Scales of Data Sharing Challenges in the Digital Humanities

Lindsay Poirier
Science and Technology Studies, Rensselaer Polytechnic Institute
Research Data Alliance Data Share Fellow

Abstract

As part of the RDA Data Share Fellowship, I will spend the next six months examining the barriers to adoption of RDA outputs in the digital humanities. Based on early research, in this poster, I elaborate a framework, designed by Kim Fortun, for analyzing the scales that buttress contemporary problems. I show how the challenges that DH researchers face as they work towards adopting RDA outputs are the result of interactions between issues at many scales. In this sense addressing the barriers to output adoption in the digital humanities is not just about developing more appropriate technologies, nor is it just about providing the humanities with more technical support. It will require an agenda for tackling issues at many different scales.

Scales

Meta

- How do we build discourse around the promise of open science – particularly at a time when global support for research itself is under attack in many settings?
- How do we build discourse that encourages investment in the development and maintenance of research infrastructure?

Macro

- How do we ensure that data sharing practices work within the parameters of existing laws – such as copyright and requirements for human subjects research?
- How do we ensure that appropriate financial structures are in place to support and sustain the development and implementation of data infrastructure?

Meso

- How do we ensure that appropriate incentives are in place to encourage researchers to participate in data sharing and infrastructure development?
- What governance structures best support collaborative research projects?
- What sorts of support systems and organizations are needed to both enable and encourage collaborative work?
- How do we network diverse national and interdisciplinary efforts to build data infrastructure with each other and local data sharing communities?

Micro

- How does data sharing change the research workflow?
- How do we build data workflows that align with diverse research methods and practices?
- How do we prepare researchers to work with data infrastructure – to integrate data best practice into their day-to-day work?

Techno

- How do we select which standards and technologies get enlisted in shared data infrastructure, or in other words, which become "best practice"?
- How do we network these technologies in ways that work for interdisciplinary communities?
- How do we factor the time to build technologies into research projects?
- How do we ensure there are (affordable) places to put new technologies and data?
- How do we ensure that the infrastructure (Internet access, bandwidth, and processing power) needed to support data sharing technologies are accessible to DH researchers?

Data

- How do we design data architecture to protect the diversity of analytic modes and thought styles across research traditions?
- How do we design and structure data architecture and configurations to fairly represent information?

Nano

- How do we convince humanists that data-sharing is a worth-while endeavor, particularly in fields that are technology-averse or tend to work in isolation?
- What educational programs would orient the next generation of DH researchers towards seeing data sharing as, not only a valuable endeavor, but also an imperative one?
- How can researchers best deal with the paradoxical need for research infrastructure customized in keeping with their own research tradition, yet with enough standardization to allow research sharing among different communities?

Digital DH researchers' Recommendation to RDA

- For each output, RDA instate one traveling consultant, responsible for actively working with research groups and helping to translate the technical aspects of the output into terms that could be more broadly understood.
- Develop a registry with a vetted pool of technologists, modeling DHCommons, that DH researchers could leverage whenever they got a bit of money.
- Develop more user-driven tools for implementing RDA outputs, such as a tool for automatically implementing PIDs
- Produce appendices to all outputs listing and describing supporting services and organizations in different geographic contexts.
- Create a space on the Web platform where organizations that have implemented outputs can post implementation notes, describing services they either used or developed.
- Develop model documents with language that could be framed for governments, administrators, etc to convey the value of building digital infrastructure for the humanities. The RDA could vet such documents. *Based on data collected at the RDA/ADHO Workshop at DH2016
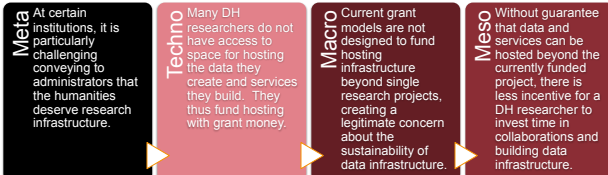
Further Research

How can RDA tap into existing DH support venues to promote output dissemination and adoption? Which proposals should be put forward?
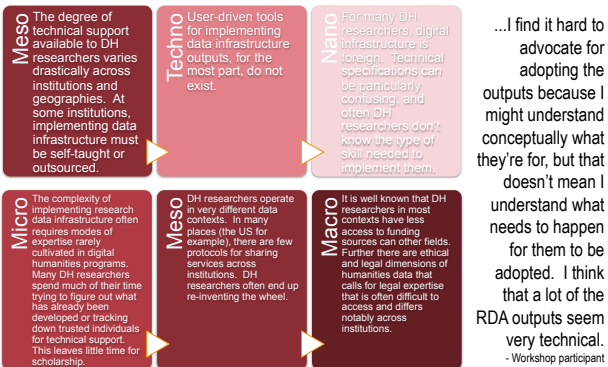
Open Questions

Does leveraging this framework to understand the barriers to data infrastructure adoption change how we approach solutions?

Acquiring Sustainable Hosting Infrastructure

Meta — At certain institutions, it is particularly challenging conveying to administrators that the humanities deserve research infrastructure.

Techno — Many DH researchers do not have access to space for hosting the data they create and services they build. They thus fund hosting with grant money.

Macro — Current grant models are not designed to fund hosting infrastructure beyond single research projects, creating a legitimate concern about the sustainability of data infrastructure.

Meso — Without guarantee that data and services can be hosted beyond the currently funded project, there is less incentive for a DH researcher to invest time in collaborations and building data infrastructure.

Support for Implementing Outputs

Meso — The degree of technical support available to DH researchers varies drastically across institutions and geographies. At some institutions, implementing data infrastructure must be self-taught or outsourced.

Techno — User-driven tools for implementing data infrastructure outputs, for the most part, do not exist.

Nano — For many DH researchers, digital infrastructure is foreign. Technical specifications can be particularly confusing, and often DH researchers don't know the type of skill needed to implement them.

Micro — The complexity of implementing research data infrastructure often requires modes of expertise rarely cultivated in digital humanities programs. Many DH researchers spend much of their time trying to figure out what has already been developed or tracking down trusted individuals for technical support. This leaves little time for scholarship.

Meso — DH researchers operate in very different data contexts. In many places (the US for example), there are few protocols for sharing services across institutions. DH researchers often end up re-inventing the wheel.

Macro — It is well known that DH researchers in most contexts have less access to funding sources can other fields. Further there are ethical and legal dimensions of humanities data that calls for legal expertise that is often difficult to access and differs notably across institutions.

> ...I find it hard to advocate for adopting the outputs because I might understand conceptually what they're for, but that doesn't mean I understand what needs to happen for them to be adopted. I think that a lot of the RDA outputs seem very technical.
> - Workshop participant

Adapting Data Infrastructure to DH Methods and Theories

Nano — Often DH researchers bring different theories, methods, and assumptions to their work than those designing digital infrastructure.

Data — The architecture of digital tools, platforms, and methods can be out of sync with humanistic assumptions, methods, and theories.

Micro — With data architecture out of sync with humanistic inquiry, it can be particularly challenging for DH researchers to map their workflows onto digital infrastructure.

Meta — With little language for how digital infrastructure can be adapted for the humanities, it becomes difficult to advocate that the humanities deserve digital infrastructure or for more traditional humanists to see its value.

Venues for Pushing Outputs to the Digital Humanities Community

DiRT for indexing tools
- Directory of research tools for digital DH researchers
- Developed out of US-based Project Bamboo
- Allows researchers to track and compare tools
- http://dirtdirectory.org/

DHCommons for collaborators
- Platform designed to connect DH projects with skilled collaborators
- Matching service for DH collaboration across institutions
- Under the centerNet initiative, the platform also supports face-to-face meet-ups, and sponsors micro-grants
- Publishes a journal for peer-review of in-progress DH work
- http://dhcommons.org/

DARIAH for consultancy and services
- European infrastructure offering technologies and services tol DH researchers (such as PID and data management services)
- One working group working to establish certification for trusted digital repositories
- Also offers a course registry and scholarships

NEH (and others) for financial support
- Offers start-up and implementation grants for DH projects
- Other prominent funders include the Andrew W. Mellon Foundation, Humanities in the European Research Area, and the European Research Council

ADHO for presenting and publishing
- Sponsors the annual DH conference
  - Hosted a RDA-themed panel and workshop at DH2016
- Publishes the DHQuarterly Journal