IJDC | *Conference Pre-print*

Piloting a Community of Student Data Consultants that Supports and Enhances Research Data Services

Jonathan S Briganti
Virginia Tech

Andrea Ogier
Virginia Tech

Anne M Brown
Virginia Tech

Abstract

Research ecosystems within university environments are continuously evolving and requiring more resources and domain specialists to assist with the data lifecycle. Typically, academic researchers and professionals are overcommitted, making it challenging to be up-to-date on recent developments in best practices of data management, curation, transformation, analysis, and visualization. Recently, research groups, university core centers, and Libraries are revitalizing these services to fill in the gaps to aid researchers in finding new tools and approaches to make their work more impactful, sustainable, and replicable. In this paper, we report on a student consultation program built within the University Libraries, that takes an innovative, student-centered approach to meeting the research data needs in a university environment while also providing students with experiential learning opportunities. This student program, DataBridge, trains students to work in multi-disciplinary teams and as student consultants to assist faculty, staff, and students with their real-world, data-intensive research challenges. Centering DataBridge in the Libraries allows students the unique opportunity to work across all disciplines, on problems and in domains that some students may not interact with during their college careers. To encourage students from multiple disciplines to participate, we developed a scaffolded curriculum that allows students from any discipline and skill level to quickly develop the essential data science skill sets and begin contributing their own unique perspectives and specializations to the research consultations. These students, mentored by Informatics faculty in the Libraries, provide research support that can ultimately impact the entire research process. Through our pilot phase, we have found that DataBridge enhances the utilization and openness of data created through research, extends the reach and impact of the work beyond the researcher's specialized community, and creates a network of student "data champions" across the University who see the value in working with the Library. Here, we describe the evolution of the DataBridge program and outline its unique role in both training the data stewards of the future with regard to FAIR data practices, and in contributing significant value to research projects at Virginia Tech. Ultimately, this work highlights the need for innovative, strategic programs that encourage and enable real-world experience of data curation, data analysis, and data publication for current researchers, all while training the next generation of researchers in these best practices.

Submitted 18 December 2020 ~ *Accepted* 19 February 2020

Correspondence should be addressed to Anne M Brown, University Libraries, Virginia Tech, 560 Drillfield Drive, Blacksburg, VA 24061 Email: ambrown7@vt.edu

This paper was presented at International Digital Curation Conference IDCC20, Dublin, 17-19 February 2020

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: http://www.ijdc.net/

Introduction

Statement of Problem

In a previous paper (Brown et al, 2018), we describe the data and informatics services offered by the University Libraries at Virginia Tech. These services were developed and updated in response to identified researcher needs. As we began to scale the consulting services described in Brown, et al. to match the needs of researchers in a university setting, we quickly realized that based on the number of personnel, we had a choice to make: either focus on cultivating more advanced, in depth research partnerships, or focus on achieving a breadth of service across the University. Often, our consultants are concurrently working with several partners on campus while also contributing to other University Library initiatives such as digital literacy and data education. In essence, we found ourselves confronted by the very problem we were trying to solve: like the researchers we were trying to assist, we also needed a solution that would scale effectively and allow us to support the "T" shape of breadth and depth across the research ecosystem (ibid.). Recognizing the need for more resources, specifically in individuals trained in data wrangling, visualization, and analysis, but with limited funding sources to hire more specialized, advanced-degree holding consultants, we looked towards the willing workforce present at the university - the students.

Utilization, cleaning, analysis, visualization, and curation of data are quintessential skills for students entering the workforce[1], regardless of discipline. Many students seek opportunities to participate in skill-building, resume-supporting positions. At Virginia Tech, the emphasis on undergraduate research and experiential education is becoming more of a driving force for students to gain hands-on, applied skills to make them more competitive for the job market, professional schools, and beyond (Börner et al, 2018). Although undergraduate research as a field has developed multiple on-boarding, compensation routes, and benefit structures to student and advisor (Hayward et al, 2017), students most commonly receive class credit that impacts their overall GPA and credits towards graduation. A smaller percentage of students are able to be hired for wage or summer fellowships that provide monetary compensation. Unfortunately, given the breadth of backgrounds and financial needs, many students are not able to fully participate in undergraduate research due to the necessity of monetary compensation and the small number of paid research positions. Often in these situations, a limit is also placed on the number of students able to join any one program, and these programs are in high demand (Linn et al, 2015). Many undergraduate research programs and labs have only a small number of positions, and even fewer offer the option for either paid or credit-based work. At a large, state-funded, R1 research institution, these positions and experiences are not feasible for all students to participate (Desai et al, 2008), and often completely out of reach for students from underrepresented and marginalized groups. While DataBridge cannot solve all these systemic problems at the university level, our program seeks to actively recruit students who may not have a clear path towards these experiential, hands-on opportunities, which have proven vital to their post-undergraduate successes (Lopatto, 2004; Healey, 2005). We have organized and designed the program pilot of DataBridge to begin to address some of these institutional challenges in order to create resources for further adaptation, to support students from any major, and to utilize hierarchical mentoring for scaling and student professional development. A structured approach (Brown et al, 2016), multiple types of project engagement opportunities

[1] Committee on Future Career Opportunities and Educational Requirements for Digital Curation; Board on Research Data and Information; Policy and Global Affairs; National Research Council. Preparing the Workforce for Digital Curation. Washington (DC): National Academies Press (US); 2015 Apr 22. 4, Preparing and Sustaining a Workforce for Digital Curation. Available from: https://www.ncbi.nlm.nih.gov/books/NBK293662/

(e.g. long-term, student-derived, consultation-based, or knowledge building projects), and a focus on the development of workforce skills are core features of the program to-date.

As DataBridge grows, we anticipate that offering students the ability to choose their method of compensation, dependent on resources, will help us encourage, support, and promote equity for students of underrepresented, underprivileged, and underserved groups on campus. Taking into account these pedagogically-backed questions and theories on undergraduate research and high-impact practices (HIPs) for students has already proved requisite in our pilot deployment of DataBridge. While the original "need" of DataBridge was to provide researchers with more human power to aid with their expressed research data needs, we sought to incorporate these pedagogical practices early in program design to create a sustainable, valuable to students, and scalable program. With the tenants of using, finding, analyzing, visualizing, and curating data of any kind and from any discipline, we hope to support and grow these necessary workforce-ready, durable skills for all students participating and positively influence the next generation of individuals using data in their chosen majors and domains. Herein, DataBridge addresses the growing needs of the research data community, regardless of discipline, by providing pedagogically-backed, HIP experiences for students to aid in current, on-going research data projects, and to also serve as a starting point for training cohorts of students in data best practices.

Evolution of DataBridge

Building on the successes of the data visualization studio in University Libraries (Ogier, et al 2018), and the known need of researchers for assistance and partnership on their data and computation heavy projects, we designed DataBridge as a program to train students with a focus on applied, data science skills. While we can train students in basic data science skills, the applied portion is necessary to individually characterize. Critical thinking, "does it make sense" type of logic thinking is necessary to address and provide minimal risk environments for, so that students can authentically grow and develop these skills. After completing a micro-project in the training phase of the DataBridge program, students can then apply those skills to projects the Data Services faculty in University Libraries at Virginia Tech either lead or are consulting and partnering on with more confidence and empowerment to deploy best practices regarding the project.

Using a framework and a scaffolded approach, we sought to align two needs of the university - to provide more sophisticated and embedded research services while utilizing and training a student workforce to aid in the "heavy lifting" some research projects require. This program has been named "DataBridge", in an effort to provide "bridging" services that connect data experts in service-based units to the research and teaching community at a university, as well as providing a "bridge" for students between knowledge gained in their courses and the practical application of that knowledge in their careers. Here, we describe the role of such a program on the data lifecycle at a research university, the skills introduced to students to work in the DataBridge program, and the lessons learned and future implications of such a program in the realm of data curation and education. DataBridge is designed to be adaptable by other institutions and highlight the influential and impactful role that students can play in research data services and the university. Service units, such as Libraries, IT, and core service centers, can potentially conceptualize how to integrate students into these roles with the structured program we have designed.

Initially, DataBridge was envisioned to support the middle of the data lifecycle, primary data cleaning, transformation, and analysis. Focus was placed on DataBridge students entering the role of an external consultant paired with a project that was underperforming or progressing outside the expected timeline. The student would familiarize themselves with the workflow, tools and technology, and the importance of the work at hand, leaving the domain expertise on the research or research group. As DataBridge moved through the pilot phase, we noted that

projects were consistently not at the stage that the clients assumed or conveyed. Due to unfamiliarity with new concepts, unawareness of past mistakes or potholes, and assorted assumptions, researchers were undertaking projects that ultimately needed auxiliary and unanticipated support to complete, especially revolving around the data of the project. It was realized that DataBridge students could act as consultants, able to jump in at any point in the data lifecycle, including project planning and scoping, data and dataset discovery, data visualization, data wrangling, API usage, digital curation, script production, database prototype and development, and more. Juxtaposed with this was a decision to focus on longer-term, in-depth consultations as the core of our services. This enabled Data Services, the department where DataBridge is housed, to get a better foothold on the trajectory of a project, integrate other support services such as Data Management Planning (DMP), spot issues further away (e.g. the types of visualizations requested were not possible with the types of data collected), and provide the student with a more encompassing experiential learning opportunity and exposure to how research works in both the academic and industry setting.

To support the transition to complete data lifecycle support, we created a series of modules that guide students through learning and applying essential data science skills. These modules cover fundamentals of data science, accessing and wrangling spreadsheets, how to ask a question (and then answer it with data), interactive and flat visualizations, data ethics and accessibility, data curation, and programmatic data science. Each student, regardless of background or prior knowledge, is required to complete all modules and deliverables to ensure consistent proficiency and equitable practices. In a module, a student will learn the technical skills and then work to apply them to an overarching micro-research project. This micro-research project acts as their experiential project throughout the training phase (e.g. first semester in the program) and ensures the student can apply the skills gained in a novel environment. Students are encouraged throughout the training phase to identify pro and con points in relation to the field and the tools used. The curriculum emphasizes the use of tools and techniques that are relevant to their desired career, integrating specialized skill growth naturally in a way that promotes student wellbeing and success alongside digital curation. Currently, we utilize the following programs/languages: MS Excel, R, Python, Tableau, Adobe Creative Cloud, AWS, and JavaScript. These are a main focus, with the instruction to students that other programs, preferably open-source/open-license, could also be used depending on the type and format of data and the desired output. In these cases, we utilize peer-peer and peer-mentor support for these questions, weekly group meetings, and ongoing, shared documentation in our student training materials.

In this way, the resulting trained students enter the working phase of DataBridge with continually developing specialties and interests. The scaffolded process provides flexibility for students to experiment, and provides a stable base of students trained in a wide variety of skills, tools, programming languages, and more. The process of taking any major, requiring foundational skill equilibrium but specialization possibilities, and providing real-world projects continually reinforces the transdisciplinary nature of data science. DataBridge students must function within an environment comprised of vastly different skill sets and backgrounds, mimicking industry expectations.

Pushing forward in the evolution of DataBridge, we are continuing to grow our paid student positions. While this still represents a small subset of our total student population, our ability to pay students, whether by operational or grant funds, provides students with powerful learning opportunities that can help to offset some of the financial burden of college education.

Impact on Research Data Services

The role of research data services programs varies across institutes and universities. Often, these services are housed in multiple departments and locations, each of which may offer a wide-range of services and expertise while serving only a select community within the whole. Positioning

DataBridge inside the Data Services unit of an academic research library wholly benefits the entire community in improving research data practices and offers true trans-disciplinary experiences and advantages to both students and faculty. In addition, there are significant benefits to research data services, including the opportunity for data management and curation specialists to interact with more individual projects, and to train students, the next generation of researchers, in best practices.

Examples of the influence of the DataBridge program on Data Services impact include the ability to do more consults and at varying levels of depth. To date, ~30 collaboration or partnership type consultations have been performed with the aid of a DataBridge student. These consultations have ranged in type, service, and discipline. Students have worked on projects that involve creating a database of historical election data, creating a JavaScript-based, CC-BY licensed pKa tool for instructors to use in a classroom,[2] creating open content for using programs and documentation via an Open Science Framework Page,[3] using natural language processing (NLP) on academic and popular sources for trends involving antimicrobial resistance and food safety, finding and analyzing song lyric data for trends on word uniqueness over time, and more. Two of these projects (pKa Analyzer and Chicago Elections Project) are detailed more below to highlight the student's role in the project. These projects have allowed both the faculty leading DataBridge and the students to keep up-to-date on new tools and techniques, has resulted in four funded grants that include a DataBridge student (seven submitted), and has scaled from 4 students (Spring 2019) to 18 (Spring 2020).

Another area of influence on research data services is the ability to prototype and pilot new services domains. Often, once a protocol, library, or tool kit is explored, those techniques can be applied more easily and quickly to other consultations. An example of this is natural language processing (NLP). Originally, researchers approach Data Services to access academic papers in batch for a text data mining project. A solution was developed to utilize a program that pulls open-access academic papers as a start to the process for this research project. This would provide a reusable and reproducible aspect than using only behind-the-paywall academic papers. From here, text data mining and NLP processes using both Python and R were utilized, with students determining pros and cons for each program. Researchers were originally surprised on the level of support provided, but it ultimately was necessary as researchers needed more expertise and help in using these types of data analysis (NLP) and visualization. Projects in a different field have now requested Data Services collaboration in a similar project, where similar protocols can be utilized, ultimately speeding up the process and providing the support needed across campus.

Another area of growth and impact on research data services is database prototyping. Data Services has begun to enhance its service scope, providing database prototyping to the Chicago Elections Project.[4] This type of support service has been long-requested by many faculty, but has not be sustainable given the time, scope, and technical expertise needed. DataBridge faculty developed an extensive questionnaire in order to best assess needs of individuals requesting databases. Often, they might not need a database but rather a place to store data. This aids in the research data services consulting practice and with scope definition of research projects. These recognition of need for these types of materials (e.g. questionnaires to determine scope) have been useful in also acclimating student consultants to research projects.

Finally, the last area that research data services has been impacted is in the ability to have students do more experimental consults. Sometimes, it is useful to have exploration before committing to a project and to ensure the client knows the experimentation is necessary in order to determine feasibility. Students have done this in determining the use and scope of popular GIS-related programs including kepler.gl, R, and ArcGIS. By having a student explore and present the pros and cons of each program on the same project, it allows consultants and clients make more informed decisions.

[2] https://databridgevt.github.io/pka-analyzer/
[3] https://osf.io/82n73/
[4] https://twitter.com/chielections

Creating a student-centred program

Experiential education, which involves undergraduate research, internships, service learning, and other high-impact practices as described by AAC&U (Kuh, 2008), can be extremely beneficial for students. While creating DataBridge, we sought to maintain all elements of an experiential experience, similar to one a student would experience in any other department, lab, or environment, and incorporate data science training and pedagogical concepts to enhance program experience. While more "hands" are necessary to support researcher data needs on a university campus, having the program be beneficial, with a focus on student-centered learning, was also an important factor. This necessitated intentional framing and implementation to properly balance both factors.

DataBridge is an embedded student research program built to provide academic and real-world research experiences to undergraduate students and to support high impact practices within the University Libraries. This program connects the theoretical and practical application of data science skills by offering students an opportunity to work collaboratively with a team of interdisciplinary informatics consultants who are solving problems that make research processes more efficient and impactful. Faculty in Data Services oversee these students and are leaders on project collaborations with faculty partners across the University. In this way, students can gain valuable knowledge and experience while faculty can ensure project success. Collectively, as a team, we are more able to provide depth in services that are often sought after by researchers. Uniquely in our department, we have domain-specific experts that allow further depth and breadth of services to effect to research lifecycle of curating, finding, using, analyzing, and storing data. However, first, students must be trained in necessary skills in order to work on consultations and partnership projects. This training should be accessible, universal, and focused on needed and wanted data science tools to solve researcher's data needs.

Training, doing (researching), and reflecting are three of the key stages in an experiential experience, as defined by Kolb (2007). To provide a consistent experience within DataBridge, students iterate through the stages of experiential learning in stratified data-driven skillsets. These skills, broadly, include research data manipulation, visualization, analysis, and database prototyping. Research groups are asking for more training and curriculum for fundamentals such as data wrangling and analysis, while also considering integrating new tools and research plans (primarily visualization tools and quickly iterative databases). By focusing on the foundational data-science skill required to be a data steward, producer, curator, or researcher, we seek to teach undergraduate students' data-driven and problem-solving skills that can be integrated into existing research models. To address these skill gaps and ensure consistent student success, we have created a model for facilitated experiential learning that allows students to grow in responsibilities as skill sets develop.

The model (Figure 1) requires all students to iterate through the same program regardless of skill level or entry point. We have found that this locates gaps in knowledge, provides a consistent foundation, and reduces stigma of working in a 'lower' stratum. In each tier, the student has four goals that must be met before progressing to the next level. These goals are: working through a detailed guide that provides necessary technical skills, a self-reflective assignment on the data ethics surrounding the tool or skill set, a group discussion regarding ethics, and an independent case study. These skill-driven goals are coupled with concepts and practices to facilitate this process and maintain an inclusive community of practice for all students.

To monitor independent student progress, we have integrated many concepts from a traditional academic classroom. Such concepts include a syllabus, weekly updates, final presentations and papers for all students, as well as reflections and deliverables for each training and curriculum set. These aspects -- syllabus, using a learning management system, and weekly updates are essential in our program and its scalability. For anyone replicating or incorporating a similar program with students, it is highly recommended to include these aspects for both student and program director benefits. Students benefit from the structure in that it looks like a typical course and what they are accustomed to in those domains. They equate the same level of

dedication and work and can easily access all deliverable due-dates and modules for training as they do a typical college course. This stratification of learning and scaffolding provide an environment where students can develop uniquely with room for modification as needs arise. Finally, structured approaches are also important in making the program available and inclusive to all students wanting to participate (Tanner, 2013). It is recommended for any institutions or departments wanting to implement a similar program that these aspects be incorporated.

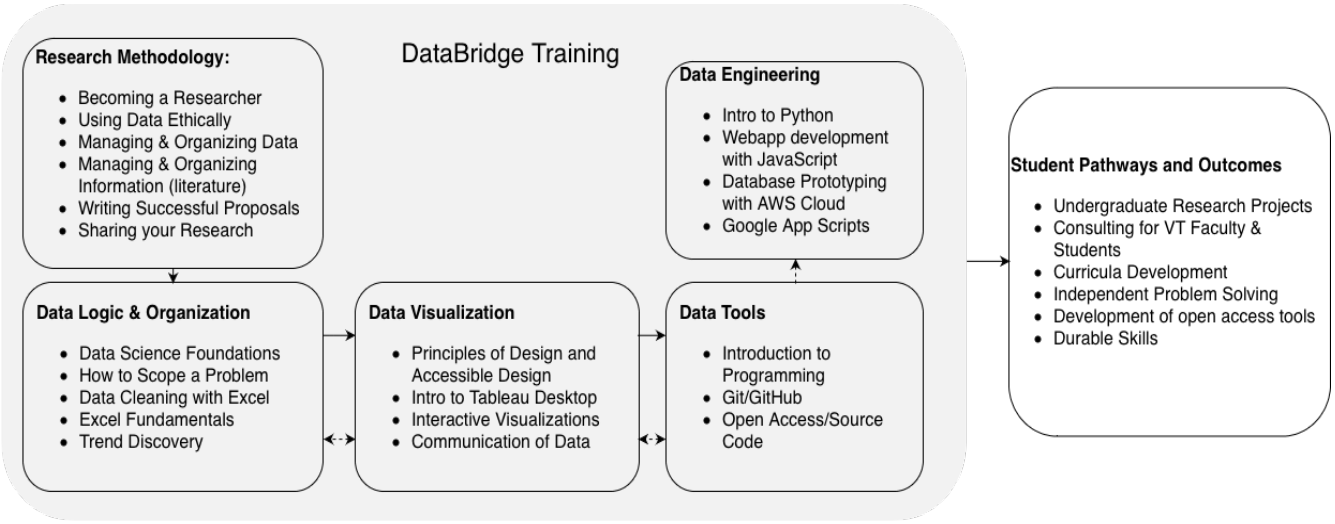

Figure 1. Scheme representation of the DataBridge training and on-boarding process.

Prior to incorporation, it was difficult to track students, students were often unaware of their progress or how they were doing, and deliverables were not at the level expected. These organization/structured aspects allow students to have and feel a "common" or "familiar" approach when dealing with hard research questions. Curriculum modules and necessary trainings include data tables (wrangling, managing, trend discovery, and program utilization), data visualization (utilizing pre-built software and libraries to accurately visualize information), and programming for data science (using Python, JavaScript, or R to build and use libraries to create and carry out complex queries), with each set acting as a strata in the DataBridge model. We are currently working to make these modules accessible and usable under a CC-BY license after they are appropriately tested out on two cohorts of students. As students complete the DataBridge model, they are partnered with Virginia Tech community members based on the researcher's need, student aptitude, and experiential opportunities. The development of modules is an initial time commitment to the program but decreases over iterations, making the program more sustainable and not requiring more faculty or staff employment.

Ideally, DataBridge serves as a hub for interdisciplinary, data-centric learning, with students engaging in research projects across all disciplines. These research projects stem from research consultations and student-driven collaborations which have promoted the growth of diverse real-world projects for students to complete. We have begun to explore how the students in DataBridge are acting as data stewards or data champions. We have begun to integrate more data ethics into training and into weekly discussions, in addition to planning for a section on DMP training. Our current DMP coordinator is consistently booked with individuals asking for assistance in writing DMPs. We are considering what student support looks like in that role and what kinds of materials we can create that will be better at portraying our messaging surrounding best data practices, incorporating FAIR data principles, and considering the long-term location of data. Regarding the role of a student being a data champion and stewards, we are not necessarily aligned with most technical definitions. However, what we are observing and promoting is that our students are using and working with data in a variety of contexts, discussing successes and failures in a group environment, and hopefully advocating with their

peers and in their practices outside our research group. An ideal scenario would be students revamping their current data storage and documentation practices (e.g. data management), telling their peers about their changes and encouraging implementation, and finally mentioning these learned best data practices in other work or lab environment so other faculty are also aware or more encouraged to participate. This is our current working goal and we are brainstorming ways to collect or capture this process, assess students in their data practices and impact of the program. We currently get many of the reflections of "how many times they have overwritten a file", but we seek to see if we can quantify that behavior changes after participating in DataBridge. Based on reflection and qualitative responses, we have indication that this is true, but more formal analysis is the next stage of this program and project.

"Bridging" the gap in researcher and student needs

DataBridge operates in the space between the groups around Virginia Tech that are asking for help with research data and students who wish to develop data science and applied, critical thinking skills. By working in the domain-agnostic space provided by University Libraries, we can connect students with peers and faculty from across the university while also ensuring foundational data skills are consistently taught. The need and desire to work with data is poised only to rise, requiring more hands and minds to work on continually multidisciplinary projects. This requires consultants that can help create solutions across the data-science spectrum. By pairing this technical curriculum with real-world applications, high impact practice (HIP) assessments from AAC&U, and an emphasis on experiential learning, we have added crucial skills and structure into a program that is based on collaborating and enhancing the use of data in an academic university. Once on a project, these interdisciplinary "unstructured" research experiences prepare students to act as problem solvers who are able to iterate and implement their skills rather than knowing single solutions. Additionally, we are providing diverse and robust services to researchers regarding their data and research impact.

To show the current utilization of DataBridge students in the research data lifecycle, we highlight two projects that involve creating, manipulating, analyzing, visualizing, and sharing of data for both academic and public consumption. These projects involve the generation of an online, CC-BY licensed pKa tool for instructors to use in a classroom,[5] and a long-term project seeking to develop a database of Chicago local, state, and federal election data from the 1800's to current.

Students involved in the pKa project were partnered with chemical engineering professor who wanted to create an open-access tool to quickly calculate expected values after an acid-base titration for a course taught to non-majors. This required the DataBridge student to familiarize themselves with titrations, the chemical reactions discussed and data to be incorporated, consider the best language and libraries to use to ensure open-access and accessibility, and to create a JavaScript based tool that could operate as a stand-alone app and as a web app. Design thinking was incorporated into this project and the accessible design of the interface and the student got to experience the full cycle of finding the data (e.g. pKa values), organizing it into searchable formats, integrating in calculations, and accessible visualizations. This project was presented at the Virginia Tech Institute for Creativity, Arts, and Technology (ICAT) Creativity and Innovation Day 2019. Data that was currently available in 2-D representations in textbooks and indices is now organized and visualized in a way to relate back to concepts discussed in the course and to make learning more intuitive and hands-on, rather than static. The build of the tool allows for the integration of new chemical moieties and values as the instructor or others need, highlighting usability long-term.

The Chicago Elections Project saw students collaborating with historians to create a database and public facing website that houses digitized records of election and voter data for more than 100 years of records. This project required students to develop the best stack choices

[5] https://databridgevt.github.io/pka-analyzer/

for a cloud-based application, develop and maintain data schemas, and create digitization processes that can be readily implemented by people without data-science skills. As this project seeks to incorporate many pieces of data from several sources, that were created largely prior to the idea of digitization practices, students had to refine strategies to account for unexpected edge cases. We have also created GIS-based visualizations of the data, alongside other quantitative analysis and visualizations. While presenting the Chicago elections data alone can provide meaningful access to information, these visualizations help promote interest and questions surrounding the Chicago elections. Long-term plans for this project include further grant submissions to national agencies, partnering with other universities to expand the scope and data content of the project, and recently a small portion of this data has been published via the Virginia Tech repository, VTechData. This project especially highlights the role of digital curation and data longevity.

Both projects pushed students to work outside of their comfort zones, explore domains that is not in their primary major-scope, and work to implement their skill sets in data science in a new field. Both projects were very well-received by the clients, more collaboration projects initiating with these and referrals to other faculty. Faculty have incorporated the works produced by students into professional presentations, grant applications, and more at this stage in the program. Working with such diverse projects helps students gain confidence, critical thinking skills, and a continually developing appreciation for the vastness of data science. Students working on both projects have or will present at on-campus research symposiums, are now funded via grants that support this work, and are using these opportunities to enhance their resumes and interviews for their next steps after graduation.

Lessons Learned and Future Implications

Ultimately, this work seeks to discuss the need for more training and support of research data services and how that need is met by engaging students in a formalized, well-curated program hosted through University Libraries. We had two areas to address -- providing more support in the data services domain for researchers and engage students in an authentic inquiry research experience.

Support for the entire research data lifecycle by researchers in a university setting, specifically deeper engagement by data experts, is of need and request of faculty staff and students. This request and need is observed across disciplines, including social sciences, sciences, engineering, and the humanities and has been met with the expansion and development of Data Services to hire domain-specific informatics experts. To further match demand, student person power has been employed in order to provide consulting services in the range of consultation, collaboration, and partnership projects. Additionally, we have partnered on several grants that allow us to aid in the broader impacts of these grants by engaging in an education/training impact while also accomplishing research goals. In order to truly impact research data, by training students in best and FAIR data practices, are we able to begin to make a difference in data utilization and sharing.

We have learned the logical and organizational needs for piloting and quickly scaling such a program. Those include treating the experience similar to a course so that students can have structure and solid foundation when approaching sometimes hard-to-solve data challenges. We have incorporated hierarchical mentoring in order to enhance peer-peer networks and provide leadership opportunities for more senior students. Additionally, this allows for program sustainability as well, with peers aiding in the training process and allowing the "team leads" to answer more novice questions which can be time consuming for faculty. To date, students have found success in working in DataBridge and have outcomes such as presenting at local conferences, summer fellowships, and being authors on papers. They also have a much broader sense of data best practices and what data means to them in their classes, field, and hopefully their futures. We are currently developing more quantitative methods to assess this response and

anticipated growth in becoming data champions. Qualitative and anecdotal responses to date highlight that students are integrating these practices into their schoolwork, especially emphasized by past issues with data handling that have led to headache. While not fitting the typical definitions of a data steward or champion, but adding n training in data ethics, management, and the observation of others challenges with data, we hope to have created a cohort-based program of individuals that will be the pioneers in best data practices in their respective fields.

We present here the work on utilization and impact of such a program like DataBridge in research data support services and highlight practices incorporated that have led to scalability and improved student engagement when piloting this program. Future endeavors include creation of quantitative assessments that can showcase the long-term impact of such a program on student workforce development and in research data support services.

References

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, Pennsylvania USA, July 2018 (PEARC18), 6 pages. https://doi.org/10.1145/3219104.3219126

Committee on Future Career Opportunities and Educational Requirements for Digital Curation; Board on Research Data and Information; Policy and Global Affairs; National Research Council. Preparing the Workforce for Digital Curation. Washington (DC): National Academies Press (US); 2015 Apr 22. 4, Preparing and Sustaining a Workforce for Digital Curation. Available from: https://www.ncbi.nlm.nih.gov/books/NBK293662/

Börner, K., Scrivner, O., Gallant, M., Ma, S., Liu, X., Chewning, K., ... & Evans, J. A. (2018). Skill discrepancies between research, education, and jobs reveal the critical need to supply soft skills for the data economy. Proceedings of the National Academy of Sciences, 115(50), 12630-12637.

Hayward, C. N., Laursen, S. L., & Thiry, H. (2017). Why Work with Undergraduate Researchers? Differences in Research Advisors' Motivations and Outcomes by Career Stage. CBE life sciences education, 16(1), ar13. doi:10.1187/cbe.16-07-0229

Linn, M. C., Palmer, E., Baranger, A., Gerard, E., and Stone, E. (2015) Undergraduate research experiences: Impacts and opportunities. Science 347, 1261757.

Desai, K. V., Gatson, S. N., Stiles, T. W., Stewart, R. H., Laine, G. A., and Quick, C. M. (2008) Integrating research and education at research-extensive universities with research-intensive communities. Adv Physiol Educ 32, 136–141.

Lopatto, D. (2004) Survey of undergraduate research experiences (sure): First findings. Cell Biol Educ 3, 270–277.

Healey, M. (2005). Linking Research and Teaching Exploring Disciplinary Spaces and the Role of Inquiry-Based Learning. Reshaping the University: New Relationships Between Research, Scholarship and Teaching, McGraw-Hill/Open University Press, Maidenhead, pp. 67–78.

Brown, A. M., Lewis, S. N., & Bevan, D. R. (2016). Development of a structured undergraduate research experience: Framework and implications. Biochemistry and Molecular Biology Education, 44(5), 463-474. doi:10.1002/bmb.20975.

Ogier, A. L., & Stamper, M. J. (2018). Data Visualization as a Library Service: Embedding Visualization Services in the Library Research Lifecycle.

Kuh, G. D. (2008). Excerpt from high-impact educational practices: What they are, who has access to them, and why they matter. Association of American Colleges and Universities, 19-34.

Kolb, D. A. (2007). The Kolb learning style inventory. Boston, MA: Hay Resources Direct.

Tanner K. D. (2013). Structure matters: twenty-one teaching strategies to promote student engagement and cultivate classroom equity. CBE life sciences education, 12(3), 322–331. doi:10.1187/cbe.13-06-0115.