Thematic Research Collections: Libraries and the Evolution of Alternative Digital Publishing in the Humanities

Katrina Fenlon

Abstract
The growth and evolution of digital scholarship in the humanities has produced new genres of scholarly work and publication, reliant upon new ways of representing and sharing evidence, analysis, and interpretation. Meanwhile, extant systems of scholarly communication, including publication, discovery, access-provision, maintenance, and preservation, too often exclude digital research products, to the potential detriment of the entire scholarly record. This paper considers one genre of digital humanities scholarship: the *thematic research collection*, a digital collection of primary sources gathered to support research on a theme. This genre is recognizable and increasingly common, yet wildly heterogeneous in precise form, function, and purpose. This typological analysis aims to identify and describe types of collections as a way toward comprehending the range, variation, and complexity of the whole genre. The research considers what thematic research collections are, how they work, and what challenges confront the provision of effective and ongoing access to digital scholarship.

Introduction
Thematic research collections are an evolving genre of digital scholarly production in the humanities. They are, at root, collections of primary sources gathered by scholarly effort and made available online to support research on a particular theme (Palmer, 2004). There are hundreds of such collections in various stages of development on the internet, usually made available through digital humanities centers. Exemplars range from well-known digital archives, such as the William Blake Archive, to small

collections of historical or literary evidence within a thematic niche, such as Nineteenth-Century Disability: Cultures and Contexts.

Despite widespread acknowledgment of digital collections as significant scholarly products, we do not know enough about this mode of digital production, the nature of its contribution to humanities discourse, or how best to integrate collections into systems of publication, 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 of scholarly production.

This paper aims to build a foundation for study of thematic research collections through typological analysis. It takes 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 and management of digital scholarship in the humanities?

My proposed typology, derived from a systematic analysis of collections, distinguishes five kinds of collection according to differences in their underlying data models. The goal of typological analysis is to perceive the breadth and diversity of a genre and identify variations omitted from conventional conceptions (Koch, 2000). The argument of this paper is not that our typology bears ontological weight nor that clear lines need to circumscribe evolving genres of scholarship, but rather that identified types and their differences may shed light on the characteristics of digital scholarship and their potential implications for libraries.

Thematic Research Collections and Libraries

Thematic research collections have long been acknowledged as one form of scholarly production in the humanities (Flanders, 2014; Palmer, 2004; Price, 2009; Thomas, 2015; Unsworth, 2000). More than a decade ago, Palmer (2004, n.p.) anticipated that "scholar-created research collections are likely to increase in number as the work of producing them becomes more widely accepted as legitimate scholarship." As predicted, the genre has grown in the intervening years, despite scholars' continued reliance upon traditional forms of production, especially journal articles and books (Acord & Harley, 2013; Housewright, Schonfeld, & Wulfson, 2013). Some collections constitute the end goal of a project; others emerge as valuable secondary outcomes of a research process. Thematic research collections may be understood as one among an array of rapidly evolving, often experimental modes of digital production and publication in the humanities. In practice, the hard edges of these genres seem to dissolve, and there is considerable overlap in their forms and functions. All confront steep uphill climbs toward integration into existing systems of scholarly evaluation, centralized access, and long-term sustainability or preservation.

Two important shifts in scholarly communication contextualize this research: a shift in what scholars want to disseminate, and a related shift in the library's role toward more active participation in and publication of research. Libraries, of course, have long played a number of roles in the processes of scholarly communication, including providing research and authoring support; describing scholarly products; enabling discovery and access; and preserving or otherwise providing sustainable access to scholarship over time. The last few years have witnessed a dramatic rise in an additional role for libraries in scholarly communication: libraries are increasingly involved with publishing, especially open-access scholarship and experimental forms that fall through the cracks of traditional publishing systems (Lippincott, 2015). Even when publishing is not the direct province of the library, libraries are increasingly called in as collaborators for experimental digital scholarship. For example, the Andrew W. Mellon Foundation recently funded a spate of projects to help university presses build capacity for publishing alternative digital scholarship, many of which projects entailed significant collaboration with the university library (AAUP, n.d.). Along with an emergent literature on enhanced publications (see, for example, Bardi and Manghi [2014]), services and tools have emerged for the publication of collections of humanities sources. Consider not only Omeka, which has grown into a popular web-publishing platform designed with collections in mind, but also the University of Michigan's Fulcrum (McGlone, 2016), in beta development, which aims to publish multimedia scholarly evidence as collections alongside monographic publications. Developments in digital publishing would benefit from an improved understanding of scholarly collections, especially those aspects that pose challenges to traditional publication processes, such as the fact that collections are dynamic and tend to grow and change over indefinite periods of time; and that collections and their items frequently interlink in complex ways with external sources and publications.

After coming online, scholarly collections pose challenges for libraries and other entities in terms of evaluation, description, access-provision, and long-term maintenance or preservation. Questions of how to create standards for the evaluation of diverse digital forms are increasingly urgent. It is also unclear whether existing descriptive metadata standards are adequate to accommodate new forms. Also, how should libraries integrate digital scholarship into their collections for discoverability, organization, and maintenance? Finally, questions of best approaches to maintaining and preserving new genres are paramount. While there are limited precedents for library roles in preserving digital humanities scholarship (Rosenzweig, 2007), these questions remain largely unanswered.

Data Modeling

This study relies upon data modeling (Flanders & Jannidis, 2015) as a concept for framing how we understand digital resources that humanities scholars create. The data at the heart of thematic research collections usually comprise some type(s) of original evidence in digital form: digitized documents and images in different formats, along with constituent parts or derivatives, all surrounded by information serving to describe and interrelate these components. We may understand a *data model* within this context as an instantiated format or structure that a resource employs to represent its content. A data model often describes the entities and relationships in a conceptual space. In practice, it enables and constrains the uses and functions of content. Collections made available on the web necessarily integrate layers of data models, which may include instantiated descriptive schemas, markup standards, models for linking among resources, and so on (Flanders, 2014; Flanders & Muñoz, 2012). Exploring how data models combine and interact to create a collection is one way of exploring how collections are constituted, which is a fundamental question confronting this emergent genre (Flanders, 2014).

Typology

Method

A great diversity of things meets our working definition of *thematic research collection*. What types may be usefully distinguished from one another, and what may these types and their characteristics reveal about the development and management of scholarly collections?

This research takes a formal typological approach to identifying different kinds of collections. Typology offers one picture of a landscape, somewhat artificially differentiating kinds by constellations of properties, with the goal of identifying relevant and useful types. Our typology relies upon aspects of data models underlying digital collections, because those models serve to embody scholarly interpretation, help determine potential uses of collections, and affect their long-term accessibility and maintenance.

This study began with an exploratory survey of the digital humanities landscape in order to identify a set of digital humanities resources that meet our definition of *thematic research collection*. Sources for our survey include digital humanities centers creating or supporting the creation of thematic research collections; library publishing programs, which are increasingly involved with the publication of alternative genres; common tools and platforms for digital publishing, including Omeka and Scalar; and scholarly collectives and peer-review organizations, including NINES. While not comprehensive of the digital humanities universe, the survey

identified a set of approximately a hundred collections that are representative of the eclectic range of collections.

We performed typological analysis on this set of collections, following the formal process described by Kluge (2000):

- Identify the class of objects to be "typed" and its members. This step was accomplished through our pilot survey.
- Develop relevant analyzing properties, the bases of division of objects into types. Properties are selected to reflect our intuitions about interesting differences among collections, within this context of scholarly work and use. In particular, we sought variations in the data models underlying the collections. For example, how do collections differently facilitate access to items? How do collections represent items in relation to one another, and to contextual information and other resources?
- Group the members by the relevant properties.
- Analyze meaningful relationships and construct types.
- Repeat earlier steps as necessary to accommodate things that do not fit.

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

Analysis and Findings

Our proposed typology of thematic research collections relies upon two properties in particular of the conceptual data models of collections. The first property asks, what priority does the collection's model give to primary sources in terms of their visibility and accessibility in the collection relative to other scholarly content? In short, are primary sources the main content of the collection or are they ancillary? This property is usually reflected in how a collection is navigated and how search results are presented. The second property concerns whether the collection employs advanced markup, which may be considered deeply descriptive or semantic markup that enables functionality beyond basic keyword searches.

These properties, which are just two of myriad ones that may be ascribed to data models' underlying collections, determine how collections are developed, what subsequent data models are employed, the uses to which collections can be put, and much more. Table 1 provides more details on these two properties, which may be understood as binary properties.

A third aspect of collections became important because it also helps determine forms and functions of collections, albeit from a different angle—namely, the purposes of collections. Collection purpose may be determinative of data modeling choices rather than being specific to one aspect of a collection's data model. In this way our third property is distinct from

Table 1. Summary of two binary properties of the conceptual data models of collections

Direct item-level access to primary sources	*Indirect or mediated access to primary sources*
In collections with this property	In collections with this property
• it is clear what the primary sources are, and they constitute the main content of the collection in the sense that the site affords them priority of visibility and use; • primary sources are usually *items* in the sense of being the unit of gathering, and also the unit returned by search or individually browsable.	• while primary sources are constituents of the collection, they are accessed indirectly through other content or advanced functionality; • search and browse do not operate directly on primary sources, or they operate in a purposefully limited fashion (for example, are limited to guide users on a specific path through a collection); • access to and visibility of primary sources are mediated by an analytic or interpretive layer, which relies upon data derived from primary sources, and this layer comprises the main content of the collection.
Advanced markup	*Minimal markup*
In collections with this property, items (usually textual) are encoded with richly descriptive, semantic, or other markups that enable interpretation, curation, and functionality beyond basic visibility and keyword access.	In collections with this property, items are encoded minimally, to the extent they must be for presentation on the site or to enable keyword searches across texts.

the other two properties that shape our typology. In particular, we divided collections based on whether their purpose is primarily concerned with research or teaching, and also whether a collection solicits or actively engages in the creation (and collection) of new, original evidence. Naturally, many collections do all of these things at once: a collection may include archival materials while gathering additional historical anecdotes or artifacts from users, for example; or a collection may be used by its creator for teaching purposes while acting as a hub for a research community. This property concerns the primary focus of the collection, as reflected in its design and self-description. *Thematic research collections*, by our definition, support research; there are some, however, that give primacy to teaching or the solicitation and presentation of new (as opposed to archival) evidence. The rationale for a further purpose-based division of types is that collections with fundamentally distinctive purposes participate differently in scholarly communication, are put to different uses, and function dif-

ferently. We hypothesize that they may require different accommodations from developers, evaluators, custodians, and others throughout their lives of participation in scholarly work. Thus they are accorded separate types in our typology. Given our properties of *access to primary sources*, *markup provisions*, and *main purpose*, we can imagine a three-dimensional (3D) property space into which discrete types will fall (fig. 1).

Our typological analysis identified five discrete types into which collections may fall. They are differentiated by distinct combinations of the properties identified above. The first three types are designed for the purpose of supporting research. Type 1 collections combine direct access to primary sources with advanced markup. Type 2 collections combine direct access with minimal markup. Type 3 collections provide indirect or mediated access to primary sources. We do not see fit to divide type 3 collections into subtypes by their level of markup, because only one collection evidently employs advanced markup to mediate access to primary sources (more on this below). The final two types identified are distinguished not by any particular aspects of their data models but by purpose: type 4 collections are primarily pedagogical, and type 5 either provide original evidence or solicit new evidence.

Figure 2 is a matrix of properties, showing how they combine to form types in our typology. Within each type, numbers represent the number of occurrences of type in our survey of collections. Figure 3 shows how each
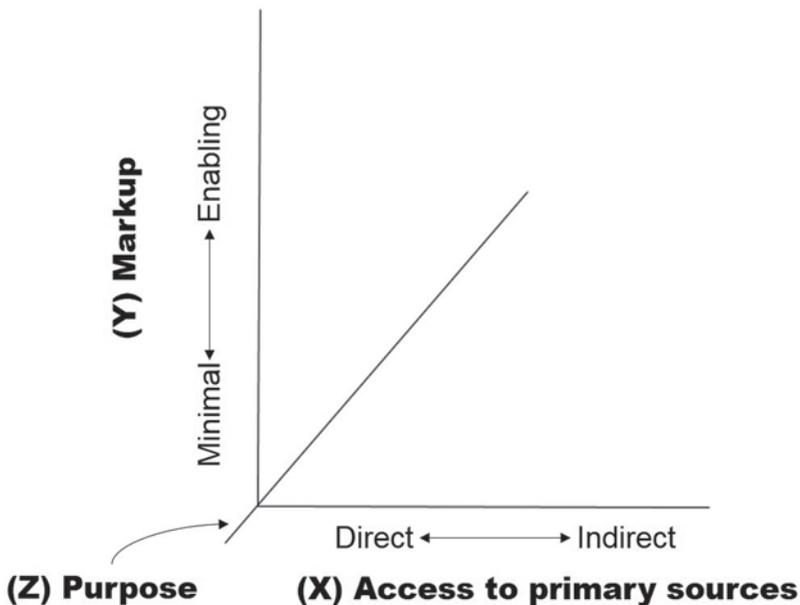

Figure 1. Three-dimensional property space.

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 / Type 3
(E) Pedagogical	6	2	Type 4
(F) Original or soliciting	18	1	Type 5

Figure 2. Property matrix indicating the types, and the numbers of collections of each type.

type can be understood to fit into our 3D property space. Below, we step through each of these types, describing their common characteristics and providing examples. In the next section we will consider the implications of these types for the development, evaluation, use, and ongoing management of collections.

Type 1 collections provide direct access to primary sources along with advanced markup, which may be semantic markup or markup that enables access to texts beyond rendering them and affording keyword searches. Marked-up primary sources constitute the main content of these collections, which predominantly collect encoded texts, although many include extensive image content and other sources that are devoid of markup. Secondary sources and other kinds of information contextualize and supplement the primary-source content in this and other types of collections. To scholars working with literary texts, these should be familiar types of collections; among them are the most well-known, oft-cited, and longest-running thematic research collections, including Thomas MacGreevy Archive, Walt Whitman Archive, and World of Dante. Primary sources are directly accessible through search and browse, and often through additional interpretive or analytic functionalities, such as multimodal views that foreground various encoded parts of texts. Markup is designed to support fine-grained access to texts through advanced searching, reading, annotation, comparative views, and other uses of primary sources. Type 1 collections are often, although not always, self-described as archives and aim to be comprehensive authorities on the works of a particular creator, or group of creators, circumscribed by time period, proximity, or social relationship.

Type 2 collections also provide direct access to primary sources, but these collections afford minimal markup for various reasons. Many of these collections gather heterogeneous media, including images and nontextual content, that are not readily encoded. When gathering text they tend to place less emphasis on affording fine-grained access to texts (as in type 1 collections) and more on providing other kinds of research
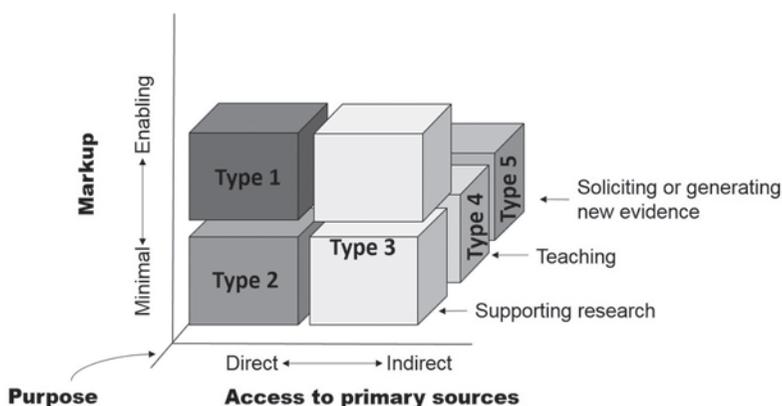
Figure 3. Types within three-dimensional property space.

support, such as comparative views of high-quality images or embedding primary sources within scholarly narratives.

Type 3 collections provide indirect or mediated access to primary sources. These may be understood as data-centric or derivative-centric collections, and they are of great interest because they stand at the very edge of our usual understanding of collections. While these collections include primary sources, they are not always directly accessible as such, or they are not prioritized in the navigation and design of the collection. Rather, access to primary sources may be mediated by an analytic or interpretive layer, such as an interactive map or 3D model, or else the collection may afford access to more granular derivatives of primary sources. In this way, the collection primarily offers or makes most visible data gleaned from primary sources. At least one collection (to be discussed in detail below) was found to rely upon advanced markup in order to facilitate fine-grained access to historical records, along with derivative information. We do not separate this case from the other data- and derivative-centric collections in terms of type in part because it is unique, and also in part because the already-distinguishing property—of providing indirect or mediated access to primary sources—is enough to make all collections with that property distinctive. However, we acknowledge that this case manifests attributes of both type 1 and type 3 collections, and with more exemplars might come to constitute a new type or subtype. Some of the implications we raise for type 1 collections (see below) will apply to this collection. Exemplars of type 3 include Voting Viva Voce: Unlocking the Social Logic of Past Politics and *Aquae Urbis Romae*: The Waters of the City of Rome.

Type 4 collections make teaching (as opposed to research) a central focus, or are built for pedagogical purposes. Collections may cater to any level of education or the general public. They may provide either direct

or indirect access to primary sources, and they may use markup or not (although all of the collections surveyed employed minimal markup). The rationale for distinguishing pedagogical collections as a separate type, despite the particularities of their data models, is that this research is concerned with collections not only in theory but also in practice. The distinctive purposes and audiences determine how such collections are developed, evaluated, used, and managed; therefore *purpose* is understood to determine subsequent data-modeling choices. Examples of type 4 collections include Salisbury Project and I Am a Man: The Memphis Sanitation Workers' Strike.

Type 5 collections are partly or primarily intended to solicit or generate new evidence. At the same time as new sources are created, they are collected. This sets them apart from collections that curate, re-present, or aggregate existing archival or literary evidence. These collections may provide direct or indirect access to primary sources, but, as with type 4 collections, they tend to manifest only minimal markup. Original or soliciting collections may thus resemble one of the above types, but they are distinguished by the scope and processes of their development, which differ fundamentally from those of other collections. Several are oral-history collections. In some collections of this and other types, solicitation or crowdsourcing of new evidence is merely one component of the collection, and the new evidence is juxtaposed with archival evidence. As with type 4 collections, we deem it important to separate these from other collections, because the methods of development and evaluation, of both items and collections, are fundamentally different and will likely carry different practical implications (see, for example, Bracero History Archive).

We have briefly characterized types 4 and 5 here, but further discussion of these is beyond the scope of this paper, which is concerned primarily with the data models that underlie collections. Future publications will address these collections in greater detail. As types 1–3 are differentiated by aspects of their data models, the following discussion considers the implications of each type.

Discussion of Types

Collections that endeavor to make primary sources directly browsable and searchable (types 1 and 2, in our typology) are the most familiar to both digital humanities scholars and libraries. Flanders (2014) acknowledges the familiarity of such collections as "an extraordinarily common way we fund and organize and interact with digital scholarly resources, a great deal of the time. [As a result,] they have taken on a certain self-evidence: we recognize and use the genre without questioning its terms" (p. 166). Part of the objective of this research is indeed to question the terms of thematic research collections. Type 1 and 2 collections comprise the great majority of such collections. When we first set out trying to understand the

variety of this landscape, we did not expect to find things that meet our definition of collections but lack direct search and browse mechanisms. We also expected that some items in every thematic research collection, which is to say the units across which search and browse and other functionalities operate, would be primary sources. These are generally unspoken assumptions of conventional understandings of collections. Type 3 collections challenge these assumptions by mediating users' experiences of primary sources in novel ways, and we discuss their challenges below. When we consider collections as integrated layers of data models, we find that type 1 and 2 collections, for all their familiarity, also pose challenges for would-be custodians, especially from a preservation perspective.

We can understand the two characteristics that distinguish type 1 collections as informing two levels of collections' underlying data models. Consider, for example, three collections that fall into type 1 and that are the among the most well-developed and well-documented thematic research collections of this sort: Shelley-Godwin Archive, Walt Whitman Archive, and William Blake Archive. Table 2 highlights relevant aspects of these collections' technical summaries so that the reader may gain a sense of how these collections are composed. Their technical summaries have pronounced commonalities, especially when considered in the abstract. Based on the commonalities among these collections, we offer a simple conceptual model that highlights common structures underlying type 1 collections (fig. 4). While this model is not comprehensive of the kinds of structures that form type 1 collections, it resonates with most of the observed collections.

At root, a type 1 collection gathers a set of digital primary sources, often digitized from artifacts at the page level, along with accompanying metadata that serves to describe and organize the digital images. Even at this base level, shown at the bottom of the "stack" in figure 4, there may be a couple of different data models working in cooperation to represent primary sources; for example, the digital-image file format and structural and descriptive metadata standards. Initial data-modeling choices made at this level affect the quality of the resource and may privilege certain uses over others. (See, for example, Flanders and Jannidis [2015] on the difference between research- and curation-driven data modeling.)

At the next level of the stack, primary sources are transcribed into text and marked up, often according to a customization of the Text Encoding Initiative (TEI) guidelines. Encoding is usually done manually or semi-automatically. At this level also there is significant variation in the specific data models used, because choices here are made in careful reflection of the chosen purposes of the resource. This level is particularly interesting for all that it can accomplish in constituting a collection. In collections that employ advanced markup, the data model(s) used to encode primary sources often bears a great deal of responsibility for how the collection

Table 2. Technical summaries of select type 1 exemplars

Walt Whitman Archive	William Blake Archive	Shelley-Godwin Archive
• Scanned images of Whitman's oeuvre saved as tiff files and derived jpegs • Transcriptions encoded in a TEI extension • Manuscript-finding aids encoded in Encoded Archival Description • Structural information for archival resources, which ties page images and so on together, encoded in Metadata Encoding and Transmission Standards	• Digital images of originals, saved as tiffs • XML files encoded in custom Blake Archive Description format to represent archival text • XML files encoded in TEI to represent supplementary materials • CSV files encoding data, such as relationships and relationship types obtaining among items, works and groupings of works, homepages and links, and so on.	• Shared Canvas data model to facilitate the description and representation of physical artifacts in terms of linked open data (to be superseded by the International Image Interoperability Framework) • Documentary transcriptions encoded in TEI customization, including "Module for Transcription of Primary Sources"

Modes of access, views, and interactive functionalities afforded by collection website:
-Search (metadata and keyword)
-Browse (multiple facets)
-Multimodal search reliant on custom encoding
-Annotations and annotation tools
-Exhibits
-Multimodal and comparison viewing of texts, images, and their components
-Visualizations and text analysis tools
-API access to transcriptions and page images

Contextual information in a variety of formats including:
-Relevant secondary sources, both internal and external to site
-Editorial and technical documentation
-Narrative interpretation, criticism, or context

Transcribed primary sources encoded in XML (most often a TEI customization), including:
-Embedded or external (linked) item metadata in XML
-Embedded image descriptions
-Embedded or external (linked) critical apparatus
-Links to other items within the collection, embedded in transcriptions or metadata
-Cues for annotation tools embedded in encoding

Original works (primary sources), usually digitized as page-images, with accompanying external or embedded structural and descriptive metadata in XML
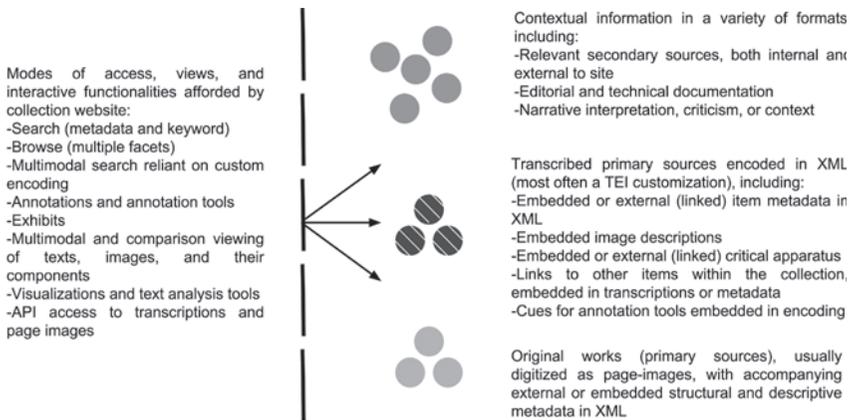
Figure 4. Common aspects of a data-model "stack" underlying type 1 collections.

hangs together, how it ultimately works. Interpretive encoding serves to open up the primary sources to advanced exploration. Encoding may also forge connections among primary sources, as information about the interrelationships among items may be embedded as links in the encoding and metadata. This layer prepares items for additional potential uses; for example, by building cues for annotation tools, comparative text viewers, and external data sources into the encoding.

While encoded primary sources constitute the main content of this type of collection and may perform the brunt of the work of holding a collection together, expressing interpretation, and enabling a variety of uses, a collection (as it is experienced by a user) is more than the items it gathers. The realization of the whole resource depends not only on the context of accompanying editorial and technical documentation, along with secondary sources, but on data models and processes that operate in the interstices among digital objects (not depicted in figure 3). These may take the form of database designs, search algorithms, management systems, cobbled-together transformation processes, and so on. In type 2 collections, which lack advanced markup, contextual information and seemingly peripheral data models (or models that are external to those used to represent digital objects) may perform still more of the work of constituting the collection.

While type 3 collections gather primary sources, their main offering takes the form of indirect or mediated access to those sources. These collections prioritize analytical or interpretive treatments of underlying sources over straightforward access provision. In most of the cases examined, these collections afford minimal or no markup to items, although some base their analyses on specialized encoding. These collections may be understood to focus on derivatives of primary sources rather than direct or one-to-one representations. While primary sources are gathered at the heart of these collections, they are not always directly accessible as such. Access to primary sources may be mediated by an analytic or interpretive layer, such as an interactive map or 3D model; otherwise, direct access is afforded to more granular derivatives of primary sources. In this way, the collection primarily offers or makes visible data gleaned from primary sources.

These collections are oriented toward a different kind of "epistemic outcome," to use Flanders's (2014, p. 166) term; beyond serving as platforms for further research, they instantiate outcomes of explicit interpretation and analysis. As we have discussed, all thematic research collections manifest interpretive or analytic work. As recognized by Palmer (2004) and many others since then, curating a collection and digitizing, encoding, and representing items online all entail some amount of interpretive work. While type 3 resources are organized as collections, often relying upon architectures similar to those of types 1 and 2 collections, they tend to exist in order to undergird representations of further analyses or interpretations. Analytical results represent the main, or perhaps the most visible, contribution of the resource, even though the collection is also intended to support research.

Consider one exemplar of type 3 collections: O Say Can You See: Early Washington, D.C., Law & Family. This collection "reconstructs the social world of early Washington, D.C., especially its multigenerational family

networks, by collecting, digitizing, making accessible, and analyzing legal records and case files between 1800 and 1862" (see Appendix). It is built on the following components (which represent some of the data models at play):

- TEI encoding of documents to capture people, places, outcomes, relationships, and so on
- CSV (comma-separated values) files that encode relationships manually inferred from documents
- Derived RDF (resource description framework) data in custom format for querying

As in types 1 and 2 collections, the main presentation of this one is structured by guided search and browsing mechanisms. However, searching and browsing prioritize abstract entities derived from case documents (primary sources): people, cases (generalized from documents), family networks, places, and so on. Case documents are available via browse and search, but the main contribution of this collection appears to be the more granular access afforded by advanced encoding of named entities, and manually derived relationships among those entities. In addition, analysis, specifically network analysis, is a goal of the project, and the analytic emphasis is one aspect of the project that sets this apart from types 1 and 2 collections. Analytic results (in the form of a graph of relationships) are appended to item results. This collection may be perceived as situated between types 1 and 3, because it relies upon advanced markup. However, the CSV files represent manual analyses that build on and transcend the analytical encodings, and this makes the collection distinctive.

While this type is less common than types 1 and 2, it is especially interesting for the ways in which it challenges our understanding of collections as an organizational form. It is more immediately evident in type 3 than in other types that these collections are dynamic and performative, similar to digital artworks or games. The layers of their data models that manifest interpretation and analysis and that rest on a base collection of evidence are often intended for interactive use and may react dynamically to underlying data and input from users. Once again, the primary sources that constitute the collection are not the whole of the collection; rather, the collection is realized by the interrelationships among those sources, their multiple derivatives, and the functional layers built on them. It is at this edge of the genre where collections begin to bleed into other genres of digital humanities scholarship, including undefined and experimental territories. And it is here where the most difficult practical questions arise.

Implications and Future Work

Flanders (2014, p. 168) proposes that data models determine the "boundedness and internal coherence" of collections. Our study of types of col-

lections and their technical underpinnings illustrates how variations in conceptual data models enact differences in how the items within a collection cohere, how collections work as unified wholes, and ultimately in their epistemological outcomes. Being organized and recognizable as collections of evidence, thematic research collections are all more alike than they are different. But there is more variation among them, and among their purposes, than anticipated: different types of thematic research collection are intended to support and present research in different ways, to different ends, and these ends are carried out by the data models' underlying collections.

We have long considered collections as gathering and being constituted by items. However, it is clear that thematic research collections, evolving into something in-between dataset and complex research publication, do more than contain a set of items. Their contributions as scholarly products stem in part from their items: both as curated selections from among the universe of available, potential evidence and in terms of the quality of the digital reproductions and the interpretation baked into transcriptions and encodings (in the case of type 1 collections). On that first level of collection, these resources add another layer of contribution in the form of relationships that the collection forges, both implicitly and explicitly, among items and between the items and contextual information. Connections are made through integrated data models. In addition, contributions are affected and effected by the layers of affordance built on the content. These layers range from thin—basic affordances of navigation and visibility—to thick, including mediations like maps and 3D models.

The biggest practical concern for the future of these and other products of digital scholarship has to do with their longevity: what are the challenges for libraries and other responsible entities in maintaining thematic research collections over time, or in archiving and preserving them? This question needs more research, and the answers will not be exclusively or even primarily technical. Preserving the discrete digital objects that constitute a collection's contents may be enough in some cases, particularly if the collection's contributions lie largely in the curation and encoding of content, rather than in the functionality or interactivity of the resource. In these cases libraries may draw on strategies for humanities data curation (see, for example, Muñoz [2013] on building capacity for humanities data curation as an aspect of library publishing initiatives). In the simplest cases digital objects may fit into existing institutional preservation systems for multimedia objects, if not institutional or discipline-specific repositories. Many cases, however, including some type 1 and most type 3 collections, may demand more complicated maintenance or preservation strategies that retain the complex connections and performative aspects of these resources. Sustaining the connections among data models is more difficult than sustaining discrete, self-contained digital objects. The challenges in

this arena resemble those in the preservation of other complex, socially functioning digital objects, such as games (Bettivia, 2016) and digital artworks (Anderson, 2016). Preserving digital humanities scholarship, and thematic research collections in particular, will likely rely upon context-specific, creative marriages of digital preservation and data-curation strategies.

The typology we propose is meant to serve as a foundation for two ongoing threads of investigation of this genre: a detailed content analysis of representative collections, which aims to further clarify their characteristics and underlying data models; and a set of interviews with people who help make and maintain these collections. This typological analysis has attempted to pull on the threads of potential types of thematic research collections in order to understand the whole fabric of the genre more fully. I do not intend to proselytize this typology as ontologically representative of the genre, but rather I hope that this systematic, albeit limited characterization has afforded useful observations about what thematic research collections are, how they work, and what challenges confront our provision of effective and ongoing access to digital scholarship.

APPENDIX: THEMATIC RESEARCH COLLECTIONS

Aquae Urbis Romae: The Waters of the City of Rome (http://www3.iath.virginia.edu/waters/)
Bracero History Archive (http://braceroarchive.org/)
I Am a Man: The Memphis Sanitation Workers' Strike (http://dlxs.lib.wayne.edu/iamaman/)
Nineteenth-Century Disability: Cultures and Contexts (http://www.nineteenthcenturydis
    ability.org/)
O Say Can You See: Early Washington, D.C., Law & Family (http://earlywashingtondc.org/)
Salisbury Project (http://salisbury.art.virginia.edu/)
Shelley-Godwin Archive (http://shelleygodwinarchive.org/)
Thomas MacGreevy Archive (http://www.macgreevy.org/index.jsp)
Voting Viva Voce: Unlocking the Social Logic of Past Politics (http://sociallogic.iath.virginia
    .edu/)
Walt Whitman Archive (http://www.whitmanarchive.org/)
William Blake Archive (http://www.blakearchive.org/)
World of Dante (http://www.worldofdante.org/)

REFERENCES

Acord, S. K., & Harley, D. (2013). Credit, time, and personality: The human challenges to sharing scholarly work using Web 2.0. *New Media & Society*, *15*(3), 379–397.
Anderson, D. P. (2016). Preserving hybrid objects. *Communications of the ACM, 59*(5), 44–46.
Association of American University Presses (AAUP). (n.d.). Collaborative and capacity building projects. Retrieved from http://www.aaupnet.org/aaup-members/news-from-the-member ship/collaborative-publishing-initiatives
Bardi, A., & Manghi, P. (2014). Enhanced publications: Data models and information systems. *LIBER Quarterly*, *23*(4), 240–273.
Bettivia, R. (2016). Where does significance lie: Locating the significant properties of video games in Preserving Virtual Worlds II data. *International Journal of Digital Curation*, *11*(1), 17–32.
Flanders, J. (2014). Rethinking collections. In P. L. Arthur & K. Bode (Eds.), *Advancing digital humanities* (pp. 163–174). Basingstoke, UK: Palgrave Macmillan.
Flanders, J., & Jannidis, F. (2015). *Knowledge organization and data modeling in the humanities*. White paper on the Workshop on Knowledge Organization and Data Modeling in the

Humanities, Brown University, Providence, Rhode Island, March 14–16, 2012. Retrieved from http://www.wwp.northeastern.edu/outreach/conference/kodm2012/flanders_jan nidis_datamodeling.pdf

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*. Retrieved from http:// guide.dhcuration.org/intro/

Housewright, R., Schonfeld, R. C., & Wulfson, K. (2013). *Ithaka S+ R US faculty survey 2012*. New York: ITHAKA. Retrieved from http://digital.library.unt.edu/ark:/67531/metadc234917 /m2/1/high_res_d/Ithaka_S+R_US_Faculty_Survey_2012_FINAL.pdf

Kluge, S. (2000). Empirically grounded construction of types and typologies in qualitative social research. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research, 1*(1). Retrieved from http://www.qualitative-research.net/index.php/fqs/article/view/1124

Koch, T. (2000). Quality-controlled subject gateways: Definitions, typologies, empirical overview. *Online Information Review, 24*(1), 24–34.

Lippincott, S. (Ed.). (2015). *Library publishing directory 2015*. Library Publishing Coalition. Retrieved from http://www.librarypublishing.org/sites/librarypublishing.org/files/docu ments/lpc_dir_2015lpd.pdf

McGlone, J. (2016, October 28). Michigan Publishing announces beta launch of new publishing platform, Fulcrum. Announcements, University of Michigan Press. Retrieved from https://www.publishing.umich.edu/2016/10/28/michigan-publishing-announces-beta -launch-of-new-publishing-platform-fulcrum/

Muñoz, T. (2013). Data curation as publishing for digital humanities. *Journal of Digital Humanities, 2*(3), 13–22.

Palmer, C. L. (2004). Thematic research collections. In S. Schreibman, R. Siemens, & J. Unsworth (Eds.), *A companion to digital humanities*. Oxford: Blackwell. Retrieved from http://www.digi talhumanities.org/companion/view?docId=blackwell/9781405103213/9781405103213 .xml&chunk.id=ss1-4-5&toc.depth=1&toc.id=ss1-4-5&brand=default

Price, K. M. (2009). Edition, project, database, archive, thematic research collection: What's in a name? *Digital Humanities Quarterly, 3*(3). Retrieved from http://www.digitalhumanities .org/dhq/vol/3/3/000053/000053.html

Rosenzweig, R. (2007, July 2). Collaboration and the cyberinfrastructure: Academic collaboration with museums and libraries in the digital era. *First Monday, 12*(7). Retrieved from http://firstmonday.org/ojs/index.php/fm/article/view/1926

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*. Hoboken, NJ: John Wiley & Sons. Retrieved from http://online library.wiley.com/doi/10.1002/9781118680605.ch36/summary

Unsworth, J. (2000, December). *Thematic research collections*. Paper presented at the Modern Languages Association Annual Conference, Washington, D.C. Retrieved from http://www .iath.virginia.edu/~jmu2m/MLA.00/

Katrina Fenlon is a doctoral candidate at the School of Information Sciences, University of Illinois at Urbana-Champaign. Her dissertation research focuses on collections as one among an array of emergent and rapidly evolving genres of digital scholarship that pose new challenges and opportunities for scholarly communication.