Data Problems in the Humanities, or "When everybody is special, no one is"?

Woods, Nathan D.
nathan.woods@uleth.ca
University of Lethbridge, Canada

Bordalejo, Barbara
barbara.bordalejo@uleth.ca
University of Lethbridge, Canada

O'Donnell, Daniel Paul
daniel.odonnell@uleth.ca
University of Lethbridge, Canada

That the humanities has ‘a data problem’ is now a common refrain amongst many communities. Humanists often argue that humanities data is a problem because they don’t have or work with data (Borgman, 2010; Posner, 2015). Librarians and information professionals, by contrast, believe that humanists have data, but assume they don’t realize it — meaning that the problem is that they must be trained to appropriately recognize and work with data they most surely have (Flanders, Julia, and Trevor Muñoz, 2012; Ikeshoji-Orlati, Caton, and Stringer-Hye, 2018). Digital humanists know that they (at least) have data, but believe that their data are special, distinct both from those used by scientists and (often) those “not recognized” by traditional humanists — and that these data require special strategies and techniques as a result (Drucker, 2011; Schöch, 2013). In a rough and general way each premise informs a mélange of assumptions, advice and best practices that comprise the emerging literature on research data management (RDM) in the humanities.

In this paper we argue that this focus on the discovery and definition of what is “special” about the definition or recognition of humanities data is a mistake. Humanities data are not special because of what they are, but rather because of how they are used (Borgman, 2017; Leonelli, 2015). Data are data, in other words, whether they are produced and used by scientists or humanists. The “problem” with Humanities data — and the thing that makes them special — lies in the use-case. Humanities data are both designed and structured by systems to meet scholarly or systematic ends. While many humanities researchers engage, as users, with designed research infrastructure, humanities scholars also innovate, design, and build scholar-led research infrastructure that have specific systems requirements, but often without clear requirements modeling. We draw out the implications of this argument on two fronts, looking at the issue of research, and the implications for humanities infrastructure development.

First, we report on the preliminary results of ongoing research examining how humanists think about data as they build, navigate, and utilize research infrastructure for scholarly purposes. Reported results derive from a series of comparative case studies (Eisenhardt, 1989; Maryl et al. 2020) examining four cases of scholar-led projects that deploy digital research infrastructure for scholarly ends. Cases were purposively selected using the following criteria: a. Are scholar-driven; b. Involve representation of ‘real world’ objects involving small, intensively curated datasets; c. Have a documented record of accomplishment and innovation.

Although humanists have a long history of designing and building research infrastructure, both analog and digital (Abbott, 2012; Bod, 2013) this has been less appreciated in the literature on humanities infrastructure design. By comparing elements from our case studies, we build on previous studies of humanists as both users of information resources (Bates, 1996a; 1996b; Bates et. al, 1993; Buchanan, Gow, Blandford, Rimmer, and Warwick, 2006; Palmer,Teffeau,Pirmann, 2009; Stone, 1982; Rimmer, Warwick, Blandford, Gow, and Buchanan, G., 2008; Watson-Boone, 1994), and digital tools (Gibbs and Owens,2022) to consider issues of use-case construction, both formally and informally, over time.

Emphasizing the role of scholars as designers (Lamb and King, 2003; Millerand and Baker, 2010), we highlight the importance of the ‘use-case’ as a means to culturally frame and model system requirements, functional and non-functional, as a type of socio-technical fact, mediating between user-agency, the purpose of a scholarly project, and the design, development and maintenance of the research infrastructures innovated to meet those goals. Originating in software engineering, use-case modeling (Jacobson et al. 1992) is a means of specifying, validating, and eliciting system requirements, where use case models describe, communicate, and facilitate all of the ways a user interacting with a system or product may work to realize a desired end. While there is a robust literature on how to write and design use cases (Kulak, Daryl, and Guiney, 2012) there is little empirical study of use-case development or implementation in software engineering (Anda,2003), and nothing that systematically examines the diffusion of the concept outside these arenas of expert work.

By comparing the roles of humanist researchers as designers, we complicate previous research on humanities work practices (Blanke and Hedges, 2013; Pachecho, 2022; Unsworth, 2000) and studies of interdisciplinary scholarly information practice more broadly (Palmer, 2013), drawing on the notion of ‘infrastructural practices’ (Baker and Millerand, 2016; Baker, Duerr, and Parsons, 2015; Edwards, 2019; Karasti and Blomberg, 2018) in scholarly information ecosystems (Altman and Cohen, 2022), and hence on the dual role of scholars as infrastructure designers-builders and users (Lamb and Kling, 2003; Millerand and Baer, 2010, Mongili, 2014).

While we acknowledge a few exceptions for specific tools and technical standards for the Digital Humanities (e.g. Voyant, TEI, IIIF, CIDOC CRM), the Humanities generally lack the proliferation of a robust, distributed, yet centrally networked, ecosystem supporting generalizable infrastructure development process, linking standards with the infrastructure practices central to scholar originated infrastructure projects (c.f. Bosman and Kramer, 2015). Rather, humanities researchers have worked far more improvisationally (Ciula, 2022), and in the absence of a systematic orientation to project requirements, have instead created custom knowledge infrastructure. As a result, the identification of discipline-wide standards for tools and infrastructure is far more difficult.

From the perspective of infrastructure design, development, and maintenance, therefore, the “problem of Humanities data” is how to support and interface with scholarly use-cases, the scenarios and problem sets scholars are concerned with, and engage with the custom infrastructural strategies they have developed to speak to them. In our conclusion, we point to some ways where a focus on use case modeling and system requirements planning can complement efforts to refine project development, particularly with refe-

rence to recent work on ‘data communities’ (Cooper and Springer, 2023) and community-oriented frameworks (Lyon, et. al 2012; Jeng and Oh, 2016; Jeng and He, 2022) to bring researchers and infrastructural partners into structured dialogue regarding requirements development.

Bibliography

Abbott, Andrew, 2011. "Library research infrastructure for humanistic and social scientific scholarship in the twentieth century." Social knowledge in the making, pp.43-88.

Altman, M., and Cohen. P.N. ,2022. "The scholarly knowledge ecosystem: Challenges and opportunities for the field of information." Frontiers in Research Metrics and Analytics 6 85.

Anda, B . 2003 . "Empirical studies of construction and application of use case models." University of Oslo, Norway.

Baker, K.S. and Millerand, F., 2016. Infrastructuring ecology: Challenges in achieving data sharing. In Collaboration in the new life sciences (pp. 133-160). Routledge.

Baker, K.S., Duerr, R.E. and Parsons, M.A., 2015. Scientific knowledge mobilization: Co-evolution of data products and designated communities. International Journal of Digital Curation, 10.2

Bates, M.J., 1996. Document familiarity, relevance, and Bradford’s law: The Getty online searching project report no. 5. Information Processing & Management, 32(6), pp.697-707.

Bates, M.J., 1996. The Getty end-user online searching project in the humanities: Report no. 6: Overview and conclusions. College & Research Libraries, 57(6), pp.514-523.

Bates, M.J., Wilde, D.N. and Siegfried, S., 1995. Research practices of humanities scholars in an online environment: The Getty online searching project report no. 3. Library & Information Science Research, 17(1), pp.5-40.

Blanke, T. and Hedges, M., 2013. Scholarly primitives: Building institutional infrastructure for humanities e-Science. Future Generation Computer Systems, 29(2), pp.654-661.

Bod, R., 2013. A new history of the humanities: The search for principles and patterns from antiquity to the present. Oxford University Press.

Borgman, C.L., 2017. Big data, little data, no data: Scholarship in the networked world. MIT press.

Borgman, C.L., 2010. The digital future is now: A call to action for the humanities. Digital humanities quarterly, 3(4).

Bosman, J. and Kramer, B., 2015. 101 Innovations in Scholarly Communication: How researchers are getting to grip with the myriad of new tools. Impact of Social Sciences Blog.

Buchanan, G., Gow, J., Blandford, A., Rimmer, J. and Warwick, C., 2007, June. Representing aggregate works in the digital library. In Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries (pp. 247-256).

Ciula, A. and Smithies, J., 2023. Sustainability and modelling at King’s Digital Lab: between tradition and innovation. In On Making in the Digital Humanities: The scholarship of digital humanities development in honor of John Bradley (pp. 78-104). University College London Press.

Cooper, D. and Springer, R., 2019. Data Communities: A New Model for Supporting STEM Data Sharing . Ithaka S+R. 13 May 2019. Web. 14 January 2023. https://doi.org/10.18665/sr.311396 .

Drucker, J., 2011. Humanities approaches to graphical display. Digital Humanities Quarterly, 5(1).

Edwards, P.N., 2019. Infrastructuration: On habits, norms, and routines as elements of infrastructure. In Thinking infrastructures. Emerald Publishing Limited.

Eisenhardt, K.M., 1989. Building theories from case study research. Academy of management review, 14(4), pp.532-550.

Flanders, J.H., Muñoz, T. and Senseney, M., 2012. Digital Humanities Data Curation. DH Curation Guide: A community resource guide to data curation in the digital humanities.

Gibbs, F. and Owens, T. 2012. Building Better Digital Humanities Tools: Toward broader audiences and user-centered designs. Digital Humanit. Q. 6 (2).

Jacobson, I., Booch, G. and Rumbaugh, J., 1999. The UNIFIED Software Development Process· Addison Wesley Longman. Reading, Massachusetts.

Jeng, W., He, D. and Oh, J.S., 2016. Toward a conceptual framework for data sharing practices in social sciences: A profile approach. Proceedings of the Association for Information Science and Technology, 53(1), pp.1-10.

Jeng, W. and He, D., 2022. Surveying research data-sharing practices in US social sciences: a knowledge infrastructure-inspired conceptual framework. Online Information Review.

Ikeshoji-Orlati, V., Caton, M.A. and Stringer-Hye, S., 2018. Tiny Data: Building a Community of Practice around Humanities Datasets.

Karasti, H. and Blomberg, J., 2018. Studying infrastructuring ethnographically. Computer Supported Cooperative Work (CSCW), 27, pp.233-265.

Kulak, D. and Guiney, E., 2012. Use cases: requirements in context. Addison-Wesley.

Lamb, R. and Kling, R., 2003. Reconceptualizing users as social actors in information systems research. MIS quarterly, pp.197-236.

Leonelli, Sabina. "What counts as scientific data? A relational framework." Philosophy of Science 82, no. 5 (2015): 810-821.

Lyon, L., Ball, A., Duke, M. and Day, M., 2012, October. Developing a Community Capability Model Framework for data-intensive research. In Proceedings of the 9th International Conference on the Preservation of Digital Objects (pp. 9-16).

Maryl, M., Dallas, C., Edmond, J., Labov, J., Kelpšienė, I., Doran, M., Kotodziejska, M. and Grabowska, K., 2020. A case study protocol for meta-research into digital practices in the humanities. Digital humanities quarterly, 14(3), pp.1-20.

Millerand, F., and Baker, K.S., 2010. Who are the users? Who are the developers? Webs of users and developers in the development process of a technical standard. Information Systems Journal, 20(2), pp.137-161.

Mongili, A., 2014. Designers as users: Blurring positions and theories in creative practices. Information Infrastructure (S): Boundaries, Ecologies, Multiplicity, pp.2-25.

Pacheco, A., 2022. Digital humanities or humanities in digital: revisiting scholarly primitives. Digital Scholarship in the Humanities, 37(4), pp.1128-1140.

Palmer, C.L., 2013. Work at the boundaries of science: Information and the interdisciplinary research process. Springer Science & Business Media.

Palmer, C.L. and Neumann, L.J., 2002. The information work of interdisciplinary humanities scholars: Exploration and translation. The Library Quarterly, 72(1), pp.85-117.

Palmer, C.L., Teffeau, L.C. and Pirmann, C.M., 2009. Scholarly information practices in the online environment. OCLC Research .

Posner, M., 2015. Humanities data: A necessary contradiction. Miriam Posner’s Blog, 25.

Rimmer, J., Warwick, C., Blandford, A., Gow, J. and Buchanan, G., 2008. An examination of the physical and the digital qualities of humanities research. Information Processing & Management, 44(3), pp.1374-1392.

Schöch, C., 2013. Big? Smart? Clean? Messy? Data in the Humanities? Journal of the Digital Humanities, 2(3).

Stone, S., 1982. Humanities scholars: information needs and uses. Journal of documentation, 38(4), pp.292-313.

Unsworth, J., 2000, May. Scholarly primitives: What methods do humanities researchers have in common, and how might our tools reflect this. In Symposium on Humanities Computing: Formal Methods, Experimental Practice. King's College, London (Vol. 13, pp. 5-00).

Watson-Boone, R., 1994. The information needs and habits of humanities scholars. RQ, pp.203-215.