Toward a Typology of Digital Thematic Research Collections

Katrina Simone Fenlon
kfenlon2@illinois.edu
University of Illinois at Urbana-Champaign
United States of America

Introduction

This paper considers an evolving genre of digital scholarship in the humanities, the thematic research collection, which is distinguished among other kinds of scholarly production as a collection of primary sources, gathered by scholarly effort and made available online to support research on a theme (Palmer, 2004). There are hundreds of such collections on the Web, ranging from well-known digital archives to small collections of historical or literary evidence within a thematic niche.

Despite recognition of the genre (e.g., Price, 2009; Flanders, 2014; and Thomas, 2015), we do not know enough about this mode of production, how it contributes to humanities discourse, or how it relates to systems of peer review, discovery, and long-term maintenance. The evolution of public-facing humanities scholarship, long-term access to collections, and the completeness of the scholarly record depend in part on a more systematic understanding of this and other emergent genres.

Through typological analysis, this research aims to build a foundation for rigorous study of thematic research collections. The goals of typology are to understand the breadth and variety of a genre, and identify unanticipated variations. We take up the following questions: What types of collections can we usefully distinguish, and what can these types and their characteristics reveal about the challenges and opportunities confronting the growth of digital scholarship in the humanities?

Method

We conducted a pilot survey of the digital humanities landscape to identify a set of resources meeting our definition. Sources for the survey included digital humanities centers, library publishing programs, tools and platforms for digital publishing, and scholarly collectives/peer review organizations. While not comprehensive of the digital humanities landscape, the survey produced a set of 98 diverse collections.

Our typology followed the formal process described by Kluge (2000):

1. Develop relevant analyzing properties. Properties reflect our intuitions about interesting differences between collections, within this context of scholarly work and use.
2. Group members by distinct combinations of properties.
3. Analyze meaningful relationships and construct types.
4. Repeat earlier steps if needed to accommodate collections that do not fit.

We iterated our analysis, refining our sense of properties and resultant types, until we were satisfied that our types speak to important and revelatory differences among collections.

Analysis

Our proposed typology of thematic research collections relies on the following four properties of collections, which are basically determinative of their potential uses: (A) Whether primary sources are the main content of the collection, or are ancillary; (B) Whether the collection employs advanced markup, to enable use beyond basic keyword search; (C) Whether the collection's primary purpose is pedagogical; and (D) Whether the collection solicits, or actively engages in the collection of new or original evidence.

The 98 collections in our set resolve into 5 types, per different combinations of these properties. Figure 1 shows how types are derived from a matrix of properties, along with the number of collections that fall into each type. Figure 2 visualizes the types in a projected, three-dimensional property space.

We can briefly describe the types as follows:

- Type 1. Traditional collections with enabling markup: Marked-up (usually textual) primary sources constitute the main content of the collection, and are accessible directly by search and other functionalities.
- Type 2. Traditional collections without enabling markup: These are more heterogeneous in content, but primary sources are still directly accessible as such.

- Type 3. Data-centric or derivative-centric collections: While primary sources are a major component, they are not directly accessible as such. Rather, access is mediated by an analytic or interpretive layer, such as an interactive map or 3D model.
- Type 4. Pedagogical collections: They resemble one of the above types, but are distinguished by their intended purpose and audience.
- Type 5. Original or soliciting collections: They resemble one of the above types, but are distinguished by the scope and processes of their development (specifically, they are collecting new primary sources).

Properties:	(A) Direct access to primary sources	(B) Indirect or mediated access	
(C) Advanced markup	26	1	Type 1
(D) Minimal markup	34	9	Type 2
(E) Pedagogical	6	2	Type 3 / Type 4
(F) Original or soliciting	18	1	Type 5

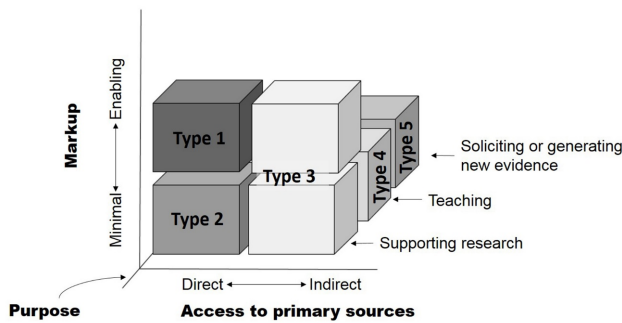Figure 1. Property matrix indicating numbers of collections in each type

Figure 2. Types visualized in three-dimensional property space

Discussion

We hope the systematic identification of properties of a broad range of collections and preliminary types may serve as a foundation for ongoing study of how collections work, and how they may be served (or not) by systems of evaluation, discovery, and long-term maintenance. The full version of this poster details methods, properties, and types with vivid examples, and considers the implications of types for effective and ongoing access to collections.

Bibliography

Flanders, J. (2014). Rethinking Collections. In P. L. Arthur & K. Bode (Eds.), *Advancing Digital Humanities* (pp. 163–174). Palgrave Macmillan UK.

Flanders, J., & Jannidis, F. (2015). *Knowledge organization and data modeling in the humanities*. Providence, R. I.: Workshop on Knowledge organization and data modeling in the humanities, Brown University.

Flanders, J., & Muñoz, T. (2012). An Introduction to Humanities Data Curation. In *DH Curation: A Community Resource Guide to Data Curation for the Digital Humanities*. Champaign, IL.

Kluge, S. (2000). Empirically Grounded Construction of Types and Typologies in Qualitative Social Research. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, *1*(1).

Palmer, C. (2004). Thematic Research Collections. In *A Companion to Digital Humanities*. Blackwell Publishing.

Price, K. M. (2009). Edition, Project, Database, Archive, Thematic Research Collection: What's in a Name? *Digital Humanities Quarterly*, *3*(3).

Thomas, W. G. (2015). The Promise of the Digital Humanities and the Contested Nature of Digital Scholarship. In S. Schreibman, R. Siemens, & J. Unsworth (Eds.), *A New Companion to Digital Humanities* (pp. 524–537). John Wiley & Sons, Ltd.