partners (e.g. ARC). Consultants gain a network of collaborators across disciplines across campus, clients gain enhanced impact and scope, and, due to the neutrality of the "library," can more effectively leverage resources in support of their collaborative work.

Areas of attention that arose during deployment of these services (and discussed herein) could be considered by others at academic institutions who are considering or offering similar services. These areas are often related to scripting, data processing, text analysis, and data parsing. Other areas of need have become apparent, centering on working with data collected from open, massive social media datasets as well as voice and video processing. Engineering technologies will continually evolve; there will always be new techniques emerging from academia or industry. Therefore, we try to stay on or ahead of the curve for emerging technologies that could someday be considered or adopted by our clients. In addition to our current skillset, we continually work to acquire new skills in trending technologies, such as machine learning and deep learning, that could be of interest to researchers. We see a future value to researchers needing an individual that can easily adapt and integrate new and innovative tools to handle their data and information needs. Placing these resources in a library environment allows researchers access to neutral, discipline-specific, information experts who can collaborate with them on research projects, ultimately enhancing the scope and impact of their work.

One of the unique values of our multi-disciplinary team of specialists is the ability to draw on specific expertise through collaborative consulting, where a primary consultant with a specific skillset can bring in trusted experts and draw on their particular experiences. This network of allies and partners is crucial to the success of these research services and depends on clear knowledge of the boundaries between service-providing units and a clear, transparent, well-documented referral process. If a social science researcher has questions about machine learning techniques, we know where to refer them. If someone has questions about mining or processing social media data, we put them in touch with our partners who specialize in those areas. The American Soldier [25] provides one example. This complex project combines the collection and dissemination of new crowdsourced transcriptions with existing survey data. Informatics consultants have lent expertise on diverse areas such as metadata standards, text analytics, decision logic for linking surveys and text responses, dynamic data visualization, and working with both humanists and computer scientists to ensure that the data are accurate and complete. This also allows us to ensure that the data are useful to the greatest variety of future researchers and educators.

We are often able to help researchers find the right services for human-data questions, secure data questions, and questions about statistical analyses; we consider every referral and every partnership as an opportunity for a future referral to us. The data needs of the research community are increasing rapidly, and by highlighting ways to strategically enhance the research impact by hiring discipline-experts in a mostly neutral location in regard to university organization, we hope to provide ideas for implementation in other universities and start the dialogue on how to best build and enhance the bridges needed across the university in regard to impactful, interdisciplinary work in both experimental and computational methods.

5 CONCLUSIONS

While the development of new tools and services should serve to drive innovative and cutting-edge research, the cost of using these new data-intensive tools and services may be too high for some researchers or groups. In these cases, services provided by modern research Libraries, like the University Libraries at Virginia Tech, can help to mitigate these costs by facilitating specific parts of the research process. By building a team of data-savvy disciplinary experts, as well as research data management teams, in the library, we can work with researchers across campus. This allows us to work closely and utilize other on campus resources (HPC) to enable collaboration and interdisciplinary methods. Ultimately, we have found that these bridges to enhance research impact have been well-received and fulfill a need of researchers, highlighting a niche for adaption in the academic research environment. When we have a few more years of service provision, and thus a longer track record, we look forward to more fully and quantifiably assessing the efficacy of our services.

ACKNOWLEDGMENTS

We thank our University partners and Advanced Research Computing at Virginia Tech for their advice and services.

REFERENCES

[1] Evans, B., Druken, K., Wang, J., Yang, R., Richards, C. and Wyborn, L. A Data Quality Strategy to Enable FAIR, Programmatic Access across Large, Diverse Data Collections for High Performance Data Analysis. *Informatics*, 4, 4 (2017), 45.

[2] I Estimating the reproducibility of psychological science. *Science*, (2015), 349, 625.

[3] Harney, J., Lim, S. H., Sukumar, S., Stansberry, D. and Xenopoulos, P. *On-demand data analytics in HPC environments at leadership computing facilities: Challenges and experiences*. City, (2016).

[4] Foster, E. D., & Deardorff, A. M. Open science framework (OSF). *Journal of the Medical Library Association*. (2017), 10**5**, 38.

[5] Surkis, A., & Read, K. Research data management. *Journal of the Medical Library Association : JMLA*, (2015), 103(3), 154–156. http://doi.org/10.3163/1536-5050.103.3.011

[6] Perrier L, Blondal E, Ayala AP, Dearborn D, Kenny T, Lightfoot D, et al. Research data management in academic institutions: A scoping review. *PLoS ONE* (2017), 12(5): e0178261. https://doi.org/10.1371/journal.pone.0178261.

[7] National Science Foundation. Dissemenation and Sharing of Research Results https://www.nsf.gov/bfa/dias/policy/dmp.jsp. Accessed March 25, 2018.

[8] Digital Curation Centre. http://www.dcc.ac.uk/ . Accessed March 25, 2018.

[9] Research Councils UK. RCUK Common Principals on Data Policy. (2015). http://www.rcuk.ac.uk/research/datapolicy/ Accessed March 25, 2018.

[10] Fearon, David Jr., Betsy Gunia, Sherry Lake, Barbara E. Pralle, and Andrew L. Sallans. *Research Data Management Services*. SPEC Kit 334. Washington, DC: Association of Research Libraries, July 2013.

[11] Tenopir, C., Sandusky, R., Allard, S., and Birch, B. Research data management services in academic research libraries and perceptions of librarians. *Library and Information Science Research.* (2014), 36(2).

[12] Virginia Tech University Libraries. Data management and curation. https://lib.vt.edu/research-learning/research-data-management-curation.html Accessed March 25, 2018.

[13] Virginia Tech University Libraries. VTechData. (2018). https://data.lib.vt.edu/ Accessed March 25, 2018.

[14] Ensmenger, N. Informatics- Technology education for people who want to solve real world problems. https://www.informatics.indiana.edu/. Accessed March 25, 2018.

[15] Center for Informatics Research in Science and Scholarship. http://cirss.ischool.illinois.edu/index.php Accessed March 25, 2018.

[16] Yarfitz S, Ketchell DS. A library-based bioinformatics services program. *Bulletin of the Medical Library Association*. (2000);88(1):36-48.

[17] Aken SN. Health Informatics for Medical Librarians. *Journal of the Medical Library Association : JMLA*. (2010);98(2):184-186. doi:10.3163/1536-5050.98.2.016.

[18] merican Nursing Informatics Association. Fact Sheet. https://www.ania.org/about-us/fact-sheet. Accessed March 25, 2018.

[19] Shen, Y.. Research data sharing and reuse practices of academic faculty researchers: A study of the Virginia Tech data landscape. *International Journal of Digital Curation*, (2015), 10(2), 157–175. https://doi.org/10.2218/ijdc.v10i2.359

[20] Burton, M., Lyon, L., Erdmann, C., and Tijerina, B. Shifting to Data Savvy: The Future of Data Science In Libraries. (2018). Project Report. University of Pittsburgh, Pittsburgh, PA.

[21] Lyle, J. ICPSR: A Consortial Model to Advance and Expand Social and Behavioral Research. *027.7 Zeitschrift für Bibliothekskultur / Journal for Library Culture*.(2014), *2*(1). doi:http://dx.doi.org/10.12685/027.7-2-1-49

[22] M. Crosas, The Dataverse Network®: An Open-Source Application for Sharing, Discovering and Preserving Data. *D-Lib Magazine.* (2011),17, (1/2).

[23] Finke, R., Bader C., and Whitehead, A.. Innovations in the Development and Use of Social Science Data Archives. Leadership in Science and Technology: A Reference Handbook, (2012), Volume II, pp. 668-675. William Sims Bainbridge, (ed.), SAGE Reference Series on Leadership

[24] Salganik, M. Bit by Bit: Social Research in the Digital Age. (2018). Princeton University Press.

[25] [25] The American Soldier. https://www.zooniverse.org/projects/tkotwim/the-american-soldier. Accessed May 25, 2018.

.