Research Computing's Demand for Humanists, and Vice Versa

Quinn Dombrowski
quinnd@berkeley.edu
UC Berkeley, United States of America

Tassie Gniady
ctgniady@iu.edu
Indiana University, United States of America

Megan Meredith-Lobay
megan.lobay@ubc.ca
University of British Columbia, Canada

Jeffrey Tharsen
tharsen@uchicago.edu
University of Chicago, United States of America

Lee Zickel
lxz11@case.edu
Case Western Reserve University
United States of America

Over the last five years, research computing has undergone a shift from focusing on running infrastructure for highly technical researchers, primarily in the sciences, to supporting medium- to large-scale computational needs across a wide range of disciplines, where practitioners fall across the spectrum of technical proficiency.

This shift in approach has opened up new opportunities, both for scholars whose research questions are facilitated by computationally intensive algorithms (e.g. photogrammetry, natural language processing, OCR at scale), and for humanists in support positions that require translation between technical and non-technical audiences. This panel will discuss opportunities for research and career development through the perspectives of research IT staff whose backgrounds in both humanistic inquiry and computational and digital methodologies enable them to engage in outreach to, training of, and support for humanities researchers. It will also provide practical suggestions for humanists who could benefit from research computing resources but find it difficult to navigate the expectations of research IT organizations.

The communication, writing, and teaching skills cultivated through an advanced degree in the humanities align with employment trends in research IT groups. While some familiarity and comfort with computation is generally a requirement for working in research IT, advanced programming or system administration skills are crucial for a minority of positions in these groups. As the mandate of research computing groups has expanded, it has become increasingly clear that successful research computing programs require documentation that is comprehensible to a non-technical audience, hands-on workshops for non-specialists, and staff capable of understanding the fundamentals of researchers' work and identifying effective approaches to meeting their computation needs. To that end, in 2014 the US National Science Foundation funded the Advanced CyberInfrastructure - Research and Education Facilitators (ACI-REF) program, which developed a cohort of computation "facilitators" across a number of universities who could "maximize the impact of investment in research computing" by "assist[ing] researchers in leveraging ACI resources for themselves" and sharing solutions among participating institutions (NSF 2014). This approach to providing support has been highly influential among campus research IT organizations, and related workshops such as the ACI-REF virtual residency (featuring plenaries such as "Effective Communication: How to talk to researchers about their research" and "Writing Grant Proposals") have drawn large crowds (Neeman et al. 2016).

Similarly, the Extreme Science and Engineering Discovery Environment (XSEDE), funded by the National Science Foundation, has begun specifically reaching out to disciplines that have been typically under-represented in the high performance computing arena, including those in the humanities. XSEDE has a program called Extended Collaborative Support Services (ECSS) that partners humanities scholars and others from under-represented areas with computing professionals who can help them achieve their desired research objectives (Wilkins-Diehr et al. 2016). This has been both necessary and fruitful particularly in areas that are focused on data analytics (such as video analysis, image analysis, network analysis, etc.), as opposed to the more traditional simulation and modeling done by scientists and engineers. XSEDE's Novel and Innovative Projects group includes a specialist in digital humanities, as

well as specialists in pertinent related areas such as big data analytics.

Along the same lines, in January 2014, Compute Canada hired a full-time Digital Humanities coordinator to lead a national team supporting researchers wanting to engage with national advanced research computing (ARC) resources. Compute Canada, a national non-profit organization incorporated in 2012, plans and oversees pan-Canadian ARC resources used for big data analysis, visualization, data storage, software, portals and platforms for research computing serving Canadian academic and research institutes. The national support team for Compute Canada now consists of the full-time DH coordinator, and a large geographically dispersed team that meets bi-weekly to coordinate national initiatives like training at the Digital Humanities Summer Institute and national competitions such as the recent partnership between Compute Canada and the Canadian Social Sciences and Humanities Research Council. The DH team also meets locally with humanities researchers at their own institutions to help them leverage national infrastructure, and shares experiences and advice back to the national team to help in developing tools, services, and training opportunities that will benefit the national DH community (Simpson 2015).

One of the interesting ramifications of these changes in IT staffing trends is the addition of humanists to IT-based groups in high performance computing, cyberinfrastructure, visualization, and data architectures--humanists who are called upon to maintain both a deep understanding of computational systems, as well as track the needs of scholars in the humanities, which tend to be quite different from researchers in the “hard” or social sciences. Often, they are the lone humanist in a group dominated by engineers and computer scientists. There is a clear need and desire to expand the use of computational resources into "non-traditional" (i.e., non-hard-sciences) disciplines, both to justify an institution-wide investment in research computing, and as a way of building institutional capacity for supporting digital humanities scholarship (as described in the forthcoming 2017 ECAR/CNI white paper on institutional digital humanities support). Nonetheless, institutions are struggling with translating their services into comprehensible and relevant offerings for humanities researchers. Finding effective means of supporting these researchers within the traditional model of the research IT group has also been a challenge, given the ways that it differs from library-based support models with which scholars are more familiar. Panelists will reflect on projects they have worked on that successfully bridged the humanities-research computing divide, to the benefit of both groups.

As one example, at Indiana University, a workflow for teaching text analysis with R has been developed that uses web-based Shiny scripts to introduce an algorithm, highly annotated RNotebooks explaining each line of code, lightly annotated RScripts allowing for remixing and adaptation, and, finally, RScripts to leverage multicore environments. The scripts use data both from literature and Twitter, and these tutorials consistently draw the highest attendance in the series DH for Humanists, held throughout the semester. Individual research projects using more sophisticated algorithms such as NER and LDA have also grown out of this project.

As another example, at the University of Chicago, the recently-developed Visual Text Explorer provides a new type of framework for reading texts along with a range of user-customizable analytics, allowing for simultaneous close and distant reading. Other humanities research computing projects include web-based data-driven animated interactive mapping systems, tools for comparative sequence analysis across literary corpora, and automated aggregators of ranges of specific secondary data sources to inform the reading of specific texts/types of texts, all of which require no knowledge of computer programming by the user but nonetheless leverage research computing resources.

At UC Berkeley, 3D modeling work by Near Eastern Studies scholars that uses photogrammetry software running on the high-performance compute cluster has been helpful for testing new Graphics Processing Unit (GPU) nodes in the cluster.

Finally, panelists will provide recommendations for how humanities scholars can translate their research projects in ways that will make them more comprehensible and compelling for institutional research IT groups that may not have a humanist on their support staff.

Panel participants

Quinn Dombrowski is the Digital Humanities Co-ordinator in Research IT at UC Berkeley. Research IT supports research data management, museum informatics, and computationally intensive research across all domains. She is the author of *Drupal for Humanists*

and has an MA in Slavic linguistics and an MLIS from the University of Illinois.

Tassie Gniady is the manager of the Cyberinfrastructure for Digital Humanities Group at Indiana University. The CyberDH group focuses on workflows for text analysis and photogrammetry. She also teaches an Introduction to Digital Humanities course in the Information and Library Science School. Tassie has a Ph.D in early modern literature from UC-Santa Barbara and an MIS from Indiana University.

Megan Meredith-Lobay is the Digital Humanities and Social Sciences Scientific Analyst for the University of British Columbia's Advanced Research Computing Department. She is also part of WestGrid and Compute Canada, the Canadian HPC national infrastructure platform. Megan has a PhD from the University of Cambridge Department of Archaeology in which she explored the early Christian archaeology of Argyll, Scotland using GIS and early online archaeological databases.

Lisa M. Snyder is the director of Campus Research Initiatives for UCLA's Office of Information Technology, and manager of the GIS, Visualization, and 3D Modeling group for the Institute for Digital Research and Education. She has a Ph.D. in Architecture and teaches Virtual Reality and 3D Modeling in UCLA's digital humanities program.

Jeffrey Tharsen is Computational Scientist for the Digital Humanities at the University of Chicago where he is the lead technical domain expert for digital and computational approaches to humanistic inquiry. Jeffrey has a Ph.D. from the University of Chicago's East Asian Languages & Civilizations department, specializing in the fields of premodern Chinese philology, phonology, poetics and paleography.

Bibliography

National Science Foundation. (2014) "Advanced Cyberinfrastructre - Research and Educational Facilitation: Campus-Based Computational Research Support." https://www.nsf.gov/awardsearch/showAward?AWD_ID=1341935

Neeman, H., et al. (2016) "The Advanced Cyberinfrastructure Research and Education Facilitators Virtual Residency: Toward a National Cyberinfrastructure Workforce". In *XSEDE16 Proceedings of the XSEDE16 Conference on Diversity, Big Data, and Science at Scale*. ACM: New York. DOI: 10.1145/2949550.2949584

Simpson, J. (2015). "Building Support for Digital Humanities". *Compute Canada* blog. https://www.computecanada.ca/blog/building-support-for-digital-humanities/

Wilkins-Diehr, N., et al. (2016)"An Overview of the XSEDE Extended Collaborative Support Program" in *High Performance Computer Applications*, ed. Isidoro Gitler and Jaime Klapp. Springer Link: 2016. http://link.springer.com/chapter/10.1007%2F978-3-319-32243-8_1