From Data to the Creation of Meaning Part II: Data Librarian as Translator

by Kristin Partlo[1]

Abstract

While institutions, methodology and geography all present barriers for communication and development of infrastructure, sometimes the greatest barriers may be in reaching not across the world but across the hallway. Engaging in the work of unified infrastructure requires finding language that bridges modes of inquiry and meaning, so that all participants see their place in the whole. This work of finding shared language involves translation at many levels. Data librarians know that not everyone means the same thing by 'data' and increasingly they seek language that spans the practices of social science, sciences, humanities, and performing arts. This paper aims to highlight some of the ways in which data professionals are already adept at translation. Drawing on examples from work as a subject librarian and data professional at an undergraduate institution, I will elaborate on ways in which translation permeates the daily work of data librarians, from helping new researchers learn the language and methods of a field, to supporting faculty as they expand their teaching and research across disciplines. Additionally, librarians' role as semi-outsiders within the institution situates them well to help drive conversations spanning disciplinary modes of thinking, in which faculty may also find themselves as semi-outsiders.

Keywords: Data profession, Language, Translation, Data Theory, Critique

Introduction

In the first of this two-part series, Justin Joque (page 7: From Data to the Creation of Meaning Part 1: Unit of Analysis as Epistemological Problem) discussed the ways in which the problem of data harmonization is not just technical but also political, ideological, and infrastructural. In this second part, I would like to dwell further on the ways in which the expertise of data librarians is not just technical but also cultural in the sense that much of their work is about communication, specifically translation. Though my approach is further removed from the texts and language of philosophy, it is my hope that I can use and build on the problem that Joque articulated. Namely, I propose that employing the metaphor of translation to describe the work of data librarians highlights a less obvious aspect of the expertise that they bring to the work of aligning infrastructure and data.

> ...the work of data librarians involves bridging systems of meaning and acting as translators.

The work of the data librarian can be seen as situated at the point where the efficiency of data meets the human work of interpretation, decision-making and communication. As enthusiasts for the potential benefits of making data reusable, librarians are deeply familiar with the ways in which consistent methods and standards open the doors for datasets to become valuable beyond their initially intended use. But in working with patrons who wrestle with other priorities, librarians also know that not everyone

is willing to make changes to their workflow to employ those methods and standards, or follow good data lifecycle management practices. Through their relations with scholars and students across disciplines and levels of expertise with different goals and values, the work of data librarians involves bridging systems of meaning and acting as translators. In this paper I will expand on this idea of translation and its implications through the lens of my work as a data librarian and subject liaison at a small liberal arts college. Though my job is idiosyncratic, that quality is shared by many data librarian positions, so it is my hope that there are threads here that will resonate with others in the field.

Translation and Data

While the idea of translation may for many readers invoke the Google translate tool, anyone who has used it knows its limitations. It is handy for getting the gist of a text in an unfamiliar language, but it cannot fully capture the meaning and nuance of the original text, nor is it reliable enough for much beyond casual use. Likewise, anyone who has attempted to travel with a phrasebook or to translate with a dictionary runs immediately into similar problems. Though on the simplest level, translation might seem to be a sign for sign replacement, not unlike assigning value labels in a dataset (male is 1, female is 2), it is actually a more complicated process of re-describing from one system of meaning, value, culture and experience to another. The catch is that some or most of the meaning needs to remain intact after the transformation. A recent review in the London Review of Books aptly demonstrated this complexity while discussing a new translation of Finnegan's Wake into Chinese:

> "There's plenty of Finnegans Wake that I'd be stumped to put into Mandarin. Browsing at random: 'The fall (bababadalgharaghtaka mminarronnkonnbronntonnerronntuonn-thunntrovarrhounaw nskawntoohoohoordenenthurnuk!) of a once wallstrait oldparr is retaled early in bed and later on life down through all christian minstrelsy.' I'm not sure this is convertible into any language, even an Indo-European one, but Dai's translation has been a hit in China, as the Western media reported widely at the time of publication." (Yun, 2014)

Even if it were possible to render this example word for word in another language, there are other things going on in this text that would be lost. The successful translator must be deeply familiar with not just the spoken and written forms of the original and the target languages, but also the culture and history, even, like in the case above, the sounds of the language when spoken and the associations they invoke in a listener or reader. The expertise of a translator comes from experience in both worlds of meaning, of the original and target languages, and the work of a translator involves slogging through decision after decision, interpretation, and awareness that the translation will never capture all of the original. Rather than the mechanized replacement of one word for another, or Google's more sophisticated statistical analysis of previous translations[2], rich translation requires the work of a human.

This work of interpretation and decision-making is messy and, even when it is not error-prone, any translation is imperfect and involves a loss of meaning from the original. Yet translation is necessary because, despite the loss, something is also gained, some new meaning or understanding made possible by shepherding an idea or concept from one context to another. Because of the inevitable loss, translation requires the arbitration of gain and loss of meaning which, again, requires deep familiarity with both the origin and target contexts of meaning. It is possible in some cases that the loss of meaning is greater than the gain, leading to the conclusion that translation is not possible or desirable. For example, when a data librarian helps a patron dig through documentation to become familiar with a dataset in order to reuse it for their research, they are judging whether it is possible to translate that dataset into the context of the new work. Sometimes the decision can be that the data are not a good fit because the loss would be too great to justify the translation.

Just as textual translation involves fluency of multiple languages and their cultures, data librarians must be familiar with the disciplinary contexts in which data are created and used, the languages, practices, ontologies and classification debates that inform them. Working without these fluencies can lead to mistranslations that can set work back or cause librarians to lose credibility with faculty researchers, instructors or students. As appealing as it might sound, librarians know that datasets are not like so many apps in the data archive app store ready to be plugged into any research project. Datasets have their own ecosystems of sense and values and rules, and they require documentation in machine and human readable form to allow for informed decision-making about their careful reuse. The ability of researchers to make such decisions depends in part on the work of data librarians, who collect, assure the quality of, and help users interpret that documentation. The more familiar data librarians are with the types of research projects that produce and use sharable data, the better job they will do. Just as translation always involves some loss of meaning, data librarians' efforts to build systems and services are informed by a need to balance gain and loss, measuring efficiencies against decreased ease of use or meaningfulness in particular contexts.

Translating Day to Day

Translation, with its technical and cultural aspects, shares this dual quality with both data and the expertise of data librarians. Anyone familiar with the history of the U.S. Census knows that data themselves are cultural and political artifacts even as they are created for an analytical purpose. Likewise, data librarians, valued for a certain technical expertise, also have a cultural expertise that is present in and built out of their day to day work. Data librarians translate between datasets and users, students and their professors' assignments, metadata and repositories, researchers across disciplines, and librarians and other professionals. In nearly every aspect of their work - collecting, describing, teaching, providing reference assistance, building systems and informing campus data management policy - librarians work among and between cultures of data use that are distinct with their own languages and worlds of meaning that overlap in some ways and not in others.

The daily work of translation is well illustrated by looking at the work of service-oriented roles. Providing data reference services involves listening to patrons' questions and translating what they say into statements of need or inquiry that can either be addressed directly or through referral. The reference interview process involves empathizing with the patron and understanding as much as one can about the context of the question - not just trying to take it at face value. This need is then matched in particular to an understanding of the collection and how it is organized as well as more generally to the landscape of scholarly communication and the search tools available. Furthermore, in my case working with undergraduates, the question must also be interpreted in light of what I know about the professor's goals for the assignment. Librarians take questions stated in the language of a novice and provide a bridge to the works organized according to the systems of disciplines and experts. The patron might wish that everything

were organized according to the logic of their own research topic, or that search engines could be sophisticated enough to anticipate their needs, but given that impossibility, it is clear that a human must be present and ready to help those transactions take place between the language of the patron looking for data and the language of the collection, repository, disciplinary literature, or dataset documentation.

Libraries as institutions attempt to place the works of all the academic disciplines in one collection, yet in their own languages. Librarians who tend these collections and translate their value to scholars and students exist in a place between and among the disciplines. A universalizing conception of the library, like those discussed by Joque (see page 7 Joque, J. (2014) From Data to the Creation of Meaning Part I: Unit of Analysis as Epistemological Problem. IASSIST Quarterly [Online] 38(2). Available from: http://iassistdata.org/iq/issue/38/2. [Accessed: 4 March 2015]), places the library outside and above the disciplines, organizing them within an overarching ontology. Focusing on the work of librarians as translators shifts the focus of the work from crafting the universal system to something more liminal, running through the spaces in between the disciplines. Situated in this way, data librarians must always be translating, building technical infrastructure while also building, participating in, and constituting cultural infrastructure.[3] By cultural infrastructure I mean the social norms, practices, and expectations in which our systems function and make sense as well as the cast of characters who enact them. Viewing the work in this way has implications for how data librarians organize and prioritize their time, form partnerships, develop expertise, and explain the nature of their work to their bosses.

Easily seen as a disadvantage, existing in a space of imperfect translation also opens up the potential to help frame issues in new ways. Librarians, rarely as fluent in any one disciplinary language as the teaching faculty, are at a disadvantage when speaking to faculty in their own disciplinary languages. However, when those same faculty step outside their own home context, for example when doing interdisciplinary work, it becomes easier for them to rely on others and for librarians to offer help. When one knows that one is learning, the expectations are changed and it allows space for imperfect articulation. A barrier of authority and fluency is removed. Librarians, who are accustomed to finding themselves in this liminal space, can take advantage of and recognize this inversion as an opportunity to make themselves understood. Data librarians can empathize with the uneasy feeling of communicating in a language that is not their first and are positioned well to anticipate how and where they can help.

Sometimes imperfect translations serve to draw people out of their native language into unfamiliar territory making it easier, when all goes well, to see commonality. For example, as part of a gap assessment, the Research Data Services and Support Group on my campus wrote a document (2012) articulating the points where students working with quantitative information in any class might run into trouble and seek assistance. It was a simple idea, but was complicated by conflicting uses of terms like 'analyze,' 'collect,' and 'data' in different disciplines. In the end, it was written in imperfect general language, not aligned with any one of the disciplines, but was meaningful enough to trigger wide engagement on an issue that had not gained traction in the past. By finding language specific enough for people to see their experience in it, but general enough to draw people out of their own disciplinary perspectives, it allowed a different kind of conversation take place across disciplines.

Two other examples illustrate further this idea of accessing an in-between space of meaning. First, being at a teaching college, it is sometimes more fruitful to raise an issue with faculty as a pedagogical question first. The language of teaching and learning is one in which faculty expect to see multiple disciplines reflected in close proximity. I have had better success engaging faculty about how to teach students to manage their data than I have had talking with faculty about their own research or teaching data. By shifting the conversation outside of my primary expertise and outside of the faculty's research area into a shared second language, we are able to find common ground.

In a similar move, in order to engage the topic of data across the curriculum with librarians who do not work with traditionally quantitative fields, I have shared an article by boyd and Crawford (2011), "Six Provocations for Big Data," which draws out some of the broader questions about how big data in research are fundamentally changing the ways researchers ask questions across disciplines. boyd and Crawford's language is broad, for example, they pose that big data "reframes key questions about the constitution of knowledge, the processes of research, how we should engage with information, and the nature and the categorization of reality" (boyd & Crawford. 2011, p. 3). By broadening the topic and effectively raising the stakes from the relatively narrow concept of big data to issues of epistemological change, we were able together to see the impact of these ideas on all of our areas of expertise.

Implications of Considering Data Librarians as Translators

IIt is widely recognized that metadata has the greatest chance of being meaningful if it is written in the language of the creator of the data. The creator not only has the most intimate knowledge of the data, but also speaks the language of the discipline or scholarly community in which the project emerged. When researchers speak to each other within their own field, they draw on the literature, they know which terms are contentious and which are clear. Furthermore, they are familiar with and can appeal to shared values. For example, Stephanie Hampton and her co-authors (2013) make what is basically an ecological argument in favor of data stewardship and reuse. Through discussion of multiple examples of research that made use of existing data to solve stubborn problems of measurement, she demonstrates how researchers operate not alone but in a system and within an environment of existing data. By framing her argument in this way, she appeals to the professional commitments of ecology, such as reuse and tending to systems, to make a case for sharing data. It would have been difficult for a non-ecologist to be persuasive in this way without this degree of disciplinary fluency. If part of the work of a data librarian is to translate the appeal for good data practices into the disciplinary languages of faculty, then it follows that part of the job is to develop these fluencies. One might attempt to do so directly in areas with affinities to one's own expertise, or one might turn to subject librarians to get closer, by proxy, to the ideal of speaking fluently across all of the disciplines. Both of these options take time, which is easier to justify when the translator role is a visible part of the work.

Finally, making visible the data librarian's interpretive work as a translator highlights the data librarian's teaching role. Unlike

technical solutions that can be set into action and observed from a distance, bringing about cultural change involves educating researchers and teachers, emerging scholars, and other professionals about the value of managing data according to established (and emerging) good practices. It is not enough to present these practices as they have emerged in the social sciences, in their native language and context. Instead, data librarians (with subject liaison partners) do the work of making these practices appear relevant by translating them into language that is meaningful in other contexts and workflows and that speaks to the relevant intellectual motivations and values. Working in this way is slow going, decentralized, and requires room for failure, miscommunication and mistranslation.

A current focus in my own work is bringing knowledge of social sciences data management to the digital humanities initiative on campus, with which I am peripherally involved. In my own liaison areas, where I have confidence that concepts of data management work their way to greater and lesser degrees into methodology instruction, my approach to education and outreach is to complement what the professors are already teaching or aspire to teach their students. In the digital humanities, the humanities librarians and I are working on finding language and metaphors to help scholars see their existing practices with materials, digital or otherwise, as amenable to data management. For example, we have used summarized versions of the Data Curation Profiles interview instruments (Carlson 2010) and used them as a discussion exercise in several settings with other librarians and with the undergraduate digital humanities interns to introduce the concepts of data management and reshape them into a meaningful framework for re-applying the model and thinking about what counts as data in the digital humanities. Nearly always, the term data gets replaced with something like research materials, but that replacement is not sufficient to make the disciplinary leap. Without translation of these concepts in a very concrete manner to questions and considerations familiar to individuals in the humanities, too many people see data management as something that does not apply to the kind of work they do even as their work becomes increasingly digital.

Infrastructure is something most people don't see or think about until it breaks down. Through their work, data librarians make visible the challenges of aligning infrastructure, both technical and cultural. The work of data stewardship is not a back room problem, but one that is tied up in cultures of research, teaching, and processes of scholarly communications.

Data librarians engage in the cultural work of translation in many ways, and that skill is a part of their expertise. Such expertise is needed in developing data management policy at the institutional level and in the broader culture increasingly interested in big data. Professionals with the detailed knowledge of data structures and practices can help translate the value of integrating best practices to those who teach, those who collect data in the field, those who fund the research and the institutions that support it, and those who are learning to become tomorrow's researchers.

Acknowledgements

I am deeply indebted to my colleagues at Gould Library, Carleton College, who unfailingly and generously offer creative and insightful dialogue and critical editing, especially Heather Tompkins and Iris Jastram. I am grateful to my employer and my library for supporting and valuing the work of contributing to the published scholarship in our field and to IASSIST for providing a forum for it. Finally, this paper would not have been written without the prompt from (and the ensuing lively discussion with) Justin Joque (see page 7 Joque, J. (2014) From Data to the Creation of Meaning Part I: Unit of Analysis as Epistemological Problem. IASSIST Quarterly [Online] 38(2). Available from: http://iassistdata.org/iq/issue/38/2. [Accessed: 4 March 2015])to collaborate on a project drawing on our shared background in continental philosophy.

References

boyd, d. and Crawford, K. (2011) Six Provocations for Big Data. A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society. 21 September 2011. Oxford: The Oxford Internet Institute. [Online] Available from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1926431. [Accessed: 3 June 2014]

Carlson, J. (2010) Data Curation Profiles Toolkit [Online] Available from: http://datacurationprofiles.org/. [Accessed: 2 June 2014]

Costa-Jussà, M.R. & Farrús, M., 2014. Statistical Machine Translation Enhancements Through Linguistic Levels: A Survey. ACM Computing Surveys, [Online] 46(3), pp.42:1–42:28. Available from: http://dl.acm.org/citation.cfm?doid=2518130. [Accessed: 25 July 2014]

Hampton, S. E. et al. (2013) Big Data and the Future of Ecology. Frontiers in Ecology and the Environment. [Online] 11 (3). p. 156-162. Available from: http://www.esajournals.org/doi/abs/10.1890/120103. [Accessed 3 June 2014]

Joque, J. (Forthcoming) From Data to the Creation of Meaning Part I: Unit of Analysis as Epistemological Problem. IASSIST Quarterly [Online] ? (?). Available from: ?. [Accessed ?]

Research Data Services and Support Group (2012) 10 Points in Working with Quantitative Data when Students Need or Seek Support. [Online] Northfield: Carleton College. Available from https://apps.carleton.edu/campus/library/assets/10_Points_Data_Support_13.10.09.pdf [Accessed 30 July 2014]

Yun, S. (2014) Short Cuts. London Review of Books. [Online] 36 (7). p. 24. Available from: http://www.lrb.co.uk/v36/n07/sheng-yun/short-cuts. [Accessed 3 June 2014]

Notes

1. Kristin Partlo is Reference & Instruction Librarian for Social Sciences and Data at Carleton College in Minnesota, USA. She can by reached email: kpartlo@carleton.edu. This paper was presented at the 2014 IASSIST conference in Toronto, Ontario, Canada on 4 June, Session 3J, along with its companion paper by Justin Joque, "From Data to the Creation of Meaning Part I: Unit of Analysis as Epistemological Problem."
2. Readers interested in learning about statistical machine translation methods used by Google and other translation software will find a literature review and useful insights into how linguistics and computer science concepts are used together in this multidisciplinary field in a recent survey by Marta Costa-Jussà and Mireia Farrús (2014).
3. My colleague, Heather Tompkins, frequently uses this expression in an analogy about the current state of Digital Humanities being like finding yourself in a car on a rope bridge. The technical tools may be there to make certain projects possible, but the cultural infrastructure is not yet developed sufficiently to plan well and prevent disasters.