Journal of World-Historical Information
JWHI

Matt Drwenski, Christopher H. Myers

Data, Set, Match:
Reflections on Database History

Reflections and Comparisons

All of these topically, geographically, and spatially wide-ranging projects, at the very least, meet the goal we set out for this thematic cluster: to construct historical narratives through database conception, collection, and analysis. Historians can concretely and without a lifetime of study or training reach new conclusions through data. Often these new conclusions can only be, or can best be reached through database research. A broader point that we would like to emphasize is that it seems that even the historians who feel positioned farthest from quantitative or digital human can contribute to historical database research. None of these projects are complete or are independent of the other graduate work by each researcher, but by forcing people to both complete some of their analysis allowed us to engage with and reflect on the process of doing database-history.

All of the projects reflected on the assumptions and shortcomings of the datasets. For example, despite looking at two different sources or historical data, Myers' research and Sherry's work most explicitly delimit the limitations of their data. Myers recognizes the difficulties that varying amounts of information about historical individuals present for drawing collective conclusions. Sherry uses the limitations of the nation-based refugee data as the foundation for a transnational, as opposed to international or multinational, conceptualization. The projects' understandings of datasets' limits, assumptions, and analytical applicability determine the choice of descriptive or

Volume 2-3, No. 1 (2014-2015) | ISSN 2169-0812 (online)
DOI 10.5195/jwhi.2015.14 | http://jwhi.pitt.edu

complex statistical and visual tools. For example, disparities in type of data and geographic units favor more descriptive analysis of each dataset in Drwenski's work. In Behrendt's project, the study group's small size and temporal specificity precludes more complex tools. The large number of cases, harmonization across long timespans, and numerous variables in Sherry and Myers' datasets encourage the use of more complex visual and statistical analysis.

The clearest division is between levels of data-collected. Myers and Behrendt wrote social biographies of clearly defined groups to draw out broader conclusions, showing the influence of educational and socioeconomic factors, and cultural dynamics on elite labor. Drwenski and Sherry draw conclusions about less cohesive groups but are still concerned with individuals who are functioning in aggregate transnational processes. A related division is that Myers and Behrendt created databases by bringing together primary sources, while Drwenski and Sherry have aggregated and standardized others' datasets. Indeed, data can be created and aggregated from a variety of sources and places ranging from secondary literature to the manuscript archive and digital database collection. Despite these differences, these five projects construct social historical narratives which grapple with the relationship of focused data within societal dynamics that cannot be observed in the data alone. The process of creating database projects more focused on economic, politico-military, institutional, or environmental historical questions might have some clear differences, especially in the construction of narrative. Traditional quantitative history is often focused on the data as an end in itself, but our five articles center on narrative history as the end product of historical analysis.

Insights

N. Katherine Hayles suggests that narrative and database can function as symbiotic creatures. But, are narrative and database even different species?[1] First, the process of creation is similar: the categorization of ideas, people, and events. Of course, a database is necessarily more explicit and its assumptions should be more apparent. This explicitness is also limiting, that which cannot be made explicit is difficult to include in a database. Narrative is better at making assumptions more implicit. This is essential in order to describe the complexities of human history. Fundamentally, however, both approaches are parts of the same spectrum of historical investigation. We then have a tradeoff between quantification and explication on one hand, and qualification and implication on the other. Our point remains that the same process underlays both approaches. When we take this perspective, Andrew Behrendt's "data epiphany", as a contextual historian, seems less miraculous and more the result of our commonalities as historians. The divisions between contextual histories and social science methodologies and the gap between micro-historical and macro-historical, are less apparent when the process and practice of this works are put into focus. When viewed from the process of historical creation, database and narrative are not different species. Perhaps, like dogs and wolves, they share a common underlying DNA, but narrative has been domesticated and has served human purposes for some time and databases still need some taming.

Inspired by the CHIA initiative, we had set out to stretch the limits of using databases and world-historical information to see if we could find common methods, themes, and processes through the diverse topics and approaches we followed. We succeeded in that we have come to the conclusion that despite our diversity of topics these five projects cohere around a common method. They analyze social structures and dynamics at both smaller and larger levels through conception, data collection, and analysis generally following the process laid out by Manning in *Big Data in History*. The different levels of data scholarship in our papers can be broken down by space time and topic, but still share common concerns, methods, and processes.[2]

NOTES

[1] N. Katherine Hayles, *How We Think: Digital Media and Contemporary Technogenesis*, (Chicago: The University of Chicago Press, 2012), ch. 6.

[2] Patrick Manning, *Big Data in History* (London: Palgrave Pivot, 2013), 16.

Articles in this journal are licensed under a Creative Commons Attribution 4.0 United States License.

This journal is published by the University Library System of the University of Pittsburgh as part of its D-Scribe Digital Publishing Program and is cosponsored by the University of Pittsburgh Press.