



Home Economics Archive: Research, Tradition and History

Cornell University

www.hearth.library.cornell.edu

The Ithaka S+R Case Studies in Sustainability were conducted in partnership with the Association of Research Libraries and were made possible by the generous support of the Institute of Museum and Library Services. A summary report, “Searching for Sustainability: Strategies from Eight Digitized Special Collections,” is available at <http://www.sr.ithaka.org/research-publications/searching-sustainability>. This specific case study was researched and written by Sarah Pickle, Analyst, Ithaka S+R.

ABSTRACT

The Home Economics Archive: Research, Tradition and History (HEARTH) is a digitized collection of academic and popular monographs and journals comprising the core literature of home economics, or, as it is more commonly known today, human ecology. Created at Cornell University’s Mann Library, which serves



primarily Cornell’s College of Agriculture and Life Sciences and the College of Human Ecology, HEARTH was launched in 2003 with the support of a two-year IMLS National Leadership grant. Today, it is one of Cornell’s most popular online collections, and

all ongoing support for it is provided by either Mann Library or the Cornell University Library system. This internally sponsored ongoing development is possible because the user-friendly back-end program created by the HEARTH programmer enables inexperienced staff to efficiently and inexpensively add to the collection materials from an already selected core-literature bibliography.

INTRODUCTION

Home economics is a field long derided by many who believe it has pigeonholed women as servants of the domestic realm rather than valuable forces outside of it. But the digitized special collection Home Economics Archive: Research, Tradition and History, created by staff at Cornell University’s Mann Library, aims to dispel those stereotypes by preserving and sharing the most important publications related to the early areas of study of Cornell’s College of Human Ecology.¹ The collection presents more than 770,000 pages of academic and popular monographs and journals published between 1850 and 1950 representing the core literature of the field—much of which consists of scholarly articles written by women—as well as bibliographies and essays on subjects from hygiene and applied arts to consumer studies and institutional management. As a corpus, HEARTH demonstrates that, rather than a barrier, home economics was a gateway into academia for many

women. Since the completion of its IMLS National Leadership Grant in 2003, the collection has grown by almost 80 percent, which is due in no small part to the fact that a processing and formatting system created by the project’s programmer makes it an extremely efficient operation to run; as a result, Cornell has been able to support internally all ongoing development.

Cornell University is a sprawling institution with twenty libraries that serve fourteen colleges and schools in several locations throughout New York State, and it is funded both by a private endowment and by state (i.e., public) land-grant money. Mann Library rests atop East Hill in Ithaca, New York, and an extensive building renovation that built out its collaborative and study spaces has turned it into a major hub on campus. For many years, the Library has reported to administrators at the two colleges it mainly serves, the College of Agriculture and Life Sciences and the College of Human Ecology (until 1969, known as the College of Home Economics), in addition to reporting up through the Cornell University Libraries.² Its role has been supporting students and scholars at these public colleges supported by public funds, and, as a consequence, it views both the campus community and the citizens of New York State as its key patrons.

BACKGROUND

Until 2013, when its budget began coming directly from the central Cornell University Libraries (CUL), the Mann Library received its budget from the Colleges of Agriculture and Life Sciences and of Human Ecology. According to some, this may have made it easier for former Library Director Janet McCue to set her own priorities for her campus unit, including the digitization of the Library’s collections, in which she was motivated predominantly by academic and preservationist concerns.

McCue believed that Mann Library should create carefully curated online collections in order to preserve the most important historical literature from the fields it serves—agriculture and home economics/human ecology. In addition to reflecting concerns regarding the age and increasing fragility of the materials, her scholarly approach was to develop collections whose outlines were determined by the scope of the field being studied, not limited by the collections on hand. To identify which materials were most important, Mann Library staff assembled committees of experts, including scholars beyond the university, to help them through the process. As a result, current Director Mary Ochs sees each of the digitized collections created by the Library as “a goldmine of related content,” rather than simply a pile of assets produced from mass digitization of as much material from the field as possible or from scanning stand-alone, bounded special collections.



Courtesy of the Home Economics Archive: Research, Tradition and History.

Mann Library’s first digitized collection, which established a model that HEARTH later followed, was the Core Historical Literature of Agriculture (CHLA), created in 1994 to combat the “brittle books problem” through the preservation of fragile and disintegrating print materials in digital form.³ The “core literature” approach guiding the project team’s selection process was developed by Wallace Olsen, a former Senior Research Associate at Mann Library.⁴ Additionally, because it was around this time that the internet was becoming an increasingly useful tool for scholarly research, the team responsible for CHLA, including McCue and then Head of Collection Development Ochs, decided to post the materials online in 1996, both in order to assemble them in one place, as there was no distinct physical counterpart to CHLA—the scanned originals were scattered throughout Mann Library’s holdings—and to make them more accessible to scholars unable to visit the Library. With the digitization of the core literature of agriculture well underway, McCue and Ochs turned their attention

to the other college with which Mann Library works, Human Ecology, and the most historically important documents to emerge from that field.

HEARTH began as the Core Historical Literature for Home Economics (CHLHE), and its development directly paralleled the methodology behind the creation of CHLA. As with the earlier collection, the protocol for choosing the materials to be included in HEARTH followed Olsen's approach, and the list that

Ochs and Paulson aimed for efficiency, which is evident in their choices to work with the preset bibliography and to train students to do nonspecialist work

resulted reflects months of careful planning by the project coordinators. For HEARTH, Ochs and current Project Manager Joy Paulson compiled an extensive list of monographs and journals using their own local experts, bibliographies in the backs of books, and any other resources they could find.

These lists were later sent for evaluation to scholars of home economics and history from across the country who had been recruited as advisors; they ranked the sources listed from one to four based on relative value and importance. Once the evaluations were returned to Mann Library, the results were tabulated and a bibliography of core literature was compiled.

Early research revealed that the Cornell Libraries held only about half of the materials listed in the bibliography, so Ochs first sent Paulson out to used and antique bookstores to locate and purchase missing titles. Later, when digitization practices had developed to the extent that images could be made with little to no damage to the physical object, the project team took advantage of preexisting relationships with their counterpart libraries at the University of California at Berkeley, the University of Minnesota, the University of Nebraska at Lincoln, and Pennsylvania State University, whose own holdings could help fill out the digital collection. Altogether, about 10 percent of the materials currently in HEARTH were contributed free of charge by these other institutions, a collaborative practice that Paulson believes dates back to the Cooperative Preservation and Microfilming Project of the 1980s and '90s.

As Ochs and Paulson began to compile the collection, tracking down books and serials from across the country and creating basic bibliographic records for materials previously lacking such records, they were



Courtesy Division of Rare and Manuscript Collections, Cornell University Library.

well aware that much of the literature they were seeking was unknown to most researchers and that their work was not motivated by a great demand in their community for the core bibliography of home economics. Instead, their project was primarily an academic one, as their work could help preserve and build awareness of these resources as well as inform visitors about the history of the field. As Paulson explained, HEARTH was less a way to reach out to current scholars of Human Ecology, who work with contemporary research, than it was to inform social historians, historians of science, scholars of cultural studies, and other humanists about the discipline and thereby increase its visibility and improve public perception, as many of these materials were under threat of being lost due to degradation and de-selection.

Mann Library's position within the greater nest of the Cornell Library system was also beneficial. Cornell's Olin Library was home to Anne Kenney and Oya Rieger, early pioneers in developing best practices for digitization and digital preservation; working up the hill from them meant that Ochs and Paulson felt confident of the quality and durability of the surrogates they produced. They were generating stable TIFF master files of a much higher quality than is likely to ever be displayed on a computer screen, so only the delivery images—rather that

the original files themselves—need to be changed when upgrades are made to keep pace with developments in technology.

As HEARTH was developing Ochs and Paulson aimed for efficiency, which is evident in their choices to work with the preset bibliography and to train students to do nonspecialist work, thereby freeing up librarians to confront bigger challenges when necessary. Desire for efficiency also led the project team to focus almost exclusively on digitizing public domain materials (published before 1923) during the grant period, as working on CHLA had demonstrated to them just how much time can be spent on chasing down copyright permissions.

The initial grant application to IMLS for HEARTH was submitted in 2000 and was one of Paulson’s very first assignments after arriving at Cornell from the University of Michigan in 1999. The two-year, \$277,311 grant, for which Mann Library received a one-year no-cost extension, covered the following:

- Mary Ochs, Principal Investigator: 5 percent of time
- Head of Mann Library information technology section: 5 percent
- Joy Paulson, Project Manager: 25 percent
- Frances Webb, Programmer: 75 percent
- Metadata librarian: 5 percent
- Copy cataloger: 50 percent
- Collection development librarian/website designer: 10 percent (one year only)
- Preservation assistant: 100 percent

This diverse cast meant that all stages of the development of the project benefited from a range of technical skills and subject expertise. Almost all work related to the creation of the collection was done in Mann Library; scanning, which was also funded by the grant, was conducted by a firm in Montreal that was so cost- and time-efficient that almost 40 percent more pages were scanned than the 450,000 pages the grant proposal had anticipated. Including all activities necessary to add a single page to the site (digitizing, ingesting, record writing, copyright clearing, uploading, and preserving), Paulson calculated that HEARTH’s per-volume cost totaled \$262.42 and that its per-page cost was \$0.65—30 percent less than she and Ochs had anticipated. Scanning alone came to about \$0.17 a page.

In the coming years, Paulson would like to continue working through the bibliography of core literature and perhaps even expand the types of media available on the site to include videos and pictorial collections. As long as the Library keeps posting new materials and Paulson and Webb see web traffic demonstrating that those materials are being used, Paulson believes Mann Library will continue to support this work.

SUSTAINABILITY STRATEGY

HEARTH’s project team is able to sustain the work it does by operating at a low simmer, adding and highlighting new materials when they have the time and money to produce them, which is typically at the rate of about 150 titles per year. Since the IMLS grant ended in 2003, this approach has permitted them to chip away, little by little, at the titles in the bibliography that remain to be digitized, at this time roughly 66 percent of the full list.

Economic Model

Cornell University has assumed all costs for HEARTH, including what is required to continue adding to the collection. Mann Library has control over a small endowment and special gifts from graduating classes, and

Paulson and Ochs can allot these funds to scanning for HEARTH when the money is not needed for more urgent tasks. Furthermore, all funds provided to HEARTH can be devoted to scanning because, since 2003, the remaining core staff of the project—Paulson, Webb, and several others—are regular, hard-money employees of the Library, and ingesting files for incorporation in the collection is part of their typical responsibilities.

The figures below represent the total amount of time each individual spends on all three of the Library’s digital collections—HEARTH, CHLA, and The Hive and the Honeybee⁵—combined, as it would be too difficult to tease out the percentages of time each devotes to HEARTH alone.

- Joy Paulson, Project Manager: 10 to 25 percent of time
- Frances Webb, Programmer: 25 percent
- Carol Lowe, Preservation Assistant: 50 percent
- Copy cataloger: 5 percent

To be sure, HEARTH and CHLA are quite extensive, but working with only three collections and being on hard money has permitted Paulson and Webb to serve as genuine project managers and to dedicate time to thinking about how to improve them should they ever have the means to do so. Webb also follows user statistics and performs general maintenance and ingest of new materials. This sustained concentration on just a few projects is rare in the CUL system, according to Associate University Librarian Oya Rieger, and Mann Library Director Ochs is not interested in shifting her Library’s priorities any time soon.

Other costs for supporting the ongoing life of HEARTH are covered internally, as well. The virtual storage for the image files and data are paid for by Mann Library, and because the front-end DLXS interface is used by all digitized special collections created by Cornell Libraries, support for it is provided by the Libraries’ central IT department.

In light of these continuing costs, Chief Technology Strategist for Cornell Libraries Dean Krafft has noted that the libraries have considered generating revenue by licensing images from this and other collections created at Cornell, but he also confided that he and other staff were not confident that doing so would be worth the effort. Ochs explained further that any attempts to earn money from HEARTH would have to be compatible with IMLS stipulations that the materials produced from their grants be made freely available; she was also hesitant at the suggestion of advertising or other open-access-compatible models of revenue generation because of the Mann Library’s status as a library serving a land-grant institution.

Public Benefit

Ochs and Paulson have spoken about the HEARTH team’s accomplishments in internal workshops and at external conferences, focusing most often on their success with a serials processing system and with project management. Additionally, both Paulson and Ochs pride themselves not only on the selection process that produced the bibliography, but also on HEARTH’s usability and accessibility as a carefully curated collection with structural metadata and searchable text. They know that some students and faculty on Cornell’s campus have used the resource, although fewer than they had hoped when they created the site; and spikes in numbers of non-Cornell visitors during school semesters suggest that researchers on other campuses may have found their way to it, as well.

One measure of HEARTH’s success is its number of visits by individual users (over 13,000 visits per month, on average), which is several thousand higher than that of CHLA. Additionally, almost half those sessions represent overseas users, and the most common referral sites (StumbleUpon, Google, FoodTimeline.org) have

no affiliation with Cornell. This suggests that the reach of HEARTH is relatively broad, extending far beyond the campus, and it has encouraged the project team to add more popular titles, such as Harper's Bazaar and Good Housekeeping, in an attempt to keep attracting nonacademic visitors. HEARTH is also a success in the context of CUL, with the second-highest number of page views (1.2 million) among all digitized Cornell Libraries collections, almost twice as many as the third and fourth (the Cornell Daily Sun's digital archive and CHLA) most viewed collections.

Web analytics are helpful for gaining a sense of how many people are visiting and where they are coming from, but the numbers cannot reveal much about why people are using the collection. For that kind of qualitative feedback, HEARTH staff rely on what they hear from academics on campus and from a form on the site that sends comments and questions to Mann Library's circulation desk, whose staff forwards them to either Paulson or Webb. The project team receives about ten of these comment forms a week, often several from nonacademic visitors, who have written to describe how helpful the resource is. Among this group are genealogists who have discovered relatives' names in the texts, which have been made searchable via the creation of OCR files; and staff from the Oxford English Dictionary, who prior to the advent of Google Books uncovered several examples of early instances of specific words that they could use in the dictionary's annotations. An OED editor wrote to suggest that the project add search limits by date, a change that has since been implemented. Other users have requested direct access to the OCR files, and users do have the option of viewing plain, uncorrected OCR text.

KEY FACTORS OF SUSTAINABILITY

- 1. A clear editorial strategy.** Although the ongoing support staff for HEARTH and Mann Library's other two digitized special collections is small, Paulson and Webb do not find it challenging to manage their collections, in large part because of the core literature strategy. Because a list of priority content was developed in the early stages of the project, virtually no time need be devoted to editorial decision making today. When money is available for digitization and staff have time to commit to the collection, the team has a list from which to work, and they know where to locate the materials. The current online collection includes essays on various subjects within home economics, the core literature bibliography itself, and digitized materials from Mann Library and its counterpart institutions, comprising close to one-third of the overall bibliography. Since the end of the grant, the project team has reached far beyond the 1923 public domain cutoff date, adding books and serials released as recently as 1997. Journals range from the popular Harper's Bazaar (issues from 1867 to 1900 are included) and Good Housekeeping (from 1885 to 1950) to the academic Journal of Home Economics (1909 to 1993), and monographs include titles such as Letters to the People on Health and Happiness (1856) by Catharine Esther Beecher and The Business of Being a Woman (1921) by muckraker Ida Tarbell. Paulson, who is now the Digital Collections Librarian at Mann Library, continues to add new titles from the bibliography to the collection when money is available or when requests are submitted by researchers. About one thousand of the current eighteen hundred volumes available on the website were digitized with support from the initial IMLS grant; the remainder of the work has been covered internally by Mann Library.
- 2. A user-friendly back-end program.** The planning that went into the creation of a program that could bundle all data and files in the collection and support the routine ingest of items to the website has been HEARTH's greatest asset and has made it possible for the remaining project team to gradually add more to the site as the years pass. Because CUL requires that every digital project have an IT person assigned to it, programmer Webb has been on the HEARTH project from the beginning, and the system she developed has led to a workflow that is user friendly for nonlibrarians and requires little learning time from a serials cataloguer or technical staff member outside of Collection Development.

While planning the collection, Ochs and Paulson felt comfortable that they knew their target audience and the technical processes and procedures involved in digitization; less clear to them was the answer to a problem that many librarians must confront, whether they are working with digital or analog collections: how to write elaborate and consistent collections records that can capture the idiosyncrasies of older publications. Monograph metadata were relatively simple to compose, but the arbitrariness and occasional inconsistencies of journal volume and series numbers meant that the process of writing their metadata was challenging and labor intensive. Had the library had the means to assign an additional group of expert librarians to work on the project, and had Paulson and Ochs not had other responsibilities to tend to, the project team might not have had to rely on work-study students for some of the work involved in creating metadata; but because this was their situation, Webb devised a back-end interface that enabled unskilled employees—work-study students, in this case—to write structural metadata (see figures 1 and 2).

Once all the metadata are written, Webb’s program also makes it possible for the ingest and harmonizing of files, quality control checks, and copyright permissions to be conducted and reviewed in her system before an overnight upload of the data to the web. Paulson expects that this “bundling” of images and related information before delivery to the public platform means that the digital assets that make up HEARTH are not tied to the DLXS proprietary system and thus can be easily migrated to other platforms.

3. **Position within a larger, experienced institution.** With both of these strategies in place, as well as the support the HEARTH team receives from CUL IT, which manages the hardware and DLXS platform, the only ongoing development costs for which Paulson has to explicitly appeal for help are those associated with scanning. This stable situation puts the project team in a good position to respond to the few requests from scholars that they receive each month. And should Paulson, who has spent such significant time creating and supporting this collection, ever have to pass the baton of project leader, she is confident that the combination of the simple back-end program and the bibliography would make it easy to integrate new staff.



Figure 1. A screenshot of Webb’s back-end program for writing structural metadata for popular serials in HEARTH. Each row represents an article and is colored for easy reading; each cluster of rows beneath a page icon represents a scanned page (i.e., page 4 has five articles, beginning with “Planning Food Officials...”). Pages may be flagged for special features, such as “Table of Contents” or “Table of Illustrations,” which are highlighted on the webpage for the entry (see figure 2). Figure 2. A screenshot of the webpage for American food journal (8.1).

The successes of HEARTH owe a great deal to CHLA, which also had an audience mainly made up of humanists and which provided Ochs, Paulson, and Webb with a great deal of experience not only in creating a digitized collection before many other institutions had done so, but also in planning for one that was carefully curated to represent the fundamental texts of a discipline and that required careful work with extremely delicate materials. Additionally, CHLA had been a lesson in how to anticipate and avoid certain difficulties, including verifying copyright clearances before scanning and ensuring that all images are stored in a single location.

ONGOING SUSTAINABILITY ISSUES

- **Imminent migration to a new platform.** Within the next two years, Cornell Libraries will have to migrate all of its digitized special collections off the DLXS platforms when the University of Michigan, which developed the software, will end its support. But Paulson believes that this transition should be easy for HEARTH, as none of the collection's files are in proprietary formats until they are bundled together for delivery to DLXS. Her only worry is that this period of migration might turn into a missed opportunity to further enhance HEARTH by adding multimedia or integrating social media.
- **Modest funding going forward.** As Ochs put it, the digital collection's home site is "beyond middle age as websites go," but any changes beyond migration to the new platform, which will be supported by the Library, would require outside funding. In 2005, Paulson attempted to secure another IMLS grant for a second intensive round of digitization, but her application was unsuccessful. In her experience, in the age of mass digitization by companies like Google, funds are scarce for project leaders who are looking for more money merely to add more pages to their collections; innovation is key.
- **Little understanding of users.** The project team's knowledge of HEARTH's current users is still quite foggy. Although Paulson already has a strategy for how to better understand who visits the site and how they use the information provided—a report on the project written in July 2003 outlines a plan involving eight different measures (e.g., number of classes including HEARTH in syllabi and number of students in the classes) all matched with target groups, information sources, and intervals at which that information could be collected—that strategy has never been put in action.

LESSONS FOR OTHERS

1. **An editorial strategy can offer coherence for researchers and a roadmap for future growth.** A coherent, topic-driven editorial strategy can help both in project planning and in shaping a collection to be of great interest to scholars and students. It can also enable efficient, opportunistic, and gradual additions to a collection that may struggle to secure external awards.
2. **Partnerships can add to the richness of a resource.** Should the editorial strategy allow it, partnerships with like-minded institutions can help to ease the burden of creating collections by distributing responsibility across multiple institutions.
3. **Clear work protocols and a user-friendly back end can enable greater participation.** A well-designed program for inputting metadata can structure and formalize a workflow that can enable inexpensive nonspecialists to work quickly while minimizing risk of error. Dropdown menus for volume and issue numbers for serials help avoid input mistakes, copying and pasting OCR'd texts helps reduce typos, and color-coding for each article in an issue helps make the program's input screen more legible (see figure 1).

APPENDIX

Interviewees

Dean B. Krafft, Chief Technology Strategist, Cornell University Libraries

Mary Ochs, Director, Mann Library

Joy Paulson, Digital Collections Librarian, Mann Library

Oya Rieger, Associate University Librarian for Digital Scholarship and Preservation Services, Cornell University Libraries

Frances Webb, Applications Programmer, Mann Library

Resources Consulted

Albert R. Mann Library—Strategic plan (2013-2015) (March 2013).

The HEARTH Team. Final Report. HEARTH (July 2003).

NOTES

- 1 Arthur R. Mann Library, Cornell University, Home Economics Archive: Research, Tradition, History, <http://hearth.library.cornell.edu/>.
- 2 Beginning in summer 2013, Mann Library reports directly to the Cornell University Libraries.
- 3 Arthur R. Mann Library, Cornell University, The Core Historical Literature of Agriculture, <http://chla.library.cornell.edu/>.
- 4 Wallace C. Olsen, “The Current Core Literature of the Agricultural Sciences,” *Journal of Agricultural and Food Information* 2, no. 3 (1994): 99–109.
- 5 Arthur R. Mann Library, Cornell University, The Hive and the Honeybee: Selections from the E. E. Phillips Beekeeping Collection at Mann Library, <http://bees.library.cornell.edu/>.

This project was made possible in part by the Institute of Museum and Library Services, LG-06-10-0189-10.



Copyright 2013 Association of Research Libraries. This work is licensed under a Creative Commons Attribution-Non Commercial-ShareAlike 3.0 United States License. To view a copy of the license, please see <http://creativecommons.org/licenses/by-nc-sa/3.0/us/>.