

Swarthmore College

Works

English Literature Faculty Works

English Literature

2012

Swarthmore College: The Early Novels Database And Undergraduate Research: A Case Study

Rachel Sagner Buurma

Swarthmore College, rbuurma1@swarthmore.edu

Anna T. Levine , '12

Richard Li , '11

Follow this and additional works at: <https://works.swarthmore.edu/fac-english-lit>



Part of the [English Language and Literature Commons](#)

[Let us know how access to these works benefits you](#)

Recommended Citation

Rachel Sagner Buurma; Anna T. Levine , '12; and Richard Li , '11. (2012). "Swarthmore College: The Early Novels Database And Undergraduate Research: A Case Study". *Past Or Portal? Enhancing Undergraduate Learning Through Special Collections And Archives*. 279-282.

<https://works.swarthmore.edu/fac-english-lit/100>

This work is brought to you for free by Swarthmore College Libraries' Works. It has been accepted for inclusion in English Literature Faculty Works by an authorized administrator of Works. For more information, please contact myworks@swarthmore.edu.

The Early Novels Database and Undergraduate Research: A Case Study

Rachel Sagner Buurma, Anna Tione Levine, and Richard Li

Traditionally, the classroom has been associated with the canon as opposed to the library, with the idea that a special subset of texts selected as both especially excellent and representative (of the literary tradition, say, or a specific cultural moment) are all one can or should teach given the time constraints of traditional college classes and the goals of the liberal arts education. Making the library—even the special collections library—into a classroom in the way our bibliographic database project does therefore raises obvious challenges, ranging from the procedural problem of training undergraduates to do competent descriptive bibliography and library cataloging to the ethical and pedagogical issues surrounding the involvement of undergraduate students in work that many might consider the province of the graduate education. Our case study—co-authored by the database’s faculty director Rachel Sagner Buurma and its undergraduate researcher-catalogers Anna Tione Levine and Richard Li¹—describes a database project to which undergraduate researchers have made a significant and ongoing contribution, and suggests that even projects involving such relatively technical and specialized work have a role to play as part of a liberal arts education.

THE EARLY NOVELS DATABASE: A DESCRIPTION²

The Early Novels Database (END) is a bibliographic database based on the University of Pennsylvania’s Rare Book & Manuscript Library’s extensive collection of fiction in English published between 1660 and 1830. Produced by the collaborative effort of Penn li-

brarians, information technology specialists, faculty from Swarthmore College and Penn, and Swarthmore College undergraduate researchers, the completed database will include richly descriptive records of more than 3,000 novels and fictional narratives, from the very canonical to the almost unknown, from fictions that clearly announce themselves to be novels to the works of fiction (fable, travel narrative, romance) that formed part of that genre’s notoriously murky origins. Users will be able to perform both keyword and faceted searches across bibliographic records containing both edition-specific and copy-specific information about each novel.

We have designed END to complement the extensive existing full-text facsimile archives that contain early novels (such as ECCO, GoogleBook, and the Internet Archive). One of the most significant problems with recent large-scale book digitization projects has been the loss of edition-specific and copy-specific structured metadata—of information *about* and describing the book—of the kind often available in library card catalogs. The absence of this data can make it difficult for scholars and other researchers to find particular novels or sets of novels they are interested in, because even as our archive of digital texts from the seventeenth, eighteenth, and nineteenth centuries has expanded exponentially, our ability to access them in precise, controlled, and complex ways has diminished. While recent projects have begun to take on this challenge—Brian Geiger’s (University of California, Riverside) and Ben Pauley’s (Eastern Connecticut State University) Google-sponsored effort to automatically

match ESTC (English Short Title Catalog) records to GoogleBook items is a notable recent example—our project seeks to use human eyes and brains and hands to create and control bibliographic descriptions in ways that computers cannot. For example, we tag each noun, adjective, person name, place name, and object mentioned in the title of each novel; the resulting information can be keyword searched but also appears as a set of “facets” that display how often a given word in each category appears. Therefore, researchers can not only perform traditional keyword searches of the title field to turn up relevant items, but can also see the entire array of nouns appearing on all title pages sorted alphabetically or by frequency. We also include in-depth information on other aspects of the novel’s paratexts, describing the prefaces, introductions, dedications, indexes, tables of contents, copyright statements in both controlled and more discursive vocabularies.

A scholar interested in the genre of “history,” to take a hypothetical example, can not only instantly call up all 189 records of novels with this noun in the title; she also, at the click of a button, can see that of the records of novels with “history” in the title, 27 of them also include the adjective “young”; that 56 of them have prefaces; that the majority of them are

written in the third rather than the first person; and that four of them profess to be written by women but were in fact penned by men. She can sort and unsort them by year and decade of publication, and notice that most of them are published in London, but that after 1787 many of them also are published in Dublin; she can pull up records of all novels that contain prefaces, and click on each record to see the individual idiosyncratic titles of each one. She can find out instantly that 134 of them have epigraphs on the title pages, and by looking at the authors of those epigraphs she can determine at a glance how many are by “ancient” and how many by “modern” authors. And she can do all of this work in seconds, rather than in the weeks or even months it would take for her to generate this information herself. So while as a bibliographic tool END does not itself make a claim about literary history, or even represent to its users the “insides,” or texts, of the novels it includes, it makes possible the writing of new, alternative histories of the novel.³

THE WORK OF THE UNDERGRADUATE RESEARCHER

The great challenge—and promise—of this kind of project is that it unites the bibliographic description

Figure 8.1. Screenshot of the database front page

Under Development: END: Early Novels Database

Search Collection []

IMPORTANT: This DUA Site is Under Development!

About the Database

END is a bibliographic database based on the Collection of British and American Fiction 1660-1830 held by the University of Pennsylvania's Rare Book & Manuscript Library. When completed, the database will include records of more than 3,000 novels and fictional narratives by canonical authors such as Daniel Defoe to Jane Austen as well as less well-known novelists like Mary Brunton and Mary Walker. Users will be able to perform both keyword and faceted searches across bibliographic records containing both edition-specific and copy-specific information about each novel.

END is designed to complement existing full-text archives. A large (and growing) number of digitized copies of early novels exist online in both open-access and proprietary form, yet one of the well-documented characteristics of new digitization projects has been the loss of precisely the kind of metadata END aims to provide. This means that even as our archive of digital texts from the seventeenth, eighteenth, and nineteenth centuries expands, our ability to access them in precise, controlled, and complex ways has diminished. By offering human-generated and easy-to-manipulate sets of information about early novels, END offers a concrete solution inspired by sophisticated models of text searching that have existed for centuries.

By uniting twenty-first-century database and search technologies with the sensibility of eighteenth-century indexing practices, END allows users to access the collection of novels in a variety of innovative ways. It creates access to the terms, genres, and categories – from titles to indexes – by which the novels name and organize themselves, revealing the kinds of “book information” that allow users to let “the novel itself” speak. At the same time, the database does include extensive stores of more subjective and extra-bibliographic cataloger-created information, allowing users to choose whether to include or exclude such information, or even to search on a range of information outside the other.

Sort Results By:

- Possible Sorts:
 - Title [a-z][A-Z]
 - Year [a-z][A-Z]

Narrow Results By:

- Repository: browse
 - University of Pennsylvania: 703
 - PU: 102
 - Authorized: 1
 - Bryn Mawr College Library: 1
 - Unauthorized: 1
- Month: browse
 - Volume: 139
 - History: 115
 - Author: 82
 - Letter: 77
 - Novel: 71
 - Edition: 64
- view more...
- Adjective: browse
 - Two: 53
 - Several: 31
 - New: 30
 - Young: 26
 - Late: 23
 - Original: 19

of books produced in an extraordinarily idiosyncratic genre and publishing moment with the necessity to “control” the descriptive terms we use in order to make searches across large numbers of records possible and meaningful. To this end student researcher/catalogers Anna Levine and Richard Li have been involved in the process of creating a bibliographic template of information we want to capture, developing a cataloging protocol for the project, writing the glossary, guides to searching, and other website text, in addition to actually cataloging the books. We have had to ensure that Anna and Richard learn “professional-quality” descriptive bibliography, think about bibliographic description and control in sophisticated ways, and manage their time effectively all while engaged in work that can sometimes be mindnumbingly, eyeglazingly boring. At the same time, we’ve had to try to make sure that their work on the project is valuable as a part of a liberal arts education. As we theorize it, this value has to do with the very practical and immediate way the project foregrounds the necessity to aspire to some kind shared and transparent standard of description while simultaneously acknowledging (and even being suspicious of) the difficulty or impossibility of this as a perfected project.

To work on the Early Novels Database, to spend a summer (or two summers) of days creating detailed bibliographic descriptions of novels as Anna and Richard have, is to be involved in an ongoing demonstration of the impossibility of reducing even the paratext of the novel to a standard formula, and a continual rediscovery of the singularity (and resistance to full description) of the material text. Working on the database also required that Anna and Richard not only to do research themselves (for example, to determine epigraph’s author, to verify book format, to authorize a name), but also that they learn to imagine what kinds of searches in which researchers using the database might be interested. The process of creating the bibliographic template offers one example. Anna and Richard not

only helped decide what information to enter in the database, but even worked to make decisions about standardized terminology, format, and design. The necessary work of familiarizing themselves with existing conventions and how they exist in the novels we are working with has been both a challenge and a learning opportunity. In some cases, this is a mere matter of learning new terminology in order to be able to identify and name half-titles, for example, or subscriber lists. In other cases, terminology is complex, ambiguous, or nonexistent; in these case the project team worked together to create terms that are intuitive and transparent to researchers as well as faithful to the books surveyed. A classic example of a conflict between an “intuitive” classification and a “faithful” one arises in the case of recording the titles of the paratextual essays appearing in each novel. Since the terms that early novels’ paratexts use to name themselves (for example “advertisement,” “dedication,” and “postscript”) are often idiosyncratic (even by eighteenth-century standards) and inconsistent, we want to capture that variation and diversity. Yet we also need researchers to be able to quickly sort books by the types of paratexts they contain. We solved the problem by creating three related fields: a field containing a controlled set yet expandable set of terms early novels use to describe their paratextual essays, a notes field offering less expert database users a “translation” when necessary, and a field that gives the title of each essay verbatim. But the creation of controlled terms is—problematically—never-ending; since each book cataloged potentially enlarges the range of types of paratextual essays we know about, Anna and Richard must be able to make on-the-spot decisions about adding new terms. And, of course, they have to know when and how to ask expert advice—from a librarian, a professor, a reference work—when necessary.

END’s design, as we quickly discovered, offers wide latitude for the undergraduate researcher to explore her own interests without introducing unwanted clutter, since it is possible for us to tag char-

acteristics of books in the bibliographic records which our program and web editor can then exclude from the website display. Anna Levine took full advantage of this built-in opportunity to experiment without consequences. After becoming interested in the historical relationship between epistolary fiction and third-person narration in her first-year seminar at Swarthmore, she decided to create a field that tracks the narrative form of each novel (as determined—admittedly imperfectly—by a few minutes of scanning). Under the guidance of Rachel, Anna organized the experimental narrative form field into two main subfields: the first, \$a, denotes the primary narrative form of the text (if one exists), while the second subfield, \$b, contains narrative forms within the volume that are not expressed in \$a. This categorical system has proven itself useful in many cases: if a novel is written completely in letters, for example, but between the fictional exchange of letters there exists third-person narration (as in Richardson's *Pamela* to take the most canonical example), we would note in \$a that the primary narrative form is "epistolary"; we would then record the existence of third-person narration in \$b. In this situation, the subfield system is quite useful: \$b allows and highlights narratorial exceptions, while \$a honors the preeminence of the primary narrative form.

Because of the room for experimentation that the END inherently allows, we found that it was easy to incorporate Anna's experimental field into the project without compromising any other aspects of the END. We have also found, after working closely together to develop a comprehensive set of terms to define the narrative intricacies of early novels, that the narrative forms field seems relevant enough to the project to make it into a facet on the website. In these ways and in many others, END's undergraduate researchers continue to make significant contributions to the shape of the project while learning ways of thinking and skills that hopefully have relevance to life beyond END.

NOTES

1. Anna and Richard's work on END has been made possible by the generous support of Swarthmore College's summer humanities research grant program and by a Hungerford grant from the office of the Provost at Swarthmore.
2. END would not have been possible without the unwavering support and concerted efforts of the following individuals: Lynne Farrington Curator of Printed Books, Rare Book and Manuscript Library, University of Pennsylvania; Michael Gamer Associate Professor of English, University of Pennsylvania; Heather Glaser, Curator and Assistant Fine Arts Librarian, Fisher Fine Arts Library, University of Pennsylvania; David McKnight Director, Rare Book and Manuscript Library, University of Pennsylvania; Dennis Mullen, Web Developer and Designer, Van Pelt Library, University of Pennsylvania; Jon Shaw Head, Research, Training and Quality Management, Van Pelt Library, University of Pennsylvania; Laurie Sutherland, Metadata Specialist, Van Pelt Library, University of Pennsylvania; Leslie Vallhonrat, Web Managing Editor, Van Pelt Library, University of Pennsylvania. View the database at: <http://hdl.library.upenn.edu/1017/88396>
3. While END is in many ways a database of information designed to give researchers a "middle distance" view of the novel (as opposed to enabling the kind of "distant reading" of visualized large-scale sets of information about the novel which Franco Moretti and others are interested in), some of the types of macroscopic information included may eventually lend itself naturally to graphical representation. (See Franco Moretti, *Graphs, Maps, Trees: Abstract Models for a Literary History* (London; New York, Verso), 2005.) 2005. Eventually, for example, END may be able to map the frequency of epigraphs against a timeline, or even more specifically, the frequency of quotations from Shakespeare used as epigraphs against a timeline.