Canada's Future DRI Ecosystem for Humanities & Social Sciences (HSS)

Author and contact: Geoffrey Rockwell (grockwel@ualberta.ca), President, The Canadian Society of Digital Humanities / Société canadienne des humanités numériques

Authors: Matt Huculak and Emmanuel Château-Dutier, with Barbara Bordalejo, Kyle Dase, Laura Estill, Julia Polyck-O'Neill, and Harvey Quamen.

The Canadian Society for Digital Humanities / Société canadienne des humanités numériques (CSDH/SCHN) is a Canada-wide association of representatives from Canadian colleges and universities that began in 1986 as the Consortium for Computers in the Humanities / Consortium pour ordinateurs en sciences humaines. The society is an active member of the international Alliance of Digital Humanities Organizations (ADHO), which promotes and supports digital research and teaching across arts and humanities disciplines around the world.

Current Issues

By definition, Digital Humanities research uses digital tools, services, and research software to explore and research subjects and objects of relevance to the humanities & social sciences (HSS). Digital Humanities is a para-discipline that brings together practitioners from a variety of humanities and adjacent fields; the research questions and approaches are complex and varied. Digital Humanities is significant because it unites scholars in many disciplines who share an interest in exploring humanities questions with the aid of digital tools—in the past, the field was known as "humanities computing."

HSS students and scholars have varied abilities and literacies when it comes to digital tools, yet the research they undertake in the twenty-first century is predicated on these very tools: library search algorithms that privilege certain results over others; digitized texts and multimedia objects that circumscribe what is findable and researchable.

While many HSS researchers might not, at this point, have access to advanced computational tools or expertise, their potential future needs should be considered by Canada's DRI plan. Providing DRI training and infrastructure for the next generation of HSS scholars at both undergraduate and graduate levels is a high priority for the field. In short, we represent fields on the other side of the campus that do not yet have the skills necessary to take full advantage of DRI. With proper outreach, adding HSS disciplines could roughly double users on the national ARC platforms.

Needs Summary

- Systems to engage with small and large corpora of text & images
- Remote collaborative work and infrastructure for research dissemination
- Training new generations of scholars in computational methods
- Infrastructure on which to do that training
- National team of consultants & developers to use ARC platforms
- Infrastructure for long-term access and preservation of work

What are the main DRI tools, services and/or resources you currently use in your research?

The most popular DH tools on Compute Canada infrastructure are Voyant (voyant-tools.org/), a web-based reading and analysis environment for digital texts that has had over 1 million visitors since 2018; the HSS Commons; and CWRC, the Canadian Writing Research Collaboratory (cwrc.ca), a virtual research environment for literary and cultural studies that in 2020 served close to 7,000 unique users with

access to more than 395,000 cultural objects. Many digital humanists use Portage's Data Management Plan assistant (https://assistant.portagenetwork.ca/) to plan for the lifecycle of their data. These are projects that require little preliminary training to use.

Despite the success of these tools and services, Canadian digital humanists are not always using the current Canadian DRI: they are turning to paid cloud services, foreign hosting, or creating bespoke tools and projects. Few HSS researchers are using ARC infrastructure to its full potential. *The low use of the existing Canadian digital infrastructure among the DH community is partly due to a lack of knowledge of the services offered, but also to the fact that they do not meet the full needs of humanists.*

What works very well in the current state of DRI tools is that any faculty member or librarian at a university/college across Canada has access to platforms and services maintained by a national team. It is essential that these services continue and are augmented by support (consultation and development).

Humanists need Platforms as Service (Paas) which would provide infrastructure for research managed by professionals in a non-piecemeal fashion. More importantly, humanists need consulting about what platforms are available and how to use those platforms as projects begin. Regular check ins, back and forth, over the course of the project is necessary to see if the tools are being used to their full potential; ideally, we would have access to developer time for project programming based on the needs of individual projects that can then be shared back into the larger community.

Humanities data is often based on cultural artifacts and the preservation of both the items themselves as well as historical contexts/meanings/arguments of those objects (the paradigm of interpretation). Knowledge mobilization in the form of sharing data and interpretation is key to achieving the goals of many projects. In contrast to STEM data, DH data is highly curated and metadata rich which, in turn, might be used by different scholars to answer their own research questions in other fields. For digital humanists, the challenge of long-term data-sharing is central to current discourse as we strive to live up to standards and principles (such as FAIR and CARE) established by other fields.

Digital humanities projects range in scope from boutique to large: there are projects focused on a single text or object and others that engage with big data and analytics. The challenges of smaller projects are just as important as those of larger ones but rarely do they have access to similar resources: A lack of infrastructure and human support can mean that projects struggle to implement or extend existing standards in ways that support interoperability and reuse. Grant financing currently rewards "new" projects; rarely are projects able to hire personnel over years or undertake much-needed project updates. The grant cycle leads to data preservation issues where data, analysis, and the technical expertise required to extend past work are lost to future generations of researchers.

Support to digital humanities researchers is uneven across the field, and often depends on the researcher's home institution: is there a digital humanities center that offers web hosting and sysadmin support? What will the university's IT department support? Does the library actively support data management and management? Having a national support team for HSS scholars for training and consultation on national platforms would be an incredible development for researchers.

Needs Summary

- Need places to store our data for long-term preservation and access
- Need meaningful human support on projects from beginning to end
- Need standard software preinstalled (or easily accessible VMs)
- Need to make it easier for researchers to build more tools like Voyant, CWRC, etc., for the future of HSS research.

Do you have access to all the DRI tools, services and/or resources you need for your research? What are they? What is missing?

No. Because DH projects are so varied in their approaches and needs, the current systems are not sufficient for the needs of researchers. In many cases, a DH project might need something as simple as a database, or a long-term knowledge dissemination platform like a CMS. Because the costs of maintaining and upgrading database-dependent systems are so great, very few researchers have access to something as simple as a website to share their data through departments or libraries. Moreover, humanists tend to produce data that are viable/useful for study over the long term, so a long-term commitment to website hosting is needed. Other DH scholars prototype and design software as an argument within itself; for example, see Galey and Ruecker's "How a Prototype Argues".[1] Moreover, the temporary and unstable position of graduate students and early career scholars means they often lack a longstanding home for their projects and are at risk of having no infrastructure or support at all.

Individual digital humanities projects often pay monthly or yearly fees for basic digital research infrastructure, such as cloud hosting and servers. (Some institutions, for instance do not even provide their faculty with programs to manipulate a pdf file.) If NDRIO met these basic needs, grant funds (often from Tri-Council sources) could be used in better ways.

Needs Summary

- Easy, personalised web environnement to deploy web applications, service should be offered with maintenance, and security updates (sysadmin services).
- Software solutions for long-term web hosting for scholarly communication
- Sandboxes and prototyping--particularly useful for graduate training
- National, shared repository for data access, preservation, and dissemination that can also provide interpretive structures/critique of those data.

What are your biggest challenges accessing and using the DRI tools, services and/or resources that do exist and are available to you?

The biggest challenge facing HSS scholars is not access to tools; rather, it is access to service. **Our community needs servers AND services.** We have found access to technology to be readily available (ComputeCanada); however, in existing DRI models, once the infrastructure is provided, a new problem arises in training in how to use those tools. There is a need for tech onboarding and training in how to use those tools that are so necessary within the epistemological contexts of humanist inquiries. Moreover, there is also a need for long-term commitments to hosting those services. In 2015 it was calculated that 50% of projects presented at the flagship Digital Humanities conference, DH2005, were no longer available.[2]

Needs Summary

- Service support from professionals who are equipped to train humanists in using supported technologies.
- Tech onboarding: human to human training on how to use systems.
- Project Advisors: we need collaborators to participate in the initial architecting and ongoing operational oversight of projects from conception to implementation.

---

[1] Alan Galey and Stan Ruecker, "How a Prototype Argues," *Literary and Linguistic Computing* 25, no. 4 (2010): 405-424, doi: 10.1093/llc/fqq021.

[2] https://robincamille.com/presentations/mla2015/

- Long-term commitment (at least 10 years) to grant-funded projects.

Future DRI State

What is your vision for a cohesive Canadian DRI ecosystem that would fulfill your research needs?

A cohesive Canadian DRI ecosystem would provide infrastructure and services to scholars throughout the research lifecycle, from data creation, curation, to long-term preservation.

It would see Infrastructure as a service with universal accessibility to tools and support for researchers at every stage of their careers; it would also integrate closely with granting systems that allowed for personalised yet scalable services.

Needs Summary

- Accompagner le cycle de vie des données
- Better integration with the grant system so that access to computing resources (including human services) are articulated during the grant writing process so that “ready-to-go” solutions are already conceptualized should the grant be successful. DRI & support should be a foundation of the grant-writing process.

What are the types of DRI tools, services and/or resources you would like to use, or envision using, in the future?

For digital humanists, having highly qualified personnel to support projects is paramount: this includes sysadmin support that allows projects to focus on the development of web applications that are directly related to their research. We strongly believe this kind of support can be offered at a federal level if the infrastructure can develop cloud services to deploy and maintain personalised servers and VMs. Along with web hosting services for research, specialized services could also be proposed to deal with content that humanists (but not only) are often working with (e.g. text, images, videos, sounds).

One interesting model to follow could be the Huma-Num model (France): this includes research data hosting, staff dedicated to supporting digital humanities projects, and standardized research environments (such as virtual machines) that can be personalized and deployed on demand. Some basic tools that humanists need and often have to purchase from grant funds include speech recognition service and captioning; video editing suites; optical character recognition (OCR), including handwriting recognition (HCR); document management and bibliographic management (like Zotero); computer vision and tools to run and train artificial intelligence data models. File sharing, social research tools, and version control (some of which can be currently done via GitLab or Radicle).

Needs Summary

- The infrastructure needs of HSS researchers fall into two categories: 1. the physical infrastructure needed to produce and disseminate research. 2. and the support needed to train the users and actors of the community
- Humanities researchers (but not only them) need dedicated online tools and services to process text, images, videos, sound.

What challenges do you foresee while using integrated DRI tools, services and/or resources?

In addition to the hosting of web applications and the provision of specialised tools, we believe that the creation of the new infrastructure by bringing together the areas of expertise previously provided by three separate agencies for advanced computing, research data management and research software, provides

an extraordinary opportunity to foster integrated approaches in these three areas. In this respect, the infrastructure could develop services to support open science practices by accompanying the data life cycle and following the FAIR model. It must therefore propose work chains based on rich metadata supported by standards as well as tools focused on the discoverability, publication and dissemination but also the perennial archiving of research data. These solutions must be integrated or designed in collaboration with initiatives currently underway in Canada in the research library sector.

Needs Summary

- An agile technological platform that responds to the variety of workflows/pipelines used by humanists in their work coupled with on-the-ground human support.
    - Flexibility in stitching together services
- Solutions should be developed to support open science, FAIR principles, data management and long term preservation of research data
- Services should be articulated with other canadian initiative for data management and long term preservation of research data

How to Bridge the Gap

What are the tools, services and/or resources NDRIO should leverage to achieve your desired future state?

ComputeCanada has officially provided important infrastructure that has revolutionized teaching and research in Canada (especially for SSHRC-grant holders). Unofficially, the ComputeCanada team has provided training, tutorials, and step-by-step help to assist humanists with technological challenges.

How do you see NDRIO's role in addressing current gaps in the national DRI ecosystem?

NDRIO is in an unique situation to unite piecemeal services and infrastructure in Canada and to provide resources in an equitable way across disciplines in Canada. As Ray Siemens notes, humanists tend to use technology and computers differently than other disciplines.[3] Because of this methodological approach, humanists have not had access to the same resources as the hard sciences. NDRIO can ensure equitable access to resources while recognizing the diversity of approaches to technology. In departments that do not have a tradition of computational research such as the humanities, this is essential for supporting new scholars. Finally, NDRIO could be a nexus of cross-disciplinary work among colleagues. For example, in the field of AI, humanists can contribute to the human questions of technology--such as the prejudices that are replicated in AI environments.

Needs Summary

- Equitable access to technology services that acknowledges the diversity of disciplines in Canada
- Easy-to-use, cloud-based services that can be requested by Canadian scholars at any stage in their career (students to advanced researchers)
- Provide a way for cross-disciplinary work so that important ethical questions are considered in computational research

What other suggestions do you have?

A need to enshrine values of diversity and accessibility. There is a great need to recognize that data exists in a political and human world. NDRIO should have a diversity, equity, and accessibility working group to address issues in DRI.

---

[3] Hong, Joo-Wha, and Dmitri Williams. 2019. "Racism, Responsibility and Autonomy in HCI: Testing Perceptions of an AI Agent." *Computers in Human Behavior* 100 (November): 79–84. https://doi.org/10.1016/j.chb.2019.06.012.