

White Paper on Canada's Future DRI Ecosystem

Gaps in Digital Research Infrastructure for Canadian Digital Humanities Researchers

Co-authors: Lawrence Evalyn (English), Elizabeth Parke (OVPR, UTM and Visual Studies), Patrick Keilty (Information), Elspeth Brown (History, Director of the Digital Humanities Network) -- University of Toronto

Signatories:

Ishtaque Ahmed (Computer Science)	Claire Battershill (Information/ English)
Laurie Bertram (History)	Alexandra Bolintineau (Centre Medieval Studies/ Woodsworth College)
Mark Campbell (Arts, Culture & Media / Music)	Heidi Bohaker (History)
Jill Caskey (Chair, History of Art)	Beth Coleman (ICCIT/Information)
Kevin Coleman (History)	Costis Dallas (Information)
Trevor Deck (University of Toronto Libraries)	Wendy Duff (Dean, Information)
Alexandra Gillespie (Vice-President and Principal/ English)	James Ginther (Centre for Medieval Studies & University of St Michael's College)
Yi Gu (Arts, Culture & Media /Art),	Adam Hammond (English)
Timothy Harrison (Near & Middle Eastern Civilizations)	Pamela Klassen (Chair, Religion)
Alison Keith (Director, Jackman Humanities Institute/ Classics)	Will Kwan (Arts, Culture & Media)
Carl Knappett (Chair, History of Art)	Michelle Murphy (WGSI/ History)
M.E. Luka (Arts, Culture & Media/ Information)	Thy Phu (Arts, Culture & Media/ Information)
David Neiborg (Arts, Culture & Media/ Information)	Scott Richmond (Cinema)
Ajay Rao (Historical Studies/ Religion)	Natalie Rothman (Chair, History)
Seamus Ross (Information)	Timothy Sayle (History)
Phil Sapirstein (Art/ Archeology Centre)	Christine Stewart (University of Toronto Libraries)
Naomi Seidman (Chancellor Jackman Professor of the Arts)	Nick Terpstra (History)
Lawrence Switzky (English)	Chris Young (University of Toronto Libraries)
Tamara Walker (History)	

1. Current Issues in Digital Humanities Research

Humanities scholarship has been transformed by digital and computational technologies. Like other eras of historical transformation defined by technology, such as the shift from the scroll to the codex or the invention of moveable type, we live in an era where new information technologies have reshaped the methodologies, research design, and questions of humanities inquiry. Digital Humanities (DH) is an interdisciplinary field that transforms humanities scholarship by bringing digital tools, platforms, databases and other informational structures into dialogue with the questions and approaches central to humanities research. We take a “big tent” approach to the field of DH and here define it as:

1. interdisciplinary research and teaching that brings digital tools, platforms, databases and other informational structures into dialogue with the questions and approaches of humanities research;
2. a critical praxis that emphasizes collaborative, team-based projects that engage in the building blocks of digital activity, such as archiving, curation, analysis, coding, editing, visualization, mapping, modelling, versioning, prototyping, and failing;
3. interpretive or theoretical work on digitality itself.

DH research projects bring together research teams that combine undergrads, grad students, postdocs, faculty colleagues, community partners, and international collaborators. Due to the complexity of these projects, no one scholar can solve the intellectual challenges posed by the research; cross-disciplinary collaboration is required, bridging humanities scholarship and computational and information science expertise. Many DH projects function as cross-divisional humanities “labs.” Examples of just a few based at the University of Toronto include: Michele Murphy’s Environmental Data Justice Lab; Heidi Bohaker’s work with GRASAC, the [Great Lakes Research Alliance for the Study of Aboriginal Arts & Cultures](#); The Records of Early English Drama (REED) Project, and others. These DH-lab based projects make corporate pollution on Indigenous lands visible and actionable; create deeper understandings of Great Lakes arts, languages, identities, territoriality and governance; and map the progress of theatre companies across late-Medieval England.

Digital humanists are central to cross-disciplinary partnerships and collaborations with computer scientists, archivists, physicists and microbiologists addressing our most pressing research questions. By virtue of our training, humanities scholars bring a deep knowledge of the past—the long record of human history—to contemporary problems, whether climate change, surveillance capitalism, or global pandemics.

DH data may be “born digital” (such as a video games, oral history interviews, or Instagram posts) or may have started out as analogue data which has been subsequently digitized as part of the project. Important forms of “data” in DH research include films; μ CT scans; audio recordings; oral history testimony; high-resolution images; 3D models; corpora of encoded texts; maps; architectural drawings; software; and other formats which are particular to humanities objects of study. One project, for example, is Thy Phu’s *Family Camera Network*, a digital archive of 17,000 photographs, 60 albums, 37 home videos, almost 800 objects, and 42 oral history interviews that provides a unique resource for teachers, historians, and scholars to write new intersectional histories of photography, family, and Canada. Another project, Tim Harrison’s [Computational Research on the Ancient Near East](#) (CRANE) project, uses large-scale computer models and simulations of ancient social groups to reconstruct ancient history. DH researchers often work with many disparate and idiosyncratic kinds of data which may not be reducible to simple plaintext files. The data that emerges from DH project require both preservation and access. DH research presents several “big data” challenges: significant volume, varied in format, varied in quality, and contains social and cultural value. In addition, it often requires digitization at the outset of the project with digital preservation and maintenance at the end of the project.

1.1 DH requires human centric infrastructure

Like STEM research, DH digital research infrastructure hinges on subject expertise—not only in the data sciences, but also in the humanities. Humanities expertise is essential to DH projects, especially in research design, qualitative methods, and posing research questions central to humanities inquiry. Humanities scholars bring questions of meaning, power, historical change, ethics, social analysis to the formulation, execution, and archiving of DH research (data and outputs). For instance, without carefully curated OCR data created by domain specific paleographic scholars, OCR of medieval manuscripts, non-Latin scripts, or pre-Industrial printed materials is impossible. Moreover, the pressing concerns around ethical AI and the racial bias of data sets underscores the need for discipline specific training and attention to human-centered, critical research data infrastructure.

2. Current Issues: Major DRI Gaps for DH Researchers

2.1 Existing Canadian funding agencies will not fund DH infrastructure

The most foundational infrastructural gap faced by DH researchers is ineligibility for existing federal funding. SSHRC grants, which are meant to support humanities research, explicitly disallow funding for infrastructure. NSERC, which allows funding for infrastructure, will not fund humanities research. The gap between these funding agencies contrasts sharply with federal funding in the United States, where the National Endowment for the Humanities, National Science Foundation, the Council on Library and Information Resources, Library of Congress, and the Institute of Museum and Library Services allow funding for DH research, including necessary DH infrastructure. American researchers can also rely on private foundations like Mellon, Ford, and Ahmanson. For example, the SSHRC Insight Grant excludes, among other things, projects whose primary objective is “digitization of a collection, or creation of a database” -- both of which are common, and valuable, forms of DH work.

Some of our large-scale, multi-partner DH projects span funding portfolios, thereby never fitting comfortably into one program. DH projects can often ask big questions that cross divisions, bringing engineers and scientists with humanists in collaborative teams. For example, Alexandra Gillespie’s research into the surfaces, structures, and previously hidden contents of ancient books requires collaboration with 10 international research clusters, including conservation scientists who use multi-spectral imaging (MSI), Fourier transformation infrared spectroscopy (FTIR), fiber optic reflectance spectroscopy (FORS), and other non-invasive techniques to measure changes in book samples. These collaborations between humanists and scientists are not fundable in traditional tri-council funding envelopes. Canada needs dedicated funding for long term preservation and archiving of DH projects to maintain its global research edge.

2.2 Specific infrastructural support is needed for the ends of projects

DH researchers are developing new best practices to reduce the burden of long-term maintenance costs. [The Endings Project](#) at the University of Victoria is creating tools, principles, policies and recommendations for digital scholarship practitioners to create accessible, stable, long-lasting resources in the humanities. They have developed tools and best practices for new projects to be designed for sustainability, and for existing projects to be redesigned as static sites which require far fewer resources to maintain. Even this work, however, only *reduces* the burdens of long-term storage and maintenance, and does not eliminate it. Indeed, converting an older project to a more sustainable format is itself a costly process, for which limited support exists. Investment and resources in this area of research are necessary to convert these legacy projects into a format that could integrate these projects into a library’s developing DRI (e.g., Dataverse Canada) so that the shared public investment in these research endeavors is not wasted.

2.3 Existing platforms do not support the technical demands of DH data

Data management and research software infrastructure is often unprepared for the forms that DH data takes. As one challenge often posed by DH data, often the total size of a DH dataset may be smaller than, say, satellite data for research in astrophysics, but the size of each individual file is much larger than individual datapoints generated by science research. A digitized film, for example, must exist as a large file, presenting unique challenges for data storage and data transfer. As another common challenge, when DH researchers wish to study historical software, this often requires emulation. Projects at the [Canadian Centre for Architecture \(CCA\)](#), [Bitcurator](#), and [Yale's Emulation as Service](#) demonstrate the need for custom digital infrastructure which can be costly, but fundamental to research and ultimately reuse of data. Even when researchers can meet their data management needs internally, they often re-encounter them when making their collections public online. Digital preservation systems don't usually work easily with content management systems that are accessible to the public, necessitating parallel systems and thus duplicating the labour of ingesting digital assets and writing metadata for each data object.

2.4 DH researchers need training and outreach to access existing technical resources

With advances in AI and machine learning, new research software tools are being developed that can aid in analyses of the vast scope and range of DH research data. UTL's adoption of CADRE from the University of Indiana provides one example of the new research tools being developed to support DH researchers. More engagement with Computer Science researchers who focus on this avenue of research is needed, and it would facilitate Canadian development and innovation in this important area of research, with potentially enormous research synergies and social benefit. Access to training in fundamental programming languages/environments (such as R, Java, PYTHON), which is typically available to researchers in non-Humanities disciplines, should be made available for DH researchers, staff and students.

Since humanities researchers are also less likely to have pre-existing familiarity with the range of options for digital research infrastructures, additional training or accessible guides to these resources could improve researchers' access. Existing outreach regarding Canada's digital research infrastructure is often restricted to the sciences. For example, the training at SciNet, the supercomputer centre at the University of Toronto, is not accessible to humanities researchers.

2.5 The burden of digital research infrastructure is placed on individual institutional libraries

Most DH researchers currently meet the project's infrastructure needs through their institutional libraries. Much of this research support is appropriate, and well within the institutional mission of academic libraries. However, providing infrastructure at the institutional level also amplifies inequities of access between institutions, and involves duplicated efforts between universities. Over-reliance on academic libraries for digital research infrastructure has also created a gap in support for the common activity of creating websites for a DH research project, particularly when these websites incorporate interactive elements. Complex web development is not the job of an academic library, but it is crucial to public humanities outreach.

2.6 Necessary staffing infrastructure for DH researchers does not have an institutional home

Currently, the human-centric needs of DH research infrastructure fall between the silos within the university. DH scholars need to work with technical staff, such as DH developers or user experience designers. Currently, however, neither the library nor IT consider positions such as these part of their mandate. Even if such skill sets are represented in the library staffing complement, this expertise is not made available for longer term project support. This kind of staffing is also not provided by individual disciplinary departments and is thus generally patchworked together at the project level, or done without, to the detriment of the research.

For example, the University of Toronto Libraries explicitly exclude high level faculty research project support from their recommended supports for faculty research. In August 2017, the UTL Executive Committee formed

the UTL Digital Scholarship Task Force (DSTF) to make recommendations regarding how best to support digital scholarship at the UTL. Their recommendations concerned the first two of four tiers of digital research support: quick references, consultations, and instruction via workshops. Importantly, the task force explicitly did not recommend tiers 3 and 4 of research data support: Tier 3, Enhanced Research Services in the form of project consultations and customized training; nor Tier 4, High Level Project Support. These two tiers are precisely what DH scholars need to support their research, however. The task force was unable to recommend project support, even if they had the expertise on staff, due to “shifting human resources (due to leave, illness, etc.), staffing limitations, and existing resources.” (Marcel Fortin, Leslie Barnes, Mariya Maistrovskaya, and Adriana Balen, “UTL Digital Scholarship Task Force Final Report,” October 31 2017).

As U of T librarians Laure Perrier and Leslie Barnes [have shown](#), U of T researchers have concerns regarding technological obsolescence, cost, and the uncertainty regarding the policies of commercial vendors when selecting data management tools and products. U of T researchers routinely encounter obstacles when collaborating across universities, with scholars in one location unable to access drives, servers, and files at the other location. Many researchers raised concerns around data security, including worries about servers located outside Canada. While the study concludes by arguing that “the library can play an important role in offering support to academic researchers as they face challenges related to research data management,” it is also the case that the library cannot play this role without infrastructural investment in people, via staffing. In this case, as in many others in this white paper, infrastructural support is most needed in the form of staff/researcher expertise in working with faculty researchers.

3. Looking Ahead: the future of DH DRI

3.1 Create a funding stream for DH projects in Canada

Canada’s investment in important DH research resources should be made less precarious through dedicated grants that fund the transitioning of these resources to library data archiving systems. Additionally, research software tools will need to be developed to make this archived data accessible and searchable by all.

Canada needs to catch up to the US and to Europe in recognizing the specific needs of DH researchers and creating a mechanism to fund these projects in an ongoing manner. Such a funding program would enable long-term project maintenance in both staffing and data infrastructure support. This program should be open to all scholars, regardless of field, and should be designed to support cross-disciplinary research that brings together humanities scholars with researchers in the data sciences, engineering, and other fields that draw on computational methods to address research problems. Ideally, digital humanists would collaborate with digital scholarship librarians in applications to this funding stream, to ensure DH project sustainability. Important research resources could be made less precarious through dedicated grants to fund the long-term maintenance of completed projects with proven ongoing value.

3.2 Support long-term employment models

DH research develops through sustained relationships with GLAM (galleries, libraries, archives, and museums) and community partnerships over time whereas grant and contract bound employment models feed cycles of precarity. Long-term employment models in DH digital research infrastructure ensure projects’ persistence and preservation while at the same time building highly skilled researchers with software and computational skill expertise.

3.3 Recognize the importance of humanities training

DH research relies on the intersection of technical computational skills and humanistic expertise; it cannot be adequately carried out by scholars with only the relevant technical skills. A recent paper published in *Nature Communications*, “Tracking historical changes in trustworthiness using machine learning analyses of facial

cues in paintings,” for example, drew extensive and widespread criticism for its failure to address basic facts of history and art history which rendered its use of machine learning meaningless.

3.4 Support multiple platforms for varying kinds of humanities data

It is not feasible for one centralized platform to serve all kinds of needs. Recent acquisition of data set management and mining tools like [CADRE](#) indicate the need for making data available, and the argument made by [Europeana](#) in a recent post “Never just a back-up: Digitised heritage as a resource for good” underscores the interrelatedness of digitization, exhibition, preservation, and interoperability. (For similar concerns, see also the UTL white paper submitted to NDRIO).

4. How NDRIO Can Help Bridge the Gap

- Recognize that DH research is important, contributes to the social good, and creates strategic, innovative research opportunities for Canada;
- Recognize that the DH create important data resources that have DRI (ARC, RDM, and RS) requirements, and represent a strategic, uniquely Canadian asset;
- Develop and foster innovative relationships between DH researchers (and domain specialists) and Computer Science specialists who develop text, data-mining, and image analysis tools;
- Increased (and sustained) access to DRI resources for Humanities researchers, not just in the form of data storage, but also training in the use of computational tools, and equal access to ARC facilities and services, including the development of new applications;
- Increased staff support dedicated to supporting advanced research computing and the development of new applications in the Humanities, including in the regional ARC facilities;
- Increased (and sustained) accessibility of training for Humanities researchers in data analytics, programming, and advanced research computing at all levels;
- Increased incentives to develop advanced research computing applications in the Humanities;
- Provide support for the transition of older and completed DH projects that do not have the resources to transition their data to formats compatible with the developing digital archiving infrastructure of libraries;
- Provide support for the development of Research Software (RS) tools that facilitate the archiving and stewardship of DH data, and render them accessible online;
- Wide, continuing, and iterative consultation process with scholars in the humanities and social sciences (a continuing iterative process is necessary, because researchers in these fields are likely to lie outside of the regular user-group and regular communications channels for existing bodies like Compute Canada, and thus are the most unlikely to hear about the NDRIO re-organization and participate in the process, which will lead to the reification of unequal access to computing resources);
- Don’t assume that the current users of RDC, CARL, CFI, and Compute Canada include all researchers with digital research infrastructure needs: there are likely DH scholars whose research could be immediately benefitted by NDRIO’s new infrastructure but who currently look elsewhere;
- Develop specific initiatives to increase the visibility of advanced research computing services to humanities and social science researchers;
- Develop strategies to consolidate digital research tools (databases, CMSs, DAMs) so that researchers can work collaboratively on the same platform, at a reasonable cost.