White Paper

China Historical Christian Database: Mapping the Spatial and Social Networks of Christianity in China, 1550-1950

National Endowment for the Humanities
Office of Digital Humanities

Digital Humanities Advancement Grant, Level II
HAA-280992-21

Daryl Ireland and Eugenio Menegon
Boston University
December 19, 2022

1. Summary

1.1 Narrative Summary
“The China Historical Christian Database: Mapping the Spatial and Social Networks of Christianity in China, 1550-1950” is a project that was generously funded by the National Endowment for the Humanities (September 1, 2021 - August 31, 2022). The China Historical Christian Database (CHCD) is hosted by the Center for Global Christianity and Mission at Boston University.¹

The CHCD is a project that quantifies and visualizes the place of Christianity in modern China (1550-1950). The CHCD records Christian organizations, events, institutions, people, and the relationships that tied them all together. This information can be used to create spatial maps and relational networks that reveal the complex cultural exchange which took place between China and the world in the modern period. This project is the largest historical dataset on Christianity in China and breaks new ground in providing quantifiable data about modern Sino-Western relations.

More than a dataset, however, the CHCD team has also developed an online platform that allows users of all technical capabilities to explore connections in the database. The platform’s intuitive interface is available in English and Chinese, and it enables the generation of visualizations, lists, and maps for use by the general public, students and teachers in secondary education and colleges.

1.2 Project Team

1.2.1 Core Team Members
- Eugenio Menegon, Boston University
- Daryl Ireland, Boston University
- Alex Mayfield, Asbury University

¹ Center for Global Christianity and Mission, https://www.bu.edu/cgcm/
1.2.2 Project Collaborators
- Xian Liu, Renmin University
- Alexandrs Dmitrenko, Heidelberg University

1.2.3 Project Associates
- Raissa De Gruttola, University of Perugia
- Matthew Preston, Anderson University
- Amanda Clark, Whitworth University

1.3 Project Links
- Project Page (https://chcdatabase.com/)
- Online Platform (https://data.chcdatabase.com/)
- Data Documentation (https://chcdatabase.github.io/data-documentation/)
- Data Collection Documentation (https://chcdatabase.github.io/data-collection/)
- Data Download (https://github.com/chcdatabase/data)

2. Background

2.1 Origins
In 2017, Alex Mayfield asked, “What could someone learn if Pentecostal missionaries, who were highly mobile actors in Hong Kong, were put on a map?” He addressed the question to Eugenio Menegon in a directed study course and, because Mayfield had the technical skills and Menegon the background in Digital Humanities, the answer was born digitally. Using over a thousand Pentecostal periodicals, Mayfield tracked the movements and activities of Pentecostal actors in Hong Kong. The addresses of their homes, chapels, orphanages, and the like were recorded, along with the names, nationalities, and gender of each Pentecostal. The result produced 4,905 different people, ministries, and locations, and generated a thick web of 33,355 relationships connecting them. This dense network was used to quantitatively illustrate that Pentecostals were heavily invested in institution building, valued education, worked cooperatively with governmental bodies and across denominational boundaries, and avoided densely populated areas. These trends all cut across standard historical narratives that have treated the growth of global Pentecostalism as a sectarian, other-worldly outcome of global urbanization and modern industrial anomie.

2.2 Scaling Up
Daryl Ireland, the Associate Director at Boston University’s Center for Global Christianity and Mission (CGCM) recognized that this approach could have far-reaching implications. If a relatively small sample could produce such new insights, what could be learned if the sample was and expanded to include all of Christianity over a 400-year period? At that scale, could a person see what role Christians and their institutions played in the formation of modern China?
Such a question aligned with the Center’s research focus and overlapped with his own expertise. Ireland, Mayfield, and Menegon agreed to expand the original project into the CHCD, making it an initiative of the CGCM in 2018.

The challenge was how to proceed. Thus, the CGCM applied for and received a $5,526 grant that came with three months of consultation and software development from Boston University’s Software and Application Innovation Lab. In 2019, those funds built the proof-of-concept version of the CHCD, which focused on one province in China: Shanxi. An additional $6,000 from the CGCM allowed a cohort of graduate and undergraduate students to work alongside Ireland, Mayfield, and Menegon as they scraped data points from sources in French, Latin, English, Chinese, and Swedish. Although limited, this beta product proved capable of inspiring enthusiasm. Soon after, the CHCD team formed formal partnerships to further the aims of the project. Notable partners include the Institute of Qing History at Renmin University, the Ricci Institute for Chinese-Western Cultural History at Boston College, the University of Naples “L'Orientale”, and Whitworth University, among others.

2.3 Preparing for the Future

Before building a full database and user-friendly web interface, the CHCD team understood the necessity of getting input from potential users. With generous financial support from Boston University’s Institute on Culture, Religion, & World Affairs, the BU Center for the Study of Asia, and the Institute of Qing History at Renmin University, Eugenio Menegon, Daryl Ireland, Xian LIU, and Alex Mayfield hosted, “Mapping Christianity in China, 1550-1950: Developing Relational and Geospatial Tools for the Study of Christianity in China.” The virtual event brought together twenty academics and archivists as presenters from four continents to provide input and feedback on the CHCD on November 19-21, 2020. They evaluated the current data sets, thought about what is missing, and explained how they imagined using the database and web interface, so that the final product could be responsive to the needs of end-users like themselves. This online event attracted 266 scholars from 28 countries, showing the pent-up energy in the field and the need for this digital initiative.

3. Activities

3.1 Activity Summary

During the grant period (September 1, 2021 - August 31, 2022), the CHCD team’s activities can be largely split into three areas: data preparation, technical development, and public outreach. Project team members were responsible for various areas of the project’s development. In addition, Alex Mayfield was hired as the Project Director to provide both technical development support and to ensure that all project activities were completed in a timely manner. In addition to the generous support of the National Endowment of the Humanities, the project team also received $100,000 in private donations during the grant period. This additional funding enhanced the stated goals of the project, namely by enabling more data collection during the project period.
3.2 Data Preparation

As the core of the project, the process of data collection, cleaning, and integration made up a sizeable portion of all project activities during the grant period.

Data Collection
A large portion of the data for the project was collected prior to the grant period. However, the additional funding from private donors allowed for an additional round of data collection to be included. Thus, from September to April, the project team oversaw dozens of students, volunteers, and partners as they collected additional data from French, Chinese, Spanish, Russian, and English sources. Notable contributions from this round of data collection include data on Russian Orthodoxy in China, early Protestant missions in China, and New Society Jesuits in China.

Data Cleaning
The data cleaning process occurred in stages and at key junctures of the data collection process. From September to October, the data collected prior to the grant period was cleaned and prepared for the database. During this period, pertinent data from the defunct Ricci Institute’s Roundtable Database was also incorporated into the dataset. Another round of cleaning took place from January to February to incorporate data collected during the fall semester. The final round of cleaning took place from April to May. This phase focused on removing duplicates, assigning unique ids, and incorporating data collected in the early part of the spring semester.

Data Integration
In early June, the initial data was converted into CSVs and uploaded into a Neo4j graph database. After this initial upload, the project team oversaw a brief period of cleaning to correct any additional errors and to ensure that the data was in conformity with the database schema.

3.3 Technological Development

Technical development activities can be largely broken into two categories: database development and online platform development.

3.3.1 Database Development

Initially, the project team intended to create a data-input system that would enable partners to enter data directly into the database. In September, the project team was advised by members of other large database projects that such an approach would be cost- and time-prohibitive and prove less useful than desired. Thus, the project team was forced to rely upon pre-existing solutions. Thankfully, Neo4j, the graph platform used by this project includes an intuitive backend interface that enabled the project team to input and manipulate project data.

From November to December, the team implemented a draft Neo4j Community Edition database on a free Elastic Cloud Computing (EC2) instance on Amazon Web Services (AWS). Hosted using free tiers within AWS pricing schemes, this draft database hosted previously gathered data and allowed for application development. From May to June, the full-scale database was
implemented using Neo4j Community Edition hosted on an EC2 reserved instance with AWS. During this time, a CSV version of the dataset was also stored on Boston University Library’s OpenBU platform and GitHub.

The choice to deploy the database with open-source and commercial solutions was a purposeful one by the project team. Widely supported by large technology companies, the technologies should be readily updateable and resistant to technological obsolescence. The CSV files stored with OpenBU provide a permanent repository for the database.

### 3.3.2 Application Development

From September to April, the team oversaw the development of the online platform. The platform is a React JavaScript application that depends on several libraries for the design and data visualization, most notably Bootstrap.js, Leaflet.js, and D3.js. This platform was a major expansion and revisioning of the beta version that had been created prior to the grant period. During this period, the principal investigators also created a partnership with students at Anderson University who were responsible for creating additional customizable graphs and charts using the data. Unfortunately, their work was not completed by the time of the project launch and will have to be incorporated later.

From May to June, the online platform underwent a period of prolonged adjustments, bug fixing, and query tuning to ensure that the application was fully functional and user experiences were intuitive. The final product was uploaded to a GitHub repository so it can be easily utilized by other projects. In addition, the projects source code was stored in Boston University Library’s OpenBU platform.

### 3.4 Public Outreach

As a project oriented towards data literacy and public engagement, outreach and publicity is as integral to the CHCD as data collection and technological development. One of the main ways this was achieved was through partnerships. During the course of the grant period, the project team worked with partners at over nine institutions to collect data and develop the project. These institutions include Boston College, Renmin University, the University of Naples, the University of Perugia, and Whitworth University. A second avenue of outreach was through scholarly meetings. Throughout the grant period, project team members presented on the progress and scope of the project via regional, national, and international scholarly meetings. As a third avenue of dissemination, the project team also published a technical essay about the project in Digital Humanities 數字人文. Finally, the project team produced a quarterly newsletter, *The Gazetteer*, which directly updates over 350 donors and interested parties about the progress of the project.

These various avenues ensured that our relevant audiences in the general public and scholarly community were informed of the status of project.
4. Outcomes

4.1 Overview of Initial Goals

The initial grant proposed to meet the following goals:

1. Create an initial database with over 400,000 data points
2. Produce an online data entry system that allows for streamlined data entry, and
3. Produce a fully functioning web application which will allow the public to interact with the data from the database.

As will be seen, Goals 1 and 3 were not only met but surpassed by any measure. The overarching intention of Goal 2 was met, however, the form needed to be adjusted for the sake of cost-effectiveness and efficiency.

4.2 Goal 1: Database

By combining the data produced prior to the grant period with the data collected during the grant period, the CHCD surpassed its initial goal of having a database with over 400,000 data points. In its initial release, the database is comprised of:

1. 44,979 Nodes
   - 33,900 People
   - 6,488 Institutions
   - 1,155 Corporate Entities
   - 130 Events
   - 1,306 Geographic Locations
2. 208,237 Relationships
3. 1,916,717 Properties

This large number of nodes, relationships, and properties allows users of the database to create queries to study a diverse number of religious groups in China, and they can be combined with other datasets to understand how Christianity affected historical happenings in China. The sources utilized to create this dataset are documented in the image below.
4.3 Goal 2: Data Input System

Initially conceptualized as a way for partners to enter data, our experiences with bulk data-collection demonstrated that developing any backend-interface may be cumbersome and potentially slow the rate of data entry. In addition, such an interface would require a large amount of development time and an ongoing suite of trainings for partners. Upon advisement from experienced members of the digital humanities community and members of the project’s board, the goal was adjusted for the sake of efficiency and cost-effectiveness.

Rather than develop a new system for data entry, the team focused on developing data collection processes and spreadsheet templates that would ensure accurate and cross-referenced data which could be cleaned and easily integrated into the database using customized queries. This approach leveraged pre-existing technologies and lowered the technical threshold needed for project participation.


<table>
<thead>
<tr>
<th>Tradition</th>
<th>Group</th>
<th>Dates</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protestants</td>
<td>General</td>
<td>1837-1901</td>
<td>Boynton, Charles Luther. &quot;Revised Chronological List of Missionaries&quot; Columbia Archives</td>
</tr>
<tr>
<td>Catholic</td>
<td>Passionists</td>
<td>1932-1950</td>
<td>&quot;Passionists in Hunan, China.&quot; Passionist Historical Archives</td>
</tr>
</tbody>
</table>
4.4 Goal 3: Web Application

Building on the beta version produced through a prior grant, the funding from the National Endowment for the Humanities enabled a full-scale redesign and feature addition to the database. The Beta Version was a basic, browser-based application that depended on JSON files to populate the various views. The beta version did not work well on mobile devices, and it was not scalable.

In the Release Version, the web application is a fully responsive React.js application that is directly linked to the database over an HTTPS connection. This responsive design allows for students, educators, and researchers to use the application on whatever device they prefer, and the direct linkage to the database ensures the application will automatically incorporate any additional data added to the database.

The final application has three “views” which allow users to interact with CHCD data in multiple ways. In the “search” view, users can search for specific entities in the database. In the “map” view, users can utilize filters to create customized maps of people, events, and institutions. On the “network” view, users can create customizable ego network visualizations. Importantly, each view is interactive and allows a user to click on database entities and pull up lists with all of their relationships and information. An additional “fourth” view was planned in coordination with students from Anderson University. However, this view remains in development and will be completed through future collaboration.

Like the rest of this project, the source code for this web application is freely accessible. The release version is available via GitHub and an archived version has been placed on Boston University’s OpenBU platform.

5. Evaluation

5.1 Achievement & Impact

With the completion of the grant period, the CHCD has achieved all the stated goals of the grant project. As a digital humanities project, the technical achievements of the project team are straightforward:

1. A large database running on cutting-edge graph technology,
2. A JavaScript web application built to industry standards, and
3. The largest dataset on historical Christianity in China in existence.

Together, these technological achievements constitute an innovative digital humanities project that blends the best of the humanities, newer developments in database technology, and the world of application development.

Having achieved the main goals of the project, the CHCD’s larger impact on the fields of Chinese History, Global History and World Christianity will be only clearly ascertained well after the grant-period is over. At the same time, several research groups have already expressed
interest in using the initial dataset for their own research agendas. As an initial entry into this future body of scholarship, the project team is currently formulating several research questions which could be answered using data from the database. To measure future impact, the project team will use Google Analytics to monitor the use of the project website and online platform. We will also be monitoring publications, presentations at academic conferences, and citations. Regardless of what the future holds, the work accomplished during the grant period provided valuable lessons for the future and demonstrated that there is a remarkable amount of interest in the long-term goals of the project. Three notable learnings rise to the surface.

First, we recognized the necessity of diverse partnerships in developing large humanities projects. The project would have been impossible without the contributions of partners of all kinds. Stakeholders ranged from researchers with overlapping interests, to volunteer Chinese Christians interested in their own history, or undergraduate data science students who wanted to apply their skillsets to a novel problem. These diverse partnerships required a good deal of management but demonstrate the potential value and intersectional appeal that digital humanities projects can have. When they are mutually enriching, digital humanities projects can be “co-owned” by a diverse body of stakeholders.

Second, we recognized the need for a “canonical” approach in large-scale data collection. Any historical dataset will be necessarily incomplete, yet data visualizations produced from such datasets can mask this incompleteness. Through the course of data collection and application development, the project team recognized the need to find some standard measure by which to measure the relative completeness of data being collected. As such, data was collected only when a clear corpus or “canon” could be identified which contained the desired historical information. This produced a dataset with nodes and relationships that were more or less parallel with one another (e.g., all Jesuits in China and all the Russian Orthodox missionaries in China), which future researchers could identify as a clear and complete unit of analysis.

Third, we grappled with the need for data literacy in the digital humanities. Citizens of modern, technology-dependent societies are constantly surrounded and affected by algorithms and data science. We are humans in a digital world. Yet, as the project progressed, we found that very few students in the humanities had the skillsets needed to accomplish many of the data-forward tasks needed for the project. Student workers and volunteers who came from non-humanities fields like data science, engineering, or business were often far better prepared to work on the project. While the interest of non-humanities students bodes well for the continued importance and appeal of humanistic questions, it does not bode well for the humanities in general. The task to create students who can answer inquiries about the human condition in the digital world demands that humanities programs invest in developing the skillsets needed to properly frame the question. To that end, digital humanities projects like the CHCD can provide a much-needed avenue for skillset training and data literacy development.

All these broader learnings are useful for informing the future stages of the project and the discipline of digital humanities as a whole.
5.2 Challenges

The major challenges encountered were wholly related to the wide range of skillsets necessary to accomplish the project. As a digital, historical project on Christianity in China, the project required expertise in multiple languages (specifically Chinese, English, French, Spanish, Italian, Russian, and Latin), familiarity with modern and pre-modern ecclesial administration and nomenclature, data science skills (notably data cleaning and database management), and web application development. This broad range of skillsets could only be achieved through the recruitment of key team members and the formation of many diverse partnerships.

6. Project Continuation

6.1 Future Goals

The project team has already identified key avenues for continuing the work of the project in the ensuing months and years.

6.1.1 Data Collection

The current dataset leans heavily on Western and Protestant sources. Notable exceptions to this trend include Russian Orthodox Missions, the Society of Jesus, the Maryknoll Fathers and Sisters, the PIME Fathers, and the Passionists. Still, several major religious groups are still missing or are almost entirely from the database. Thus, the project team has identified several major groups and sources which will populate the next major update to the database. In no particular order, these groups are:

- Chinese Literati
- Chinese Protestant Pastors
- Franciscans
- Dominicans
- Lazarists
- SVD Missionaries
- Protestant Missionaries (pre-1900)
- Educational Facilities

Data collection on many of these groups is already ongoing thanks to the support of private donors and internal institutional funding at Boston University.

6.1.2 Data Quality

In the final stages of cleaning, several endemic issues were noted throughout the database, such as accidental merges of people with similar names. While these do not take-away from the overarching usefulness of the data, correcting these issues will provide a more useful dataset and better user experience. Collaboration with a data science lab or department would be an ideal avenue for eliminating these issues.
6.1.3 Scholarship Production

Going forward, the project team would like to strike a balance between large scale data collection and scholarly production. With the release of the initial dataset, the project team has begun to pursue research questions related to the database with the goal of publishing findings in peer-reviewed journals. In addition, the team is currently submitting a grant proposal for funding that would enable further data collection, as well as the production of a scholarly edited volume made up of scholarship which utilizes the database.

6.1.4. Web Application Development

The CHCD online platform will require ongoing maintenance and development. The partnership with Anderson University will provide one of the main sources for this work, especially in relation to the “data” view of the web application. The CHCD team hopes that these sorts of partnerships will make up a greater portion of their collaborative efforts going forward.

6.2 Looking Ahead

These far-reaching goals outlined above will sustain a good portion of the project team’s research agenda for several years to come. Yet, our partners’ and collaborators’ dedication to the CHCD has demonstrated that there is an excitement about the pedagogical and research potential of the project. As the CHCD ventures into its next phase, the project team hopes that the project continues to inspire excitement about Chinese History and the often-ignored role that religion had in it.