Data Remediation as Collaborative Process

Martin, Kim
kmarti20@uoguelph.ca
University of Guelph, Canada

Brown, Susan
sbrown@uoguelph.ca
University of Guelph, Canada

The Linked Infrastructure for Networked Cultural Scholarship (LINCS) is mobilizing Canadian scholarship through the creation, dissemination, and use of linked open data (LOD). The project goal is to allow users from fields including history, communications, music, performance, and literary studies to interweave their research material meaningfully with data from cultural data publishers (galleries, museums, archives, libraries, presses) in order to transform how it is possible to engage with culture on the Semantic Web. As the project prepares to launch in 2023, this paper reflects on the activities of core LINCS project members as forms of remediation. Not only does the project work towards what may provocatively be understood as "[t]he action of remedying or correcting" (OED) a variety of differently formatted datasets into a findable, accessible, interoperable, and reusable (FAIR) state (Wilkison et al), but the results of this reworking allow for researchers and the public to remediate their knowledge through new methods of connection, visualization, and annotation.

This paper will reflect on three instances of dataset remediation within LINCS, using this term as a keyword for exploring the possibilities of LOD for the humanities, drawing on various senses of remediation from ecology (Shepherd), media studies (Bolter and Grusin), and digital humanities (Schuster and Dunn), and on the concept of (meta)data remediation (Groat; Radio and Hanrath). We demonstrate how a dataset enters the LINCS conversion workflow, becomes a part of the LINCS triplestore, and is then accessed by researchers through a variety of interfaces. The remediation of these diverse source datasets, on advertisements, paintings, and literary history, will demonstrate the ways in which these three processes of remediation are crucial to the creation of LOD through LINCS, and to the improvement of the data environment or ecosystem.

The first process occurs before any data is exchanged: the remediation of research conceptualization. The LINCS Research Board Chair, Ontologist, Project Manager, and Data Interface Developer meet with the leads of each research project for a data intake interview. During these conversations, researchers often comment that questions from the LINCS team about intent, granularity, audience, difference, and provenance change the way they conceptualize their projects. Although this result is unintentional, eliciting the context that is needed for the LINCS team to understand how best to map and convert these datasets initiates a process of rethinking everything from research questions to publication of project results.

The second process of remediation occurs as researchers' datasets go through the LINCS workflow for conversion to LOD. Through ontology and vocabulary mappings, reconciliation against authorities and conversion to LOD, the LINCS team works both to understand the needs of individual researchers and to bring this understanding to aligning their data with our core ontology, CIDOC-CRM, or ways of extending the CIDOC base, to ensure compatibility across datasets and with LINCS tools. LINCS source datasets differ in form (structured, XML and natural language data) as well as in granularity, precision, domain, language, historical period, geopolitical location, and epistemology, in addition to researchers' own identities and disciplinary methodologies. Datasets therefore require quite generalized ontological structures that can accommodate these differences. The granular and nuanced complexities of human artifacts and human subjectivities, including the varied and interdependent, imbricated and intersectional (Crenshaw) facets of identity, given CIDOC's structure, manifest in vocabularies, which means that remediation of the data is being distributed across two different types of structure, which has implications for how the data can be used and accessed.

Finally, LINCS must create interfaces that mediate among these myriad perspectives and multiple worldviews, providing space for debates, difference, and contextualized and situated knowledge (Haraway) on the Semantic Web. As an infrastructure project funded to leverage as far as possible existing DH tools and platforms, LINCS has been challenged to make space for the desires of different research projects while adapting platforms to be usable by a broad audience and multiple projects. However, the technology stack is so complex that it is a challenge to make LOD usable for research, for reasons related to expertise, implementation, and infrastructure (Sanderson). Enabling users to make connections between data points across the Semantic Web is challenging, given the complexity of the data, meaning that the comfort of traditional media forms (images of paintings or print advertisements, for example) provides an important means of grounding the user within a novel and at times disorienting web interaction. We will demonstrate our attempts to overcome this complexity by showcasing how LINCS researchers are using LOD through two interfaces: the Context Browser plugin and our instance of ResearchSpace.

Bibliography

Bolter, J. David / Grusin, Richard (1999): *Remediation : Understanding New Media* . Cambridge, Mass: MIT Press.

Crenshaw, Kimberlé (1989): "Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics," in *University of Chicago Legal Forum*, 1989, 1.

Groat, Greta (2009): *Future Directions in Metadata Remediation for Metadata Aggregators.* Digital Library Federation.

Haraway, Donna (1988): "Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective," in *Feminist Studies* 14,3: 575–99. https://doi.org/10.2307/3178066.

Radio, Erik / Hanrath, Scott (2016): "Measuring the Impact and Effectiveness of Transitioning to a Linked Data Vocabulary," in *Journal of Library Metadata* 16,2: 80-94.

"remediation, n." *OED Online*. Oxford University Press, September 2022. Web. 4 November 2022.

"ResearchSpace." *The British Museum* , https://researchspace.org/.

Sheppard, Laurel M. (2003): "Remediation," in *Environmental Encyclopedia*, ed. Marci Bortman et al, vol. 2, 3rd ed. Gale. 1191-1192.

Sanderson, Robert (2018): "Shout It Out: LOUD." Europeana-Tech Keynote Address, 2018. https://www.youtube.com/watch?v=r4afi8mGVAY.

Schuster, Kristen / Dunn, Stuart (2020): "Introduction," in *Routledge International Handbook of Research Methods in Digi-*

tal Humanities. Edited by Kristen Schuster and Stuart Dunn. Routledge.

Wilkinson, Mark et al. (2016): "The FAIR Guiding Principles for Scientific Data Management and Stewardship," in *Scientific Data* 3.1 (15 March 2016): 160018. doi:10.1038/SDATA.2016.18.