Unique Collections and Digital Humanities Initiatives: From Concept to Creation—Exploration and Practice at the University of Pittsburgh Library System

Edward Galloway
University of Pittsburgh

Haihui Zhang
University of Pittsburgh

Follow this and additional works at: https://scholarsarchive.byu.edu/jeal

Part of the Asian History Commons, Collection Development and Management Commons, Digital Humanities Commons, and the Oral History Commons

BYU ScholarsArchive Citation
Available at: https://scholarsarchive.byu.edu/jeal/vol2023/iss177/3

This Report is brought to you for free and open access by the Journals at BYU ScholarsArchive. It has been accepted for inclusion in Journal of East Asian Libraries by an authorized editor of BYU ScholarsArchive. For more information, please contact ellen_amatangelo@byu.edu.
Unique Collections and Digital Humanities Initiatives: From Concept to Creation—Exploration and Practice at the University of Pittsburgh Library System

Cover Page Footnote
Edward Galloway is Associate University Librarian for Archives & Special Collections, University of Pittsburgh Library System. Haihui Zhang, Former Head, East Asian Library, University of Pittsburgh Library System.

This report is available in Journal of East Asian Libraries: https://scholarsarchive.byu.edu/jeal/vol2023/iss177/3
Acquiring unique collections and creating digital humanity (DH) projects have been two vitally important issues in the field of librarianship over the last few years. Libraries face the challenge of not only seeking to acquire unique collections through purchases or donations, but how to create and mine those collections for potential data that could be beneficial for students, researchers, and scholars. The University Library System (ULS) at the University of Pittsburgh (Pitt) has been critically engaged in this area, especially within our East Asian Library (EAL), through the creation of The Cultural Revolution: 10 Project (CR/10) and Contemporary Chinese Village Gazetteer Data Project (CCVG Data). This article will examine these two projects as examples to demonstrate how such collections were leveraged to harness their original documentation and how the library transformed them into DH projects with video and datasets. Topics discussed in the article will cover conditions for creating and acquiring these unique materials, library senior administrative staff’s decision-making and support process, copyright and privacy issues involved, the collaboration of internal and external stakeholders as well as librarians’ and scholars’ contributions, the new role that librarians can play, and the support needed to fund these opportunities. Some of the challenges and difficulties we encountered will also be discussed.

Reflecting on the University Library System’s long-range goals over the past two decades reveals a strong focus on the establishment of unique collections, the promotion of scholarly communication, and the development of digital collections for global dissemination. It states, “In 1997, a Digital Research Library (DRL) was established to augment the networked delivery of commercial electronic texts, as well as to create e-text projects in collaboration with faculty and graduate students.”¹ Later, one of the library’s goals for 2003–2006 was to digitize and deliver electronic versions of print resources to a large body of users worldwide. From 2007–2010, the goal became clearer, and a specific plan was articulated to digitize materials from our unique or endangered collections and provide innovative tools to enhance their availability, access, and continued use. A key goal became providing the technology infrastructure and expertise to support the creation of new digital collections. “Building distinctive unique collections” was listed as one of the library’s Information Resource and Collection Goals in 2013.² In the following years, “knowledge creation” appeared repeatedly in our goal statements stressing that the library should not only collect

---

¹ Michael Dabrishus and Haihui Zhang, “East Meets West: Digital Humanities at the University of Pittsburgh.” In Beyond the Book: Unique and Rare Primary Sources for East Asian Studies Collected in North America, ed. Jidong Yang (Ann Arbor: Association for Asian Studies, 2022), 338.
² ULS FY13 Objectives and Strategies.
information but also create knowledge. Further, the library said it would “Actively collect and preserve scholarly and cultural materials that amplify under-represented voices and global perspectives” and “Be a leader in the active, trusted stewardship of the evolving scholarly and cultural record.” With the support of library senior administrative staff, the East Asian Library has been exploring and executing different ways to build unique collections and initiate DH projects over the past several years. Through this effort, we have gained much experience and have seen the impact that such work can make on teaching and research.

What are the CR/10 and CCVG Data Projects?

The digital humanities (DH) field has grown over the last decade as digital resources become more commonly available to scholars, historians, and students. Most DH projects typically comprise digitized primary sources, text analysis, data mining, mapping, or 3D modeling. The DH projects carried out by the East Asian Library consist of born-digital video collection (CR/10) and data extraction, mining, and analysis (CCVG Data).

Created by the East Asian Library, the Cultural Revolution: 10, referred to as The CR/10 Project is a collection of oral history video interviews with regular, everyday Chinese citizens to create an aural archive documenting the range of stories that ordinary people experienced during the Cultural Revolution. The “10” in The CR/10 Project is a double meaning: it represents both the 10 years of the era of the Great Proletarian Cultural Revolution (1966–1976) and the 10 minutes allotted to each interviewee (although they often spoke for more than 10 minutes). Since the project was being undertaken by an academic library, we strove to maintain a neutral, objective stance in the interviews, rather than leading the interviewees where we thought it would be useful for them to go in their recollections.

Staff within the East Asian Library began interviewing individuals in late 2015, and two years later, the open-access digital collection debuted on a library website ([https://culturalrevolution.pitt.edu/](https://culturalrevolution.pitt.edu/)). Presently, 124 interviews have been conducted or collected and are accessible on the website. CR/10 is an ongoing project, and we continue to accept interviewees to continually build and expand the collection. This work entails transcribing the interviews into Chinese text and creating English subtitles. Future work promises to adopt advanced technology for improved search and retrieval.

The CR/10 Project was enthusiastically received when it opened to the public in October 2017. Interview videos were used in the classroom by Pitt instructors not only in a course on modern China, but in other fields such as sociology, political science, and gender studies in and outside of the university. In 2018, Dr Diana Wood, the Director of the National Consortium for Teaching about Asia (NCTA), led a group of history teachers at high schools in Pennsylvania, Maryland, West Virginia, and Delaware to discuss how to utilize the CR/10 interviews in high school classrooms. Several World History teachers commented by saying:

---

4 ULS 2021 Prioritized Strategies.
“Amazing value since each video has transcripts.”
“Greatest gift is that students have access to wide-ranging personal accounts.”
“Provides an all-encompassing picture of the Cultural Revolution.”
“Able to compare adult and youth perspectives.”
“A lesson in comprehending an historical event and shows the complexity of studying history.”
“Will use interviews to compare the US Civil Rights movement and the Cultural Revolution.”

In 2019, with the support from the Henry Luce Foundation, the University Library System produced a documentary film, *Unreconciled Memories: Reflections on China’s Cultural Revolution*, which is based on interviews collected by The CR/10 Project ([https://culturalrevolution.pitt.edu/#documentary](https://culturalrevolution.pitt.edu/#documentary)). The video interviews are considered a primary source that has led to the creation of another more suitable teaching format.

**Contemporary Chinese Village Gazetteers Data**
The *Contemporary Chinese Village Gazetteers Data Project*, known as the CCVG Data Project ([https://chinesevillagedata.library.pitt.edu](https://chinesevillagedata.library.pitt.edu)), aims to create a web repository that is openly accessible online for specific data extracted from the Chinese village gazetteers. There are two types of villages in China: administrative villages (行政村) and natural villages (自然村). Usually, an administrative village manages natural villages nearby. By November 2022, data on 2,701 villages had been extracted from village gazetteers. To ensure data consistency on Administrative agencies at the same level, in some cases, data from 100 gazetteers of natural villages were removed. East Asian Library staff and students entered data from 2,601 administrative village gazetteers into a customized database platform developed by the library’s IT department. A total of 18 categories of data are included in the datasets with 147 subcategories covering administrative village information, Natural environment, Natural disasters, Last names, Ethnic groups, Population, and migration, Military, Politics and Management, Economy, Family planning, Education, etc.

A Chinese administrative village gazetteer is a grassroots-level primary document that contains both qualitative and quantitative data on the administrative village, the lowest administrative unit in rural China. Yuanziyi Zhang, a University of Pittsburgh graduate affiliated with the Asian Studies Center, wrote, “Current sources of China’s rural data are problematic in the following ways. First, the sample size of most of the databases in the field is relatively small, which in practice often leads to large estimation errors. Second, for the few databases that do contain an adequate amount of data, the sources are usually inclusive (documents of different types) and thus it is difficult for users to conduct either horizontal or vertical analysis without admitting sufficient assumptions.”

Since the library holds the largest collection of Chinese administrative village gazetteers outside mainland China, the East Asian Library initiated the CCVG Data Project in response to scholars’ research needs. We received initial funding from the Office of the Chancellor in 2018 as a Pitt Seed Project.

---

As the project gained momentum, it was further supported by the Asian Studies Center and the University Center for International Studies at the University of Pittsburgh. It took five years from when the project was initiated in 2018 to 2023 when we released data on 2,601 villages.

The dataset is available for download in a .csv format on the CCVG Data website. An interactive map with basic administrative information on each collected village is also included on the website. An interactive search platform with filtering, cross-searching and visualization functions will also be released soon. The CCVG Data Project has received much attention from scholars as well as librarians since we uploaded the data extracted from the first 500 village gazetteers and released the beta site in 2019. After the website launched, we were invited to teach several workshops at a variety of institutions. Each workshop has had very lively discussions in the Q&A session and information exchanges. Dr Kwok-leong Tang, the Digital China Fellow of the Fairbank Center for Chinese Studies from Harvard, wrote “The workshop was really great. The discussion was one of the best in the webinars I have hosted and organized.”

Collaboration is the key

Whether it is the CR/10 or CCVG Data project, collaboration was the key and most important factor to ensure the success of this work. From the start, it was clear that multiple library units would need to be involved in CR/10. Even though the interviews were conducted by the East Asian Library, the CR/10 digital collection could not have been created without the help of the ULS Information Technology unit and the Digital Collections Coordinator. The ULS Metadata and Discovery unit was another key collaborator who helped make the CR/10 interviews discoverable within the library online catalog. The ULS Web Services and Communications department built the visually stunning public-facing website, which introduces the background of the Cultural Revolution, explains the project’s motivation, and directs users to the online collection. In the same way, the CCVG Data project’s success resulted from highly collaborative work with even more internal units involved, including the Library Management Team, Acquisitions Management Unit, Catalog Management Unit, Collection Storage, Digital Scholarship Services, Media Services, the Office of Scholarly Communication and Publishing, and Technical Services. This collaboration included purchasing village gazetteers and cataloging them; managing book delivery between Hillman Library and our off-site storage location; performing data input and extraction; refining the mechanism and offering suggestions for data extraction and input; developing the data entry platform on Drupal; managing the D-Scholarship data repository; incorporating administrative information of villages and links to gazetteers through Google Maps to construct a spatial-textual connection; and developing the website of the CCVG Data project.

Our collaboration was not limited to within the library; rather both projects have received significant involvement from external scholars. The University Library System successfully received a grant of $180,000 from the Henry Luce Foundation and invited Dr Edward Gunn, Professor Emeritus of Cornell University, and Dr Kun Qian, Associate Professor at the University of Pittsburgh, to play the role of the director and assistant

---

director, respectively, for the creation of a documentary film, *Unreconciled Memories: Reflections on China’s Cultural Revolution*. In another example, ten scholars from the University of Pittsburgh, Harvard University, St. Vincent College, The Chinese University of Hong Kong, Hong Kong University of Science and Technology, the University of California, Los Angeles, and Max Planck Institute for the History of Science served on the advisory board of the CCVG Data Project. These scholars come from diverse academic backgrounds including East Asian Languages and Literature, Library and Information Science, Computer Science, Sociology, Political Science, History, and Economics. Advisors have also provided valuable comments and suggestions to the projects as well as promoting each of them in the classroom and academia. An interactive platform was designed and built by a team from Pitt’s School of Computing and Information. In April 2022, the Conference on Rural Society and Politics in China was held and co-hosted by the University Library System, Asian Studies Center, and Department of Political Science at the University of Pittsburgh. All these examples demonstrate that not only have collaborations ensured the smooth progress of these projects, but more importantly, the collaborations have been initiated and led by the librarians.

**From collecting to creating**

In 2015, the Collection Development Office (CDO) at the Library of Congress undertook a long-term project to update the Library’s Collection Policy Statements and articulate the use of five collecting levels: Minimal Level, Basic Information Level, Instructional Support Level, Research Level, and Comprehensive Level. Among the five levels, the Research Level is “A collection that includes the major published source materials required for dissertations and independent research, including materials containing research reporting, new findings, scientific experimental results, and other information useful to researchers. It is intended to include all important reference works and a wide selection of specialized monographs, as well as a very extensive collection of journals and major indexing and abstracting services in the field. Older material is retained for historical research. Government documents are included in American and foreign law collections.”

Ranging from our printed books to datasets and cloud-based streaming content, the University Library System’s “general library collection reflects the diversity of both the intellectual interests of the University community and the evolving resources needed to support those interests. The collections support the teaching and research needs of the University.”

Similarly, the East Asian Library has not limited its acquisitions to only collecting primary sources but has created its own primary sources as these “new findings.”

Digital Humanities activities within the East Asian Library stretch back to 2002 when the library applied and received a grant from the National Endowment for the Humanities (NEH) to catalog and preserve through microfilming 3,000 acidic and rare books from our Chinese

---

8 https://www.chinesevillagedata.library.pitt.edu/conferences.html
9 https://www.loc.gov/acq/devpol/cpc.html
10 Ibid.
11 https://www.library.pitt.edu/collections-overview
monograph collection. By digitizing a small subgroup of these materials, we could offer enhanced access and limited searchability. More recently, the University Library System completed two digitization projects related to Japanese art: Tsukioka Kōgyo: The Art of Noh, 1869–1927, and the Barry Rosensteel Japanese Print Collection. Later, more Chinese studies-related digitization projects were completed including Chinese Land Records, Chinese Marriage Certificates, Chinese Overseas Student Newsletters, Political Prisoner Photo Collection, and the Sze Papers. However, all the Digital Humanities projects noted above were those carried out through scanning for preservation and improved access. In contrast, the CR/10 and CCVG Data projects are completely different types of DH projects that require librarians to be more creative.

The CR/10 Project was entirely librarian-initiated and was an experiment within the East Asian Library. It recognizes at the outset that history is complicated since memories fade or differ from one another. Its value in preserving memories of the Cultural Revolution is self-evident, but more generally, the project is concerned with how history is passed on to succeeding generations. In addition to interviewing witnesses of the Cultural Revolution, the project also contains interviews with young people who were born after the Cultural Revolution, who share their knowledge about the era, and how they came to learn about it. The project demonstrates the value of studying individual memories collected through oral histories to investigate and analyze historical events. The CR/10 interview video collection and the documentary film are now considered one of the University Library System’s most unique collections created by the library itself; staff formed the idea and concept and then worked to create and implement it with the help of other team members.

While CR/10 is a born-digital video project, the CCVG Data Project was based on the library’s unique collection of Chinese village gazetteers that began to be acquired in 2005. After more than ten years of effort, the East Asian Library now owns the largest collection of gazetteers outside mainland China today which made the CCVG Data Project possible. The project has converted data locked in print resources into a dataset with the functionality to support access and enable sorting and analysis to support teaching and research. The CCVG Data Project contains a total of 18 categories of data which are included in the dataset with 147 subcategories comprehensively covering Village Information, Natural Environment, Natural Disasters, Last Names, Ethnic Groups, Population and Migration, Military, Politics and Management, Economy, Family Planning, and Education. As the first dataset of its kind, the CCVG Data Project has received much attention from scholars.

---

14 https://digital.library.pitt.edu/collection/kogyo-tsukioka-art-noh
16 https://digital.library.pitt.edu/collection/chinese-land-records
17 https://digital.library.pitt.edu/collection/chinese-marriage-documents
18 https://digital.library.pitt.edu/collection/chinese-overseas-student-newsletters
19 https://digital.library.pitt.edu/collection/chinese-political-prisoner-photograph-collection
20 https://digital.library.pitt.edu/collection/szeming-sze-papers
and researchers worldwide. In addition to East Asian Library staff receiving invitations to provide workshops in North America, Europe, and Asia, the CCVG Data project has also promoted our Chinese village gazetteer collection. For example, Dr Zhaojin Zeng from Duke Kunshan University started research on bacterial fertilizer and the history of technopolitical development in China inspired by the CCVG Data project. Further, a PhD student from Indiana University Bloomington applied for a travel grant to support his visit to the University of Pittsburgh to use the village gazetteers to write his dissertation on Chinese minority groups.

Further, the two projects not only support Chinese studies in academia, but also contribute to the field of Digital Humanities development on activities, procedures, and products. Beyond determining what should be digitized, how to develop appropriate bibliographic control mechanisms for digital materials, and how to ensure that digital materials are preserved appropriately and accessible, we found that libraries can do more by creating their own unique DH projects by collaboratively working with colleagues within and outside the library.

**Librarians’ new role—collecting knowledge to creating knowledge**

Decade by decade, librarians’ traditional roles in academic libraries have largely been collecting and describing print materials, curating and providing access to the library’s collection, and promoting and supporting researchers in their quest for information. Projects as described in this article initiated by the East Asian Library provide another possibility related to the librarian’s role. Rather than being passive providers of information, librarians can create content that scholars can use to create knowledge.

Since The CR/10 Project opened to the public, we have found that interview videos have been used in teaching activities and included in faculties’ reading lists for their courses on contemporary Chinese history. In 2018, the CR/10 team was invited by Professor Denise Y. Ho at Yale University to present at her “China in the Curriculum” workshop at the middle and high school levels, and to discuss ways of incorporating CR/10 into China-focused courses. In Carnegie-Mellon University professor Benno Weiner’s course, students watched interview videos, asked questions, and had a lively discussion. Professor Lei Ouyang at Swarthmore College cited CR/10 interviews in her book titled *Music as Mao’s Weapon: Remembering the Cultural Revolution.*

CR/10 interview videos have been not only used in courses on Chinese history but also in other research fields. For example, Professor Rongqian Ma at Luddy School of Informatics, Computing, and Engineering at Indiana University did “a rhetorical analysis of the design features and curation processes of the CR/10 website” and “discusses the functions of CR/10 as a Warburgian memory atlas that shape the nonlinear, multifaceted narratives of a historical incident.” Her article “offers a strong case for the datafication of cultural memories and contributes to digital archives and humanities with an innovative interface design.”

---


The CCVG Data Project essentially extracts data from printed books and enters it as electronic data into a database. It is not only a change of format (from print to digital), but also from text to the dataset which is a completely different product. Even though the size may be considered small (2,601 villages) due to the availability of village gazetteers, it fills the gap in rural China studies and has become a highly desirable and accessible dataset. Before the project was initiated, an advisory board was formed with scholarly experts from a variety of research fields. In a support letter, Professor Yunxiang Yan from the Department of Anthropology at UCLA wrote, “Such an open resource database will provide a great service to scholars in various fields of China studies and beyond by enabling them to access easily and directly the primary data on village administration, local history, population change, education, and health care, family ... its scholarly value can hardly be overstated.” On April 22-23, 2022, an international conference on Rural Society and Politics in China was held at the University of Pittsburgh, which was hosted by the University Library System, Asian Studies Center, and the Department of Political Science at the University of Pittsburgh. More than 40 scholars and PhD students from North America, mainland China, and Hong Kong participated. They highly evaluated the value of village gazetteers as a unique primary source as well as the importance of CCVG Data to research. Presently, Chinese village gazetteers and CCVG Data have caught the attention of scholars. “While computational methods could never replace close reading of primary sources, they offer new forensic tools and yield fresh perspectives on a perennial question: how do we know what we know about China?” CCVG Data is also used by students for classroom activities and research assignments. Ms. Yuanziyi Zhang, a graduate student at the University of Pittsburgh, used the Chinese village gazetteers and CCVG Data project to conduct two research projects: “Visualization: Chinese Local Performances from Village Gazetteers” and “The Influence of Arable Land Area on Gross Output Value in Chinese Villages.” Since CCVG Data opened to the public, the project surveys we conducted demonstrate the diversity of fields and subjects in which our researchers approached the content. This should be of great encouragement to librarians as they pursue more creative projects and professional development opportunities where they can play the role of knowledge creator rather than just collector.

Challenges
Both the CR/10 and CCVG Data projects experienced challenges and difficulties with some common to both projects. For example, a strategically detailed workflow needed to be developed since both projects were highly collaborative within the University Library System, between the ULS and the School of Computing and Information, and among librarians and scholars. In contrast, both projects faced different challenges due to the features and characteristics of each project. Since the CCVG Data Project was the first of its kind to extract and create an interactive dataset, it faced more technical issues and obstacles, including building and designing two customized platforms for entering data as well as enabling data-cross searching and filtering for users. For The CR/10 Project, the most essential and serious challenge was how to deal with the issue of privacy. As an academic

24 https://www.chinesevillagedata.library.pitt.edu/conferences.html.
26 https://www.chinesevillagedata.library.pitt.edu/teaching.html.
institution, the library wanted to maintain a neutral and objective stance in the interviews on the topic of the Cultural Revolution since for many Chinese, the Cultural Revolution remains a sensitive topic to be talked about with worries. Upon request, the library had to respect and protect the privacy of interviewees if they did not want to disclose their names or show their faces during the interview.

Considering that video would more likely hold people’s attention, especially in the classroom, and body language and facial expressions could add meaning to stories, the CR/10 interviews were filmed. Since the Cultural Revolution is a controversial subject in China, we also knew questions about the veracity of the interviewees’ stories might arise. To avoid users suspecting we had manipulated or edited the interview content, we decided to film the interviews and not just provide access to the audio. To protect not only the person being interviewed but also those mentioned by the interviewees, all personal names were hidden or replaced by a pronoun such as a neighbor, a teacher, or a friend during interviews, translation, and subtitle work. For political sensibility reasons, participants in CR/10 can elect to be anonymous when they sign the CR/10 Consent Form; moreover, they can request to have their faces masked or blurred to maintain their anonymity in the online video. Reviewed and approved by the University of Pittsburgh Human Research Protection Office (HRPO), the CR/10 Consent Form was provided in English and Chinese. In the consent form, interviewees give permission to include their interview video in the online archive database for research and education purposes. This consent form releases the library from any claims that may occur from showing, using, or distributing the video.

As mentioned previously, the library received a one-year grant from the Henry Luce Foundation in 2019 to support the creation of a feature-length documentary film titled, The Revolution They Remember. As a collaboration between the East Asian Library and Dartmouth Library, The Revolution They Remember is based on two video oral history projects at each university (CR/10 and Down to the Countryside Movement). The documentary film features selections from the interviews conducted by these two projects, as well as photographic images contributed by interview participants, archival footage, and photos. Since the creation of this film was not originally conceived when the majority of the respective interviews were conducted, a separate consent form was needed to be signed by interviewees in order to include them in the documentary. To include additional moving image clips and photos in the film, extensive rights usage and permission were needed which took an exhaustive amount of time and resources.

Summary

The setting described in the article by Michael Dabrishus and Haihui Zhang still applies today: “East Asian Digital Humanities is very much still under development. Due to the collection’s nature in content and languages, and international copyright questions, the digital project will consist of a team of scholars, graduate students, digitization specialists, librarians, and administrators at Pitt. The challenges of discovering, identifying, and evaluating primary documents for potential digitization projects require subject scholars’ and experts' involvement. Legal consultations, a need for staff with language skills, and cultural

knowledge make the process challenging, yet also very interesting and rewarding.”28 The CR/10 and CCVG Data projects are a direct result of such innovative thinking by the University Library System, which has provided a space to explore building unique collections and develop Digital Humanity projects. We have learned that the academic librarian can collect, present, and promote information as well as create knowledge through innovative projects. As we have demonstrated, the academic librarian can further shift from a passive service role to an active service role, but achieving this requires and demands an all-around focus on collaboration as the primary key.

---