THE CHINA HISTORICAL CHRISTIAN DATABASE

An Essential Tool for the Cartography & Prosopography of Christianity in China

BOSTON UNIVERSITY

Center for Global Christianity Mission

Project Leads
Eugenio Menegon, Daryl Ireland, Alex Mayfield

ABSTRACT

The China Historical Christian Database (CHCD) is a powerful new tool for the study of Christianity in China. Through the creation of a cutting-edge geographic and relational database, an innovative online platform, and strategic international partnerships, the CHCD offers a new approach to resolving some of the classic problems of historical research on Christianity in China: the linguistic plurality and geographic spread of sources. Moreover, the technologically driven approach of the project is poised to create a field-changing resource that opens new forms of analysis in the study of Christianity in China. This white paper details how such a task is possible and currently being undertaken.

CONTENTS

DEFINING THE PROBLEM

The study of Christianity in China is often met with a riff of Tertullian's classic theological question "What does Beijing have to do with Jerusalem?" Nationalist historiography and post-colonial theory have tended to relegate Christianity to the status of "foreign," thus rendering it peripheral to the story of China. Yet, as more recent studies have demonstrated, Christianity has undergone a process of localization and "sinicization" since its arrival in China, and its historical experience can actually provide an essential window through which to understand the revolutionary changes of identity, politics, and geography that took place in China during the modern period. Yet, the role of Christianity in China cannot be so readily divorced from the rest of the world. Throughout China's history, Christian actors and institutions created new relationships that crossed regional and national boundaries. To be Christian could be both a marker of local identity and participation in global or regional networks; the line between "local" and "foreign" is fuzzy. As such, understanding the nature of Christianity in China helps both understand China's own political journey and its evolving relationship with the world at large.

That said, analysis of the dynamic nature of Christianity in China must overcome large challenges: historical agents were highly mobile, relational networks were constantly in flux, and historical resources remain disparate. These challenges have led most studies of Christianity in China to focus on localized areas, smaller groups, or specific time frames. Broader, more empirical approaches appear all but impossible because it continues to be difficult to triangulate historical sources or organize them in such a way as to make possible large-scale studies. These challenges, however, are not unique to the study of Christianity, and a turn towards leveraging technological tools has demonstrated that addressing these issues is possible. Tools like the China Historical GIS, the Ricci Institute's Roundtable Database, and the China Biographical Database have all made strides to wrangle these same sorts of complexities. Despite these laudable achievements, however, much of the fruit of these digital approaches remain out of reach for general researchers due to a lack of user-friendly

interfaces, high technological learning curves, or technological obsolescence. Moreover, no existing effort has taken on the complicated task of organizing and making sense of spatial, temporal, and relational data pertaining to Christianity in China.

In summary, the problems facing an empirically driven, large-scale analysis of Christianity in China can be grouped into three main areas:

- **Sources in Foreign Languages**
  Christian actors and institutions in China reflected the global nature of Christianity itself. Historical sources were recorded in multiple languages and reflect multiple cultural frameworks.

- **Sources Remain Disparate and Complex**
  Historical sources on Christianity in China exist in a multitude of archives and private collections across the globe. Moreover, many of these materials exist in specialized genres and/or formats that require specialized knowledge to analyze.

- **Technological Learning Curves**
  While the above complexities are not insurmountable, scholarly engagement with new technological tools remains haphazard. When datasets are created, they are often inaccessible or unusable to most scholars.

While large, these problems are not insurmountable. The advent of new tools of analysis, the digitization of archival materials, and the steady compilation of biographical and geographical references works have prepared the way for a new approach. The CHCD builds on these past scholarly accomplishments while creating a new kind of digital tool for the study of Christianity in China.

A FLEXIBLE, MULTI-PRONGED APPROACH

The China Historical Christian Database (CHCD) is a new resource that addresses the linguistic, geographic, and technical problems facing the study of Christianity in China. At the same time, it opens new possibilities for future scholarship. Not content to create data that is inaccessible or the sole propriety of one research team, the project has three major goals:

- **Creation of a Geographic and Relational Database**

  This database will be flexible in design and allow student and scholars to track Christian people, institutions, organizations, and events in China between the 16th - 20th centuries. No historical database can be exhaustive. However, the CHCD aims to capture as wide a swath of data as possible. This wide-swath approach will enable levels of statistical analysis of Christian presence that have heretofore been impossible.

- **Creation of a User-Friendly Online Platform**

  Separate from the database, the online platform aims to make the data as accessible as possible to students, teachers, and researchers through a user-friendly interface that includes both textual search and geographic visualization. When complete, this tool will be available in both English and Chinese. *Visit the Online Platform*

- **Partnership Between Chinese and Western Universities**

  The project aims to bring together multiple teams from Western and Chinese research universities. By working together to collect and analyze data, the project provides a chance to overcome the complexities of the historical sources while creating opportunities for mutual enrichment. Further, the data-driven methodology of the project enables scholars from different disciplines and fields to engage the project.

Overcoming the obstacles in the study of Christianity in China is a complex task. This three-pronged approach is both an ambitious and achievable way to launch the field forward into the 21st century.

THE DATABASE

The work of archivists and scholars over the past few decades have drastically increased the availability of once rare sources and information. In fact, one might say that students of Christianity in China could be overwhelmed by the sea of sources. The CHCD both capitalizes on this flood of information and makes it more navigable by building a cutting-edge database that allows students, teachers, and researchers to explore the geographic dimensions and relational networks of the Christian presence in China. To do this, the CHCD embraces a specific database philosophy and design that allows the project team to capture voluminous and complex historical data without sacrificing intelligibility for users of the database.

PHILOSOPHY

The CHCD has several guiding principles which inform the database structure and implementation. Any collection of large amounts of data can give the appearance of objectivity and totality, but this is never the case. Every archive, every database has a story that informs both what information is captured and how that information is arranged. The guiding principles of the CHCD are summarized below:

- **Geographic and Relational Focus**
  The CHCD is primarily interested in tracking relationships and geography. While it is "text-based" in that it relies upon historical sources to provide data, the "texts" of the database are people, institutions, and the relationships that exist between them and the spaces they inhabited.

- **Academically Responsible**
  At the same time, the CHCD can also be treated as bibliographic database in a secondary sense due to its commitment to academic rigor. The entities and relationships of the database are always accompanied by references to the sources in which they are found.

- **Early Modern and Modern Focus**

  This database only focuses on historical actors and institutions from the early modern and modern periods. Specifically, from the arrival of the Francis Xavier in 1552 CE to the establishment of the People's Republic of China in 1949 CE.

- **Contemporary, Granular Geography**

  Projects such as the China Historical GIS have shown that China's historical geography was constantly in flux; political boundaries and regional identities changed continuously. Thus, in placing individuals "on the map," one must ask "which map and from when?" The CHCD has opted to use the contemporary (as of 2009) political map of China as its primary reference. This choice is informed by two reasons: 1) students and scholars are more familiar with the current map of China and 2) historical studies often utilize modern geographies to define their scope. This means that the CHCD has endeavored to link historical place names to their modern counterparts. This approach, of course, has drawbacks that the database has tried to control through two means: 1) original place names are retained along with their modern references, and 2) by adopting a granular approach to geography. Each geographically linked entity in the database is connected to specific geographic coordinates. This granularity allows future researchers to bound the coordinates as they see fit.

- **Strategically Bilingual**

  Historical sources about Christianity are not always available in both Chinese and Western languages. Sometimes, names of Chinese Christians might only be available in English, sometimes only Latinized baptismal names are available, and sometimes non-Chinese people and institutions had multiple Chinese names. CHCD records as much information on database entities as available. Names and alternate names are recorded in Romanized Scripts, Traditional Chinese Characters, and Modern Pinyin when possible.

- **Able to Deal with Fuzzy Data**

  Trying to map historical entities onto a geographic landscape is complicated by the

fuzzy nature of historical data itself. Sometimes, a specific village or street address is provided, other times people are recorded as being in a county or province. The CHCD mirrors this fuzziness by basing its geography schema on modern Chinese administrative divisions (Province-level, Prefecture-level, County-level, Village-level). For more details on this design choice, see the appendix.

- **Cohesive & Flexible Structure**

  The CHCD utilizes cutting-edge graph database technology (more on this below). While there is a primary database structure that focuses on specific entities, the graph database format allows the database structure to morph and grow as needed. This allows the CHCD to grow with the field as new sources and new research goals arise.

- **Lower Risk of Redundancy**

  Traditional relational databases often have a high risk of recording the same information in multiple places. This redundancy of data can lead to issues in search queries and outdated information. The database format used by the CHCD radically reduces the risk of redundancy.

- **Long-term Feasibility**

  The rate of technological progress can often outpace scholarly research projects. Sadly, this means valuable scholarly knowledge can become lost if scholars ignore advancements in the technology. The CHCD utilizes a cutting-edge graph database structure to house its data. Used by tech giants, such as Google and Facebook, graph databases are quickly become the standard data structure thanks to their fast processing power and flexibility. While this does not guarantee graph databases will remain the standard format in years to come, their popularity does mean that tools to transition from graph databases to newer database structures will be readily available.

- **Freely Available**

  The data from the project will be available in two ways. First, the entire database itself will be freely downloadable. Second, students and researchers who are not familiar with database technology can use the online platform to do searches, create visualizations,

and analyze the data from the database. The CHCD project team will upload updated versions of the online platform and database downloads a regular intervals so that the scholarly community has the most up to date information available.

Throughout the database, these philosophical principles are at work informing how data is recorded and why. The project team believes these choices allow the historical data to be both readily accessible and easily understandable.

DESIGN

The CHCD is a graph database that focuses on geographic and relational connections. It utilizes the open-source and industry leading graph database platform Neo4j. A more detailed description of the database architecture can be found in the Appendix to this document. This section provides information on graph database basics and the general design of the database itself.

GRAPH DATABASE BASICS

Graph databases mimic the natural relationships that exist in the real world; their structures often look like what you might draw on a white board when trying to describe how things are related.

The example image and four definitions below offer a basic understanding of the graph database approach:
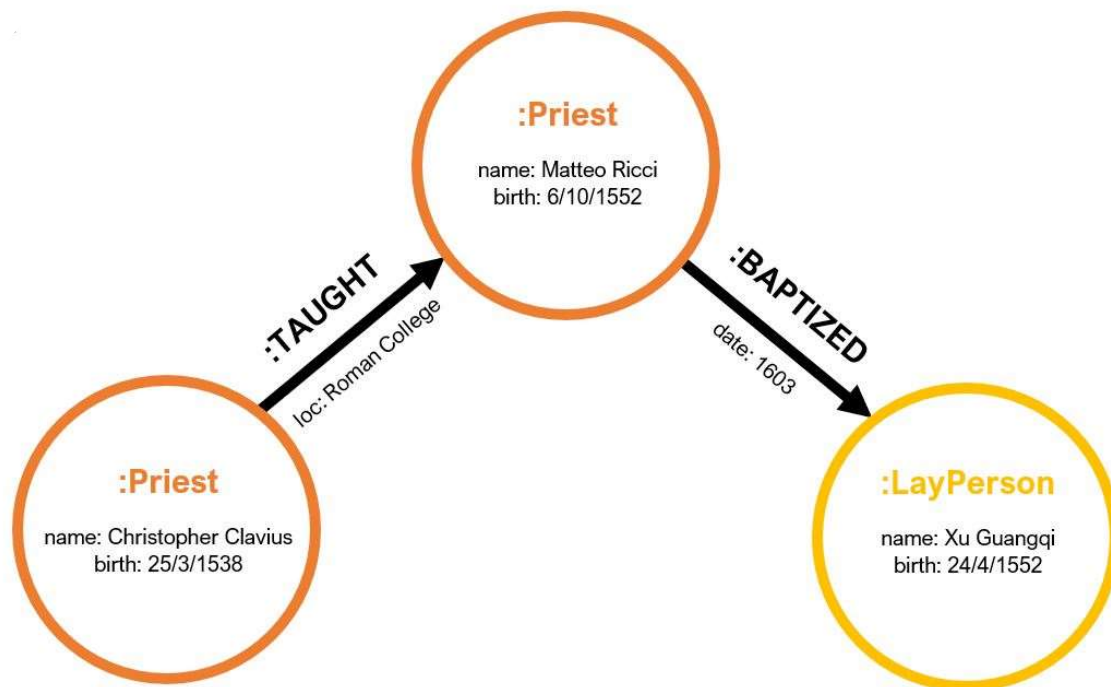
- **Nodes:** These are the primary entities of a database. (e.g. Matteo Ricci, Xu Guangqi)

- **Relationships:** These are the relationships between entities. Relationships in a graph database tend to be directional (e.g. Clavius taught Ricci, Ricci baptized Xu).

- **Labels:** These are primary markers for a node and a relationship. They also can be used to help communicate a flexible structure to the database. Nodes and relationships can have multiple labels if needed. (e.g. Ricci was a priest like Clavius, Xu was a lay person).

- **Properties:** This is additional information that pertains to a specific node or relationship. This information is not relational and does not change (e.g. Clavius taught Ricci at the Roman College).

CHCD GRAPH SCHEMA

The CHCD has four main kinds of nodes (i.e. four node labels) in the database: people, corporate entities, institutions, and events. In addition, there are five kinds of geographic nodes which represent the five different levels of geography in the database. The below schema

depicts the overall structure of the database by showing what relationships are possible between the various types of nodes.
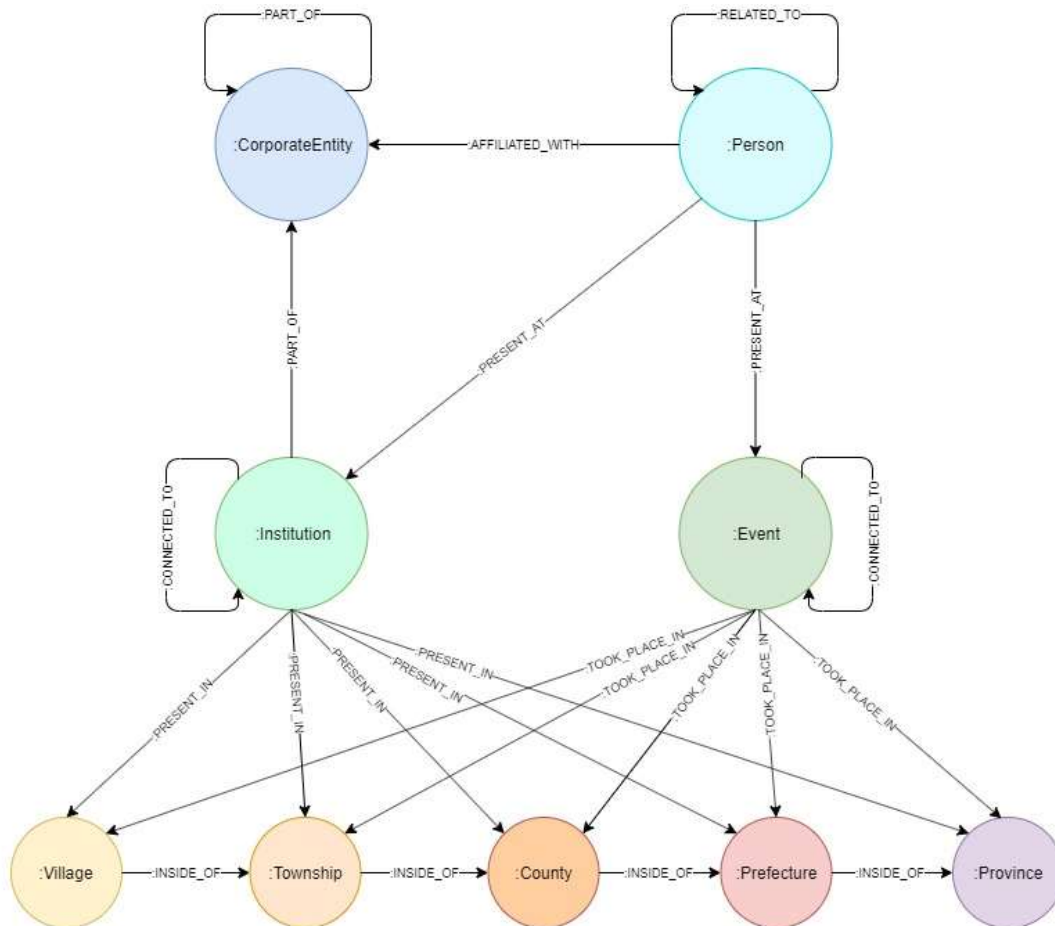
To understand the above schema, some description of the node types is required:

- **Person:** these nodes represent human beings. People are at the core of the database and thus they have the most kinds of relationships possible.

- **Corporate Entity:** these nodes represent organizations that *do not* have a direct geographic footprint. For example, the Society of Jesus is an organization, but it only exists in space through people and institutions.

- **Institutions:** these nodes represent organizations that *do have* a direct geographic footprint. Common examples in the database include churches, hospitals, and schools.

- **Events:** these nodes represent important events that took place in Chinese Christianity. Events are, by definition, temporary happenings that have specific geographic locations. Examples in the database range from Christian conferences to imperial hunting parties.

IMPLICATIONS OF THE DESIGN

The CHCD design schema is a careful balance of flexibility and rigidity. This combination allows for a wide amount of historical data collection without having to alter the core structure of the database. This has a few major implications:

- **Bridges the Catholic-Protestant Divide**
  While Protestant and Catholic organizational structures are quite different, the flexible database structure of the CHCD allows them to be recorded and analyzed together, something rarely done in the study of Christianity in China.

- **Allows for Multiple Forms of Belonging**
  Christian people moved between institutions, institutions changed locations, and corporate entities often split. The CHCD design makes it easy to track these changes while limiting redundancy.

- **Controls Complex and Fuzzy Geographies**
  Geography is regulated using two principles. First, the only nodes that have geographic coordinates attached to them are geography nodes (i.e. Village, Township, County, Prefecture, Province). Second, the only nodes which can relate to geography nodes are Institution and Event nodes. These principles, in turn, accomplish two main goals: 1) historical locations with varying levels of geographic specificity can be recorded, and 2) redundancy and errors are reduced. For more information, see the appendix.

- **Easy to Query**
  When users download the database and use its native Neo4j environment, they can utilize the easy-to-learn Cypher query language. Complex data schemas can be difficult to write queries for, making it difficult to derive meaningful data. By putting a whole host

of historical data into a single, simple schema, researchers can more easily analyze disparate sources.

- **Easy to Grow**

  The four primary node types are not the only kind of historical entities or information that could be placed in a database. While initial data collection will focus on these entities, the graph database structure can grow to include different kinds of information. This future growth will further enrich any data already recorded.

As mentioned above, more specific information on the graph database design and architecture can be found in the appendix to this document. CHCD's flexible, yet controlled structure allows for a vast trove of information to be placed into an understandable and accessible format. Moreover, this shared schema brings together information from sources that are traditionally not brought into dialogue with one another. English Protestant mission directories can be analyzed alongside Portuguese Jesuit correspondence; Chinese baptismal records can dialogue with Latin necrologies. Thus, this simple, but powerful schema opens up new possibilities for future research.

THE USER-FOCUSED ONLINE PLATFORM

Digital methodologies in the humanities often have a high learning threshold. Beyond learning the typical tools of their field, humanists often must also learn the tools of a specific approach. This often entails investment in expensive software, time learning computer languages, or additional training in new modes of analysis.

With these challenges in mind, the second major goal of the CHCD is the creation of a public online platform that will allow users to interact with the data from the database in an intuitive manner. Separate from the database itself, this online platform allows for anyone to explore the data and create visualizations that are useful for publications and personal research.

PHILOSOPHY

Researchers who are well-versed in the tools of the digital humanities will always be able to glean more from complex datasets than novices. Still, this should not mean that the major benefits of digital projects should go just to the techno-elites. As such, CHCD's online tool aims to provide analytical tools that are readily understandable to researchers. Philosophically, this means the online platform is guided by certain principles:

- **It is Understandable**

  The online platform is focused on translating the geographic and relational aspects of the database into comprehensible visualizations. When people use the platform, they should always be able to understand what the information they are looking at means.

- **It is Usable**

  Complex interfaces are not always necessary for complex analysis. As such, the online platform adheres to minimalistic and flat design principles that keep the important information and tools front and center for users.

- **It is Free to Use**

  It is hard to make a dramatic impact on the field when data is hidden behind paywalls and passwords. Like the data itself, the CHCD's online platform is online and free to use by the public.

- **It is an Entry Way**

  While the online platform will provide enough for most scholars of Christianity in China, the database can be coupled with tools such as ArcGIS or Gephi for even more detailed levels of analysis. As such, the online platform will also endeavor to resource scholars who would like to dive deeper into digital research methodologies.

DESIGN

The above principles inform the design and implementation of the online tool. Currently, the online tool provides three ways to explore and analyze the CHCD data.

- **Map View**

  The online tool's default mode of analysis is geographic. Users can search and filter persons and individuals as they relate to the geography of China. Filters such as time, nationality, affiliation, and gender allow users to set parameters for what institutions are displayed on the map. Users also have the option to employ a topographical map to the data which makes it possible to understand the relationship between Christian presence and the landscape.

- **List View**

  Some users, however, might prefer to interact with the data through a more traditional approach. "List View" allows users to apply the same filters and lenses to the data without the complex geographic visualizations. This approach to the data is ideal for helping scholars create lists of potential research subjects.

- **Network Maps**

  The website also allows users to explore the relationships between people and

institutions without referencing geography. The network section of the site allows users to create network maps for individuals in the database. These network maps allow users to explore both the direct connections that individuals had and the indirect connections they had through institutions or other people. These network views can help students and researchers discover connections that their sources may not have identified.

Within each mode of exploration, users can also explore the database through the "profiles" of each node. Much like Facebook, users can click on the name of an individual or institution to discover all the information and connections for that node in the database. If a user desires, they could then click on one of the connections to see the profile of that node, and so on and so on. This Facebook like approach is both natural to modern internet users and an easy way to navigate the complex data of the CHCD. Adherence to the right principles and these simple design choices ensure that the historical riches of the project remain accessible to all.

EAST ASIAN-WESTERN PARTNERSHIPS

The story of Christianity in China is both a local and a global phenomenon. Christianity took root in local contexts and shaped the local identities of many Chinese people. Yet, it was also the religion of many foreigners living and working in China. The complex interactions between China and the West have left a trail of archival materials that extends around the world. Furthermore, this trail is vast in size and rich in complexity.

As such, a large-scale project like the CHCD is reliant upon a host of partnerships between East Asian universities, Western universities, and archives and libraries around the globe. More specifically, however, the CHCD aims to create strategic partnerships that further four main goals.

- **Fostering an East-West Scholarly Community**
  Scholars of Christianity in China work across the globe in various disciplines and methodologies. CHCD pursues partnerships with East Asian universities in order to give scholars and students the opportunity to interact and learn from one another. Potential avenues for reaching this goal are conferences, workshops, graduate student exchanges, and visiting scholars.

- **Producing Publications**
  CHCD's rich historical data is ample ground for research papers, edited volumes, and monographs. CHCD pursues partnerships that will increase the exchange of information and ideas through the production of publications from both Western and Chinese scholars.

- **Sharing Research and Data Collection**
  Most importantly, CHCD pursues partnerships that aim to spread the workload of data collection and entry. Partnerships with teams around the world will open archival resources and expertise that any single university-based team could not muster alone.

- **Re-invigorating Scholarly Resources**

  The CHCD's wide vision means that it can also incorporate the contributions of past scholarly projects, some of which are threatened due to institutional or technological shifts. The CHCD pursues partnerships with these projects to ensure that the valuable contributions of scholars are not lost to future generations of students and researchers.

The technological and historical scope of this project cannot be accomplished without robust partnerships between scholars in the West and China. Happily, such partnerships create more than an abundance of data for research. The collegiality and scholarly friendships made possible through partnership ensure a healthier and more sustainable future for the field.

CONCLUSION

The China Historical Database (CHCD) is an ambitious project that seeks to overcome many of the challenges of studying Christianity in China. The challenge of complex and disparate resources is met by a cutting-edge graph database that is flexible enough to capture a wide array of source material and rigid enough to place that material in an understandable schema. Yet, such a technological approach also entails a barrier to entry for many who do not have the tools of digital scholarship. Thus, the CHCD online platform provides a way for students, teachers, and researchers to engage the richness of the historical data without having to spend inordinate amounts of time learning new skills. Furthermore, strategic partnerships between East Asian and Western universities can help break down the cultural divisions between the East and the West by giving scholars more chances to interact and exchange ideas about a religion that bounces between the local and the global.

In addressing these challenges, the CHCD opens a new world of analysis on the history of Christianity in China. Moreover, it creates a new kind of reference tool that is prosopographic, geographic, and bibliographic all at once. As the database grows, it will become an essential tool in the study of Christianity in China.

APPENDIX: DATABASE ARCHITECTURE

While the sections above provides a basic overview of the China Historical Christian Database's (CHCD) design, this appendix provides a detailed description of the infrastructure and architecture of the database itself. This information is provided for researchers who intend to use the database (not the online platform) for their own research. Understanding the data structure is essential for creating meaningful queries. For a model of the graph schema, see "The Database" section above.

INFRASTRUCTURE

CHCD utilizes the open-source graph platform Neo4j to create and run its database. While the data can be ported into other graph database formats, it is recommended that users familiarize themselves with the Neo4j platform. Users who are new to the world of database design are highly recommended to use Neo4j as it employs the easy to learn Cypher query language.

Documentation and more information about Neo4j can be found here.

GEOGRAPHY IN THE CHCD

Historical sources will communicate spatial data in a variety of interesting ways that are not always easy to translate into a granular data structure. Spatial references morphed to meet people's needs and relationships that existed in certain spaces are not always recorded. As such, CHCD has had to implement two unique features regarding how it applies geography to historical sources.

MULTI-LEVEL GEOGRAPHY USING MODERN ADMINISTRATIVE DISTRICTS

Historical sources might locate individuals or institutions using a street address, a county, or just a province. Capturing these fuzzy geographic references and placing them in some form of space requires varying degrees of geographic specificity. In other words, to place every historical spatial reference on a map, spatial references need to roughly match the exactness of the historical source. Administrative units are the most readily available model for mapping these various levels of geography.

This, of course, causes an immediate problem in a place like China. Which historical administrative break-down should be utilized? If historical place names and administrative units were utilized, it would create several problems: 1) it would be more difficult to track spatial relationships over time as the same place might be represented by various nodes in the database, and 2) it would require the time-intensive task of creating a gazetteer of place names and coordinates, a task that would easily take up all the time of the project.

As such, the CHCD has chosen to represent the fuzzy geography of historical sources using point data for the modern administrative map of China (c. 2009). Centroid points are created for each administrative unit and used for locations in the database. Where more granular data is available (e.g. village names and street addresses), researchers locate places individually and place them in the modern geography. While this runs the risk of reading modern geography into the past (a risk inherent in any historical project), this approach does offer two benefits: 1) historical subjects from various time frames are placed on the same geographic field, allowing researchers to analyze spatial relationships between them, and 2) researchers and students do not have to initially navigate historical administrative divisions to begin understanding the basics of the data.

CONTROLLED GEOGRAPHIC INTERACTION

Secondly, geography is only created when Institution nodes or Event nodes are related to one of the geographic nodes of the database. This initially seems counter-intuitive as historical persons also inhabited space and moved between spaces. Allowing persons to have geography, however, has two main problems: 1) it increases the chance of redundancy and errors and 2) it is more work. The CHCD solves both of these problems through by controlling interaction with geographic nodes.

**Reduces the Chance of Redundancy and Errors.**

Recording the location of people and institutions separately is a clear form of redundancy. Take the following example; Person A is present at Church B and both are in Village C. If these relationships were recorded separately the graph would look like this.
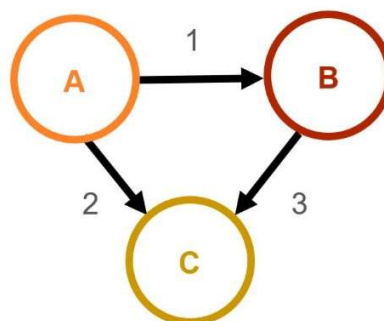
In this schema, relationships 2 and 3 are communicating essentially the same data. Moreover, if the institution were to change cities, two new relationships would need to be created. Errors are more probable because of this redundant relationship. For example, if a researchers discovered that the institution was actually location in a nearby village, Village D, they would go to the database and change relationship 3. In order to update the relevant individual location they would need to query the database for a list of all personnel who are related to Institution B and Village C, and then they would need to manually update them. Any human slip could cause a massive amount of erroneous data. The below CHCD schema eliminates this redundancy, reduces the number of relationships needed to communicate the same information, and minimizes the potential for erroneous data.
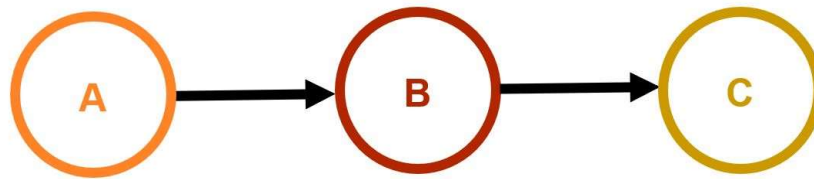
Of course, the non-institutionally and non-event linked individuals cause a problem for this schema. If all that is known about an individual is that they are in an area, the schema requires a default geographically linked institution to place them. As such, CHCD utilizes placeholder institutions in such cases. Placeholder institutional nodes are named in the following way "General Area (Location Name)" (e.g. "General Area (Beijing)").

Without care, researchers can interpret institutional data wrongly if they do not take account of these placeholder institutions which exist to allow individuals to obtain geography. Most individuals in the database, however, are linked to events and institutions

Decreases the Amount of Work.

As demonstrated above, giving individuals unique relationships to geography can dramatically increase the amount of work needed to update historical markers. In this instance, controlling the connectivity between nodes allows the data structure to remain easy to update and modify without a massive amount of repeated actions. Consider the above example again. If there were 10 individuals connected to Institution B, a direct person-to-geography model would require creating 21 relationships. If the Village was modified, it would require altering 11 of those relationships. The CHCD model only requires 11 relationships to convey the same amount of information and the same change would only require modifying 1 relationship.

NODE TYPES

The As defined above, there are four primary node types (Person, Corporate Entity, Institution, Event) in the database and five geographic node types (Village, Township, County, Prefecture, Province). Definitions of each primary node type are below:

- **Person:** these nodes represent living human beings. People are at the core of the database and thus they have the most kinds of relationships possible.

- **Corporate Entity:** these nodes represent organizations that *do not* have a direct geographic footprint. For example, the Society of Jesus is an organization, but it only exists in space through people and institutions.

- **Institutions:** these nodes represent organizations that *do have* a direct geographic footprint. Common examples in the database include churches, hospitals, and schools.

- **Events:** these nodes represent important events that took place in Chinese Christianity. Events are, by definition, temporary happenings that have specific geographic locations. Examples in the database range from Christian conferences to imperial hunting parties.

NODE PROPERTY VALUES

Each node type has a variety of property values that can be provided. It is rare for nodes to have an entry for each value. Below are tables detailing the property values for each node type with a short description of what that value is and any rule governing it.

:Person Node

family_name_en:	Family name of individual. Romanized script used by that individual.
marriage_name_en:	Family name of individual taken after marriage. Romanized script used by that individual.
given_name_en:	Given name of individual. Romanized script used by that individual.
alt_name_en:	A semi-colon separated list of alternate names for that person. Romanized script. This may include nicknames, alternate spellings, etc.
bap_name_en:	Baptismal name of individual. Uses the Romanized script used by that individual.
family_name_zh:	Family name of individual. Traditional Chinese script.
marriage_name_zh:	Family name of individual taken after marriage. Traditional Chinese script.
given_name_zh:	Given name of individual. Traditional Chinese script.
alt_name_zh:	A semi-colon separated list of alternate names for that person in traditional Chinese script. This may include nicknames, alternate spellings, etc.
bap_name_zh:	Baptismal name of individual. Traditional Chinese script.

family_name_py:	Family name of individual. Modern pinyin script.
marriage_name_py:	Family name of individual taken after marriage. Modern Pinyin script.
given_name_py:	Given name of individual. Modern Pinyin script.
bap_name_py:	Baptismal name of individual. Modern Pinyin script.
alt_name_py:	A semi-colon separated list of alternate names for that person. Modern Pinyin script. This may include nicknames, alternate spellings, etc.
birth_day:	Day of birth. Alpha numeric, 1-31.
birth_month:	Month of birth. Alpha numeric, 1-12.
birth_year:	Year of birth. Alpha numeric.
birth_city:	City of birth. Uses modern English name of location.
death_day:	Day of death. Alpha numeric, 1-31.
death_month:	Month of death. Alpha numeric, 1-12.
death_year:	Year of death. Alpha numeric.
death_city:	City of death. Uses modern English name of location.
gender:	Male, Female, Other
nationality:	Nationality of person. In English script. Uses state that existed at the time.
notes:	Other material that does not fit in the above categories.
sources:	List of sources utilized for the above information.

:Institution Node

inst_name_en:	Name of Institution. Common English format (e.g. Church of the Sacred Heart)
alt_inst_name_en:	A semi-colon separated list of alternate names. This may include nicknames, alternate spellings, the name in the original language, etc.
inst_name_zh:	Name of Institution. Traditional Chinese script.
alt_inst_name_zh:	A semi-colon separated list of alternate names. Traditional Chinese script. This may include nicknames, alternate spellings, etc.
inst_name_py:	Name of Institution. Modern Pinyin script.
alt_inst_name_py:	A semi-colon separated list of alternate names. Modern Pinyin script. This may include nicknames, alternate spellings, etc.
start_day:	Day of start. Alpha numeric, 1-31.
start_month:	Month of start. Alpha numeric, 1-12.
start_year:	Year of start. Alpha numeric.
end_day:	Day of end. Alpha numeric, 1-31.
end_month:	Month of end. Alpha numeric, 1-12.
end_year:	Year of end. Alpha numeric.
inst_type:	See section below, "Type Lists"
nationality:	Nationality of institution. In English script. Uses state that existed at the time.
notes:	Other material that does not fit in the above categories.
sources:	List of sources utilized for the above information.

:CorporateEntity Node

corp_name_en:	Name of Corporate Entity. Common English format (e.g. Society of Jesus)
alt_corp_name_en:	A semi-colon separated list of alternate names. This may include nicknames, alternate spellings, the name in the original language, etc.
corp_name_zh:	Name of Corporate Entity. Traditional Chinese script.
alt_corp_name_zh:	A semi-colon separated list of alternate names. Traditional Chinese script. This may include nicknames, alternate spellings, etc.
corp_name_py:	Name of Corporate Entity. Modern Pinyin script.
alt_corp_name_py:	A semi-colon separated list of alternate names. Modern Pinyin script. This may include nicknames, alternate spellings, etc.
start_day:	Day of start. Alpha numeric, 1-31.
start_month:	Month of start. Alpha numeric, 1-12.
start_year:	Year of start. Alpha numeric.
end_day:	Day of end. Alpha numeric, 1-31.
end_month:	Month of end. Alpha numeric, 1-12.
end_year:	Year of end. Alpha numeric.
corp_type:	See section below, "Type Lists"
nationality:	Nationality of institution. In English script. Uses state that existed at the time.
notes:	Other material that does not fit in the above categories.
sources:	List of sources utilized for the above information.

:Event Node

event_name_en:	Name of Event. Common English format (e.g. The 2nd National Congress)
alt_event_name_en:	A semi-colon separated list of alternate names. This may include nicknames, alternate spellings, the name in the original language, etc.
event_name_zh:	Name of Event. Traditional Chinese script.
alt_event_name_zh:	A semi-colon separated list of alternate names. Traditional Chinese script. This may include nicknames, alternate spellings, etc.
event_name_py:	Name of Event. Modern Pinyin script.
alt_event_name_py:	A semi-colon separated list of alternate names. Modern Pinyin script. This may include nicknames, alternate spellings, etc.
start_day:	Day of start. Alpha numeric, 1-31.
start_month:	Month of start. Alpha numeric, 1-12.
start_year:	Year of start. Alpha numeric.
end_day:	Day of end. Alpha numeric, 1-31.
end_month:	Month of end. Alpha numeric, 1-12.
end_year:	Year of end. Alpha numeric.
event_type:	See section below, "Type Lists"
notes:	Other material that does not fit in the above categories.
sources:	List of sources utilized for the above information.

:Village / :Township / :County / :Prefecture / :Province Node

name_en:	Name. Common English (e.g. Puzhou Village)
name_zh:	Name. Traditional Chinese script.
event_name_py:	Name. Modern Pinyin script.
latitude:	Latitude in decimal format.
longitude:	Longitude in decimal format.

RELATIONSHIP TYPES

Each relationship type has a variety of property values that can be provided. It is rare for relationships to have an entry for each value. Below are tables detailing the property values for each relationship type with a short description of what that value is and any rule governing it.

:RELATED_TO
This is a relationship between two people nodes.

rel_type:	See section below, "Type Lists"
start_day:	Day of start. Alpha numeric, 1-31.
start_month:	Month of start. Alpha numeric, 1-12.
start_year:	Year of start. Alpha numeric.
end_day:	Day of end. Alpha numeric, 1-31.
end_month:	Month of end. Alpha numeric, 1-12.
end_year:	Year of end. Alpha numeric.
notes:	Other material that does not fit in the above categories.
sources:	List of sources utilized for the above information.

:PRESENT_AT
This is a relationship between 1) a person node and an institution node or 2) a person node and an event node.

role:	See section below, "Type Lists"
start_day:	Day of start. Alpha numeric, 1-31.
start_month:	Month of start. Alpha numeric, 1-12.
start_year:	Year of start. Alpha numeric.
end_day:	Day of end. Alpha numeric, 1-31.
end_month:	Month of end. Alpha numeric, 1-12.
end_year:	Year of end. Alpha numeric.
notes:	Other material that does not fit in the above categories.
sources:	List of sources utilized for the above information.

:AFFILIATED_WITH
This is a relationship between a person node and corporate entity node.

rel_type:	See section below, "Type Lists"
start_day:	Day of start. Alpha numeric, 1-31.
start_month:	Month of start. Alpha numeric, 1-12.
start_year:	Year of start. Alpha numeric.
end_day:	Day of end. Alpha numeric, 1-31.
end_month:	Month of end. Alpha numeric, 1-12.
end_year:	Year of end. Alpha numeric.
notes:	Other material that does not fit in the above categories.
sources:	List of sources utilized for the above information.

:PART_OF
This is a relationship between an institution node and corporate entity node.

rel_type:	See section below, "Type Lists"
start_day:	Day of start. Alpha numeric, 1-31.
start_month:	Month of start. Alpha numeric, 1-12.
start_year:	Year of start. Alpha numeric.
end_day:	Day of end. Alpha numeric, 1-31.
end_month:	Month of end. Alpha numeric, 1-12.
end_year:	Year of end. Alpha numeric.
notes:	Other material that does not fit in the above categories.
sources:	List of sources utilized for the above information.

:CONNECTED_TO
This is a relationship between two institution nodes or two event nodes.

rel_type:	See section below, "Type Lists"
start_day:	Day of start. Alpha numeric, 1-31.
start_month:	Month of start. Alpha numeric, 1-12.
start_year:	Year of start. Alpha numeric.
end_day:	Day of end. Alpha numeric, 1-31.
end_month:	Month of end. Alpha numeric, 1-12.
end_year:	Year of end. Alpha numeric.
notes:	Other material that does not fit in the above categories.
sources:	List of sources utilized for the above information.

:TOOK_PLACE_IN

This is a relationship between events and geographic nodes.

start_day:	Day of start. Alpha numeric, 1-31.
start_month:	Month of start. Alpha numeric, 1-12.
start_year:	Year of start. Alpha numeric.
end_day:	Day of end. Alpha numeric, 1-31.
end_month:	Month of end. Alpha numeric, 1-12.
end_year:	Year of end. Alpha numeric.
notes:	Other material that does not fit in the above categories.
sources:	List of sources utilized for the above information.

:PRESENT_IN

This is a relationship between institutions and geographic nodes.

day:	Day of start. Alpha numeric, 1-31.
month:	Month of start. Alpha numeric, 1-12.
year:	Year of start. Alpha numeric.
notes:	Other material that does not fit in the above categories.
sources:	List of sources utilized for the above information.

:INSIDE_OF

This relationship type has no property values and only denotes hierarchy. It is only used by geographic nodes.

TYPE LISTS

Certain node and relationship property values need to be restricted in order to ensure that queries on the database return the appropriate value. For example, a school and university are similar but distinct types of institutions. The below lists are the most up-to date lists of types for each type list.

:RELATED_TO | rel_type:

Assistant to	Servant to	Travelling Companion
Ordained by	Spouse	

:PRESENT_AT | role:

Acting Principal	In Charge	Principal
Assistant Master	Internee	Superintendent
Assistant Mistress	Located at	Superior
Director	Officer	Teacher
Founder	President	

:AFFILIATED_WITH | rel_type:

Adjutant	Captain	Member
Administrator	Catechist	Monsignor
Aide	Chaplain	Novice
Ambassador	Commandant	Priest
Ambassador to Rome	Commissioned	Procurator
Apostolic Vicar	Doctor (Unspecified)	Provincial
Astronomer	Ensign	Registered Nurse
Avowed	Father Coadjutor	Reverend
Bachelor of Arts	Fellow	School Readiness Nurse
Bachelor of Divinity	Founder	Superior
Bachelor of Medicine	Licentiate	Vice-Provincial
Bachelor of Science	Major	Visitor
Bishop	Master of Arts	
Cadet	Medical Doctor	

:PART_OF | rel_type:

Mission of	Outpost of	Educational Facility of

:CONNECTED_TO | rel_type:

No records currently have this relationship.

:Institution | inst_type:

Church	Protestant Mission Station	Residence

:Event | event_type:

Boat Embarkment		

:CorporateEntity | corp_type:

Chinese Governmental Body	Notable Family	Roman Catholic Organization (Men)
Christian Denomination	Professional Guild	Roman Catholic Organization (Women)
Jesuit Mission	Protestant Mission Organization	Roman Catholic Vicariate
Jesuit Province	Roman Catholic Diocese	

ABBREVIATED BIBLIOGRAPHY

The following bibliography offers an abbreviated list of resource which have informed the approach of this current project and can provide some context to those unfamiliar with the intersection of digital humanities and history.

Bays, Daniel H. *A New History of Christianity in China.* New York: Wiley, 2011.

Bodenhamer, David J. "History and GIS: Implications for the Discipline." In *Placing History: How Maps, Spatial Data, and GIS Are Changing Historical Scholarship.* Edited by A. K. Knowles and A. Hillier, 219-234. Redlands, CA: ESRI Press, 2007.

Bol, Peter K. "Creating a GIS for the History of China" in *Placing History: How Maps, Spatial Data, and GIS Are Changing Historical Scholarship*, 27-60. Edited by A. K. Knowles and A. Hillier. Redlands, CA: ESRI Press, 2007.

———. "What do Humanists Want? What do Humanists Need? What Might Humanists Get?" In *Geohumanities: Art, History, Text at the Edge of Place*, 296-308. London and New York: Routledge, 2011.

Dunch, Ryan. *Fuzhou Protestants and the Making of a Modern China, 1857-1927.* New Haven, CT: Yale University Press, 2001.

———. "Beyond Cultural Imperialsim: Cultural Theory, Christian Missions, and Global Modernity." *History and Theory* 41, October 2002, 301-325.

Gregory, Ian, and Paul S. Ell. Historical GIS: Technologies, Methodologies, and Scholarship. Cambridge : Cambridge University Press, 2012.

Guldi, Jo. "The Spatial Turn in Religion." Institute for Enabling Geospatial Scholarship at the Scholars' Lab at the University of Virginia Library. Accessed 27 March 2019. http://spatial.scholarslab.org/spatial-turn/the-spatial-turn-and-religion/index.html.

———. "What is the Spatial Turn?" Institute for Enabling Geospatial Scholarship at the Scholars' Lab at the University of Virginia Library. Accessed 27 March 2019. http://spatial.scholarslab.org/spatial-turn/what-is-the-spatial-turn/.Harrison, Henrietta. *The Missionary's Curse and Other Tales from a Chinese Catholic Village.* Berkeley, CA: University of California Press, 2013.

Hunt, Lynn. Writing History in a Global Era. New York: W.W. Norton & Company, 2014.

Menegon, Eugenio. *Ancestors, Virgins, and Friars: Christianity as a Local Religion in Late Imperial China.* Cambridge: Harvard University Press, 2010.