



**ARNOLD, MATTHIAS**

**RUDOLPH, HENRIKE**

## **Network Data in the Early Chinese Periodicals Online Database (ECPO)**

*Journal of Historical Network Research* x  
(202x) xx-xx

### Abstract

This paper introduces the Early Chinese Periodicals Online database (ECPO), based at the University of Heidelberg. It offers a short overview of the development of this research-driven database as it grew to incorporate additional projects. Today, the ECPO hosts a variety of textual and visual sources from the Republican period (1912-1949), as well as biographical data of twentieth-century academic and political elites. We aim to show how this database can be used and extended to pursue historical network research, even for scholars previously unfamiliar with digital infrastructures. The concluding part discusses privacy concerns in data curation and management, which are increasingly relevant for studies of twentieth-century actors and limit the potential to release and reuse these datasets, especially if they pertain to the People's Republic of China's political history.

## 1 Introduction\*

The late Qing and Republican era (mid-nineteenth to mid-twentieth century) is considered one of the most productive periods of Chinese print culture. Books, journals, newspapers, leaflets, pamphlets, and posters printed in the urban centers circulated across the country in vast numbers. With millions of pages, the sheer quantity of Republican sources is beyond what researchers can read and digest during a lifetime. Therefore, historians were only able to subject a fraction of these publications to close reading. Digitization efforts in the People's Republic of China, the Republic of China (Taiwan), and beyond have created databases that enable historians to re-engage with these sources and pursue qualitative and quantitative inquiries that are unfathomable to previous generations of scholars. This article introduces the Early Chinese Periodicals Online (ECPO) infrastructure and its data subsets, focusing on its potential use for historical network analysis.

### 1. From Image Scans to Full Text

In the past two decades, even though Chinese full-text databases grew at an impressive speed, most digitization projects only produced image scans, even in the People's Republic of China. Specialist Greenapple Changsha's endeavor to double-key the *Shun Pao* (*Shenbao* 申報, 1872-1949) database to produce fully searchable texts and complete author-title indexes is an exception in this regard. This picture started to gradually change when the "Global Press Archive" from East View Information Services, Inc. and the "Chinese Newspaper and Periodical Index" (*Quanguo baokan suoyin* 全國報刊索引) from Shanghai Library recently began to offer individual newspapers in full text.

The "Early Chinese Periodicals Online" (ECPO)<sup>1</sup> database is no exception. The ECPO has two major precursor projects. The first introduced Chinese entertainment newspapers (*xiaobao* 小報) as an important primary source for research on late Qing and Republican period cultural and social history in an online database. Inspired by Western entertainment tabloids, *xiaobao* became

---

\* **Corresponding author:** Matthias Arnold, <arnold@hcts.uni-heidelberg.de>

1 Early Chinese Periodicals Online (ECPO), <https://uni-heidelberg.de/ecpo>.

one of the most popular media for the daily publication of entertainment news, serialized fiction, short stories, social and political commentary, and cartoon drawings in China. The “Chinese Entertainment Newspapers” database<sup>2</sup> was designed to “present information about the particular profile of a given entertainment paper, and to highlight the kinds of information that might be found there.”<sup>3</sup>

The second predecessor for the ECPO grew out of an international network of scholars in the project “A New Approach to the Popular Press in China: Gender and Cultural Production, 1904-1937,” which resulted in the database “Chinese Women’s Magazines in the Late Qing and Early Republican Period” (WoMag).<sup>4</sup> The digitization of four of the most important Chinese women’s journals from the Republican period allowed the researchers to trace common tropes, reconstruct discourses, and compare editorial styles across publications, as well as relating them to other sources and the broader historical context. The project established a web-based repository to provide access to the image scans in a readable size, compiled a detailed author-title index enhanced with analytic metadata and cross-references.

Both approaches, the “extensive” analysis of publications in general for the Xiaobao database, and the “intensive” recording of all individual items for the WoMag database, became the basis of the ECPO project where they are combined and elevated to a larger scale. As shown in fig. 1, the WoMag database with all scans and metadata is still available online, as is the Xiaobao database with its selected contents. The full scans of the Xiaobao periodicals have recently

---

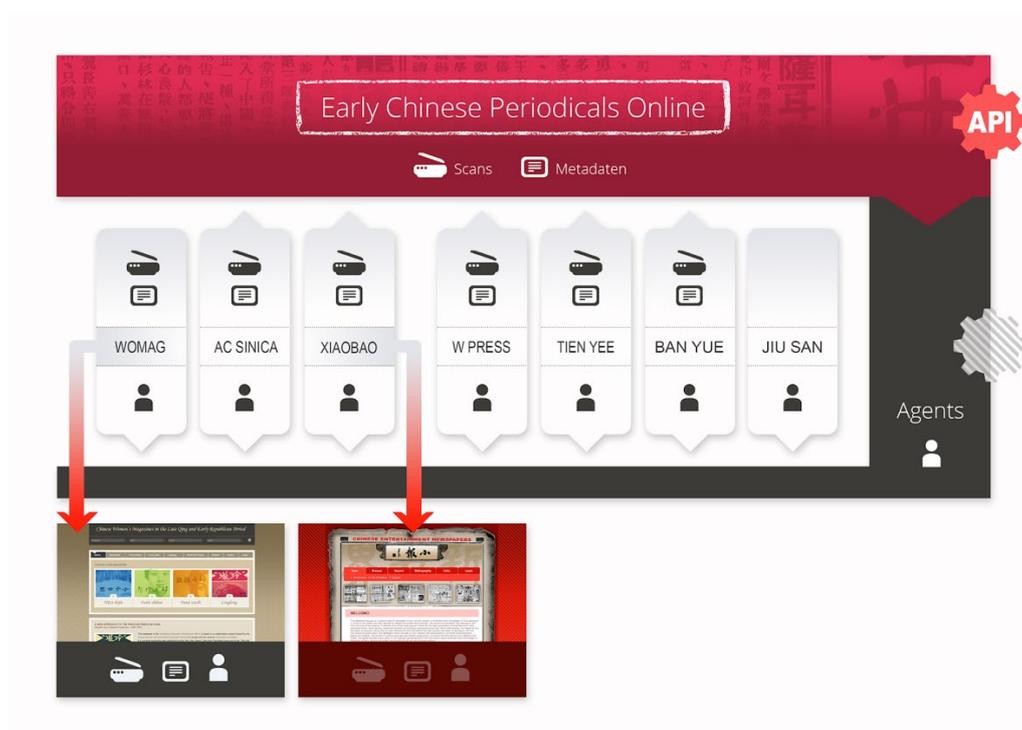
2 *Chinese Entertainment Newspapers Database* (Heidelberg University, 2006), <http://xiaobao.uni-hd.de>.

3 Catherine V. Yeh, “Entertainment Newspapers and the Rise of Mass Culture,” *Chinese Entertainment Newspapers*, 2006, <http://projects.zo.uni-heidelberg.de/xiaobao/index.php?p=intro>.

4 *Chinese Women’s Magazines in the Late Qing and Early Republican Period (WoMag)* (Heidelberg University, 2009–2020), <https://uni-heidelberg.de/womag>. See also Doris Song, Liying Sun, and Matthias Arnold, “The Birth of a Database of Historical Periodicals: Chinese Women’s Magazines in the Late Qing and Early Republican Period,” *Tulsa Studies in Women’s Literature* 33, no. 2 (2014): 227–37; Michel Hockx, Joan Judge, and Barbara Mittler, eds., *Women and the Periodical Press in China’s Long Twentieth Century: A Space of Their Own?* (Cambridge: Cambridge University Press, 2018), <https://doi.org/10.1017/9781108304085>.

## Network Data in ECPO

become accessible in the ECPO. As a result of the data recording workflow, the project recorded about 50,000 names of “agents,” which not only contains the names of authors or photographers, but also includes persons, groups, or institutions that are mentioned in the texts or depicted in images. Originally a mere by-product, these names are now organized in a dedicated database that provides an “agent service” to all ECPO modules. We use the generic term “agent” to refer to individuals, groups (such as photo studios or sports teams), institutions, or firms that are tied to an “item,” meaning any text (“article”), image, or advertisement within the journals. We are currently preparing a user frontend for the agent service to allow easy browsing of the agents’ data and provide a platform for basic export functionalities (API).



**Figure 1:** ECPO - data structure with agent service and sub-projects.

This significant database development was part of a collaborative project with the Institute of Modern Chinese History at the Academia Sinica, for which we received generous funding from the Chiang Ching-kuo Foundation for

International Scholarly Exchange. The cooperation led to a significant expansion of ECPO's content and metadata. Since then, data from several individual projects has been added to the platform. For example, the research outcome of a project on the anarchist magazine *Tien Yee* (*Tianyi* 天義, 1907-08),<sup>5</sup> a complete set of the literary magazine *The Half Moon Journal* (*Banyue* 半月, 1921-25), or the collection of "Western Publications Printed in China" by the late Professor Rudolf Wagner, including *The Canton Register* (1827-43) and *The Canton Press* (1835-44). The database is now an open-access resource with a user-friendly frontend that has been online since 2016.

Today, the ECPO incorporates entertainment journals, women's magazines, periodicals from the Academia Sinica collection in Taipei, and a collection of Western periodicals printed in China. As of September 2020, it hosts 308 publications comprising over 300,000 image scans. For 134 of these publications, we manually compiled additional content information about items, including articles, images, and advertisements.

## 2. The Agent Service

In order to trace appearances across journals, the mere recording of names as they appeared in headlines and subtitles proved to be insufficient, for the following reason: first, a single Chinese author might have published under a variety of names, ranging from his or her personal name (without surname), well-known pen names, or even pseudonyms intended to mask their identity. Second, Republican Chinese publications embraced cosmopolitanism. They covered Hollywood gossip, reprinted the speeches of Japanese intellectuals, or reported on social events in the foreign concessions of Shanghai. The journals thus frequently feature Chinese transcriptions of foreign personal, institutions, or place names.

Starting with Western names, i.e., Europeans and Americans, which Chinese sources transcribed in Chinese characters, the project has begun to assign the

---

5 Matthias Arnold and Lena Hessel, "Transforming Data Silos into Knowledge: Early Chinese Periodicals Online (ECPO)," in *E-Science-Tage 2019: Data to Knowledge*, ed. Vincent Heuveline, Fabian Gebhart, and Nina Mohammadianbisheh (Heidelberg: heiBOOKS, 2020), 106 and note 28.

## Network Data in ECPO

different names to their respective individuals or institutional bodies. Since the transcription of Western names was not standardized in the Republican era, certain individuals occur with several variant names in Chinese, not to mention the various misspellings when the Latin script was used. This is why, for example, the actress Constance Bennett has 26 different Chinese name variants recorded from resources in ECPO, most of which were hitherto unknown. From the user's perspective, whatever name variant is used as a search term, every mention of the person that this name variant is assigned to will show up in the search results.

The agent service allows us to distinguish between names (name records) and agents (agent records) and provides functionality to assign different names and occurrences to their respective agents. ECPO records all names as they appear in the publications, including pen names, studio names, or possible misspellings. We developed a module to merge names, occurrences, and additional metadata into one agent across the databases. It is also possible to separate names or extract individual names from one agent to assign them to another. In some cases, the disambiguation of individuals is very tricky, sometimes impossible. We collect these names in name records. For example, we assigned the 340 occurrences of items signed with "editor" (*bianzhe* 編者) to one single name record. We did the same with acronyms, such as Ping 平, or "Y.D." Identifying the agents "behind" names may require quite an amount of academic research, but we are steadily making progress.

As of today, the project has 54,932 name records assigned to 48,124 agents. The lists still require substantial editing and cleaning before they can be used for substantial historical analysis. Yet, recording occurrences and co-occurrences of individuals and other entities in texts, pictures, and advertisements allows us to investigate previously neglected historical actors and their networks beyond the well-researched elites.<sup>6</sup>

### 3. Relational Data in ECPO

The most basic unit of relational data in the women's magazine dataset was the connections between various agents and texts or pictures. Persons could be

---

<sup>6</sup> Arnold and Hessel, "Transforming Data Silos into Knowledge."

either actively connected to texts or photographs as authors or creators, or passively by being mentioned in an article or appearing in an image, coded, for example, as “author of” or “mentioned in.” The same logic of either active or passive appearance applies to institutions, companies, or organizations. This distinction allows us to not only create two-modal datasets, but also to make meaningful conversions into one-mode datasets to analyze co-authorship and co-occurrence networks. Person-to-person datasets can map, for example, who wrote about whom, who appeared in the same picture or was mentioned in the same set of articles.

As an example, we can examine the two-mode dataset of Song Qingling, a leading women’s activist and wife of Sun Yat-sen, the founding father of the Chinese Republic. Apart from the articles she authored, Song Qingling appears in the publications collected as part of the ECPO framework in eight instances: Four articles, one image, and three advertisements. The latter advertised the newest books of the Life Publishing House (*Shenghuo shudian* 生活書店). Thus, the books of Song Qingling were promoted alongside those, for example, of the president Chiang Kai-shek (Jiang Jieshi). As this example shows, the distinction between “mentioned in advertisement” and “depicted in image” are very different analytical categories. A co-occurrence in a publisher’s advertisement alongside other prominent figures speaks to a good reputation, but does not reveal anything about personal ties between the individuals. A personal encounter is, however, a prerequisite for being in the same photograph. The clear distinction between these categories within the dataset thus allows for a nuanced analysis of network structures.

#### 4. The Expansion of ECPO: Beyond Republican Journals

With its potential to give structure to complex historical datasets, the agent service has been further enhanced since 2018. Now, the database also allows the mapping of connections between persons, persons and events, as well as between persons and institutions or organizations without a direct reference to a digitized source. As each individual, each event, and each institution/organization is treated like an independent “agent,” the data structure

## Network Data in ECPO

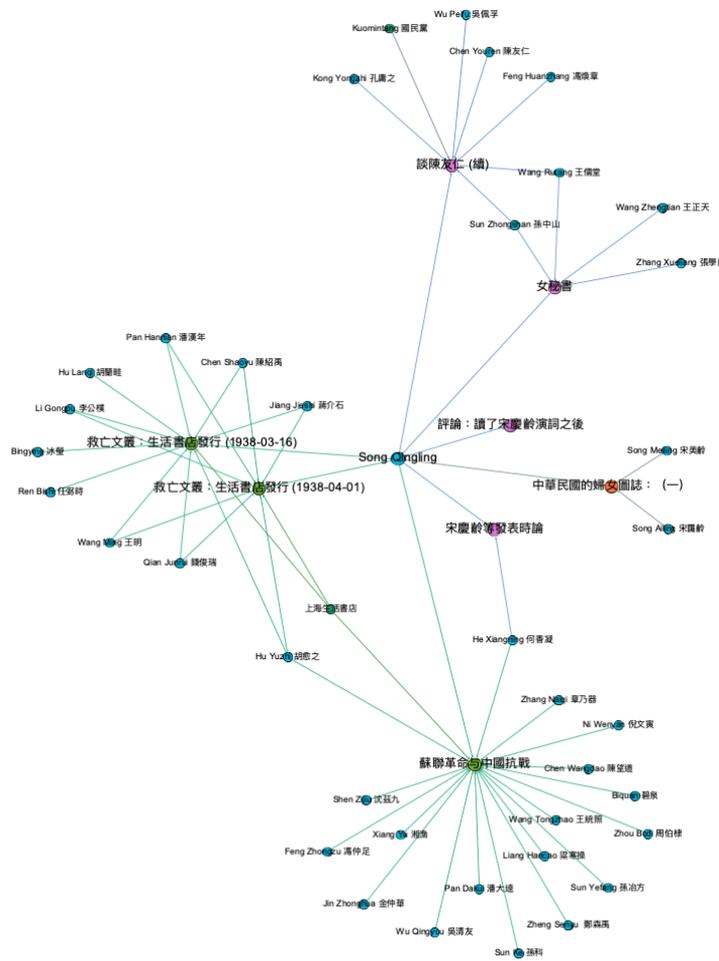
still allows us to assign a variety of names to a single agent and to merge data entries across different datasets.

This move was driven by the need to incorporate research data of a project exploring the development of the networks of the Jiusan Study Society (*Jiusan xueshu* 九三學社) from the 1940s to the early 2000s. The Jiusan is one of the so-called 'democratic parties and groups' with roots in the Republican period, which were incorporated in the new Communist regime after 1949. This project uses previously neglected archival and published party documents as well as secondary sources and explores the channels of communication and control that the Jiusan established between party members and CCP cadres. A network approach allows us to trace the temporal and spatial dimensions of the Jiusan's development, and offers new insights into its functioning throughout the turbulent years of political campaigns and economic reforms since the 1940s.

Most of the members of the Jiusan are, to this day, highly qualified medical or technical experts, scientists, or university lecturers. Therefore, each member is acting in different roles and is embedded in various social and political contexts. For example, each Jiusan member is tied to relatives, a place of work, a field of expertise, a local branch of the Jiusan Study Society, and possibly even to a political sub-group within this local branch or as a representative of the Jiusan to a regional or national representative body. We drafted a data structure that can host both attribute data (e.g., birthplace or gender) and complex relational data. The resulting dataset records both one-modal data (person-person) as well as bi-modal data (person-institution/ organization) in its temporal and spatial dimensions. Additionally, relationship types such as "teacher at," "head of," "member of," and so forth enable us to design queries that help to explore hierarchies, as well as the social and cultural capital at work within the Jiusan network.<sup>7</sup>

---

7 Generous support from the Friedrich-Alexander University Erlangen-Nürnberg and the Wikimedia Foundation, the Volkswagen Foundation, and the Stifterverband enabled the process of data coding and the expansion of the agent service.



**Figure 2:** Two-mode ego-network of Song Qingling in ECPO. Persons in blue, articles in purple, advertisements in light green, organizations in dark green, and images in orange. Visualization created in Gephi.

In projected one-mode networks, tie weights can offer additional insights. They can, for example, help to quantify if an appearance in the same picture was a coincidence or if there is strong evidence that two or more people socialized regularly. Simultaneously, the database offers possibilities to investigate social divisions within the elites. Did bankers and movie stars, for example, mingle, or did they form distinct circles? Because we expanded the agent service to allow for the storing of additional information, including the age, the date and place

of birth, and further information from a person's CV, we can address such questions. However, to date, only a small number of agents' entries possess such additional information as it has to be entered manually.

## 5. The Agent Service Data Structure

Following the structural logic of network data, the agent service is divided into attribute data ("name data" and "base data") and relational data ("ECPO assignments" and "agent relations"). For each agent, we can add an infinite number of name variants, as many as are found in the historical sources. However, for the rest of the information stored in the base data, we have to make decisions, such as in cases where sources give conflicting information on the date and place of birth.<sup>8</sup> Yet, in most cases, the birth dates deviate by only one or two years, and locations only because place names have changed. We also create ontologies that are not necessarily explicit in the sources, such as nationalities or institutional natures. The initial aim of the agent service was to trace agents across publications. In the Jiusan project, however, the sources recede into the background, and the agents' relations and interactions are of particular interest. This not only includes relationships between agents, for example family ties (person-person) or institutional affiliations (person-institution), but also personal encounters and cooperation. We thus created a special category: "events." Events can be a variety of social, professional, or political gatherings, ranging from participation in political councils to academic conferences. Furthermore, the publication of a book or article is also coded as an event. Each event is described with a start and end date, a location, and a host. In the case of an academic event, this could be, for example, the university or research institute hosting a conference. A second host institution can be added with its respective location, in case of a cooperatively organized event. For publications, the location of the publishing house is added.

---

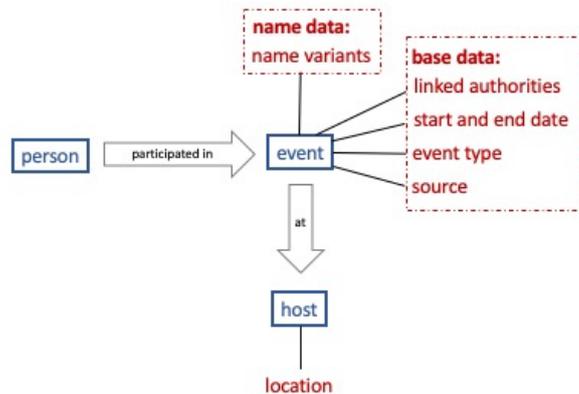
<sup>8</sup> Because conflicting source information is not a frequent problem here, we decided against a more work-intensive factoid approach that records all information as provided by different sources. Michele Pasin and John Bradley, "Factoid-Based Prosopography and Computer Ontologies: Towards an Integrated Approach," *Digital Scholarship in the Humanities* 30, no. 1 (April 2015): 86–97.

Below you can see how we code information, such as a person's participation in an academic conference, in three consecutive steps (fig. 3): First, an "agent" of the type "person" is created (if they are not already part of the dataset) with a unique ID and all relevant name and base information (see tab. 1). In the second step, in a separate table, an event entry is added with a unique ID, and again the relevant name and base data for this particular event. In a third table, the institution or organization, which was hosting the event (or, if the event is a publication, published it) is added.

Attribute data		
	Name data	Base data
1	Surname	Type (person, organization, institution)
2	First name	Institution nature (academic, political)
3	Surname in pinyin	Nationality (or national affiliation for organizations or institutions)
4	First name in pinyin	Field of expertise (biology, chemistry, etc. Only for persons)
5	Language (Chinese, English, Japanese, etc.)	Gender (only for persons)
6	Script (Chinese characters, Latin, Cyrillic script, etc.)	Date birth/founded
7	Name type (pen name, given name, other name)	Location birth/founded
8		Date death/dissolved
9		Location death/dissolved
10		Linked authorities ID (GND, VIAF, Wikidata, etc.)
11		Projects (ECPO, WOMAG, etc.)
12		Notes

## Network Data in ECPO

**Table 1.** Attribute data in the agent service. Note: For each additional name variant, name data fields 1 to 7 are added.

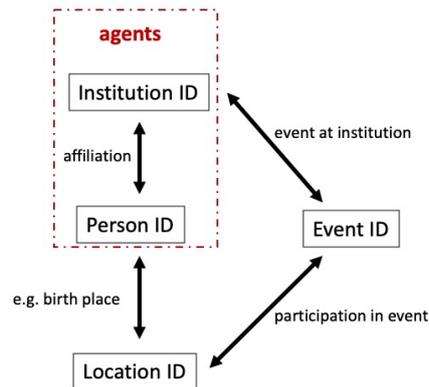


**Figure 3:** Event data structure.

The data structure thus maps both one-modal and bi-modal ties, namely direct links between people as well as links between people via institutions or events as intermediaries. This means, for example, that we would code that “Zhang > is the father of > Li” (direct tie), but *not* “Hu > participated in the same conference as > Gao.” Instead, this would be “Hu > participated in > conference XY” as well as “Gao > participated in > conference XY.” As the database grows, we can add participants to events or assign new events (such as publications) to people.

The one-modal and bi-modal ties have very different explanatory potentials. Direct social connections (one-modal), such as being the child or the Ph.D. student of someone, give us a reasonably good idea of how persons were socially connected to each other. It is much harder to make similar claims, however, merely based on congruent participation in a huge national event. Nonetheless, in analyzing the data, we can still infer networks by making assumptions about probabilities. When two people participated several times in events with a small number of participants or worked at the same institute for an extended period, it is certain that they maintained social relations, regardless

of whether their interaction was friendly, strictly professional, or antagonistic. Using the person's and event's base data, we can even try to study patterns such as possible interdependencies, like a birthplace or participation in high political bodies.



**Figure 4:** Links between data tables.

Apart from the agent service (which includes the persons and institutions or organizations) and the events table, we also created a location table. The China Historical Geographic Information System (CHGIS) laid the groundwork for historical spatial analysis by generating a reference database for place names. The CHGIS tracks changes in place names and administrative units from the third to the early twentieth century.<sup>9</sup> Unfortunately, no reference database exists for the period after 1911, although Christian Henriot and his team have already begun working on a Modern China Geospatial Database.<sup>10</sup>

The biographical data on Jiusan members was abstracted from archival documents in the People's Republic of China and contains information on the place of birth, date of party admission, institutional affiliations, past educational and professional experiences, and the place of residence. As many of the Jiusan members were known scientists or medical or technical experts, we can enrich

<sup>9</sup> *China Historical GIS*, 2001–2020, <http://chgis.fas.harvard.edu/>.

<sup>10</sup> ENP-Project, "Databases," *Elites, Networks and Power in Modern China (ENP China)*, 2020, <https://www.enpchina.eu/bases-de-donnees/>.

archival data with information abstracted from Chinese secondary sources. To trace the origin of information within the Jiusan dataset, we collect additional primary and secondary sources in a Zotero library, which links to the database entries.

## 6. Privacy Concerns

The archival membership lists of the Jiusan store a variety of personal information of everyday people. The extraction of such information poses questions regarding privacy concerns and the protection of individual data. Biographical information can also be found in published sources for some members, but it may be spread across online biographical sketches, party histories, and local biographical reference works. The aggregation of data from published sources in one database produces new knowledge, and the publication of this knowledge might already affect individuals. This is especially true if we cannot identify a person, for example if they have a very common name, as we also have no way of knowing if the person is still alive and consents to the publication of their personal data. Taking these sources and further analyzing them through statistical network tools allows us to map connections and draw conclusions that might have real-life implications on a personal or even political level for the individuals involved. Therefore, even though ECPO was initially designed as an open-access database, biographical data from Jiusan members is still not openly retrievable.

The solution could be to anonymize data for publication purposes by replacing the person's name, for example, with a numerical identifier; however, additional information such as institutional affiliations and birth dates would still make it possible to identify the individuals. As research concerning the anonymization or de-anonymization of graph data has shown, even inferring a one-mode network (where, for example, a tie between two anonymized persons expresses a joint affiliation) from the bimodal dataset would not be sufficient. To name only one exemplary work, Backstrom, Dwork, and Kleinberg demonstrated that even in a fully anonymized graph, as soon as one is able to identify the components of a sub-graph (one component of the network), the anonymization

of individuals unravels.<sup>11</sup> This means that even the publication of aggregated data abstracted from published sources might undermine the protection of more confidential information contained in archival sources. Therefore, the highly sensitive nature of the Jiusan dataset poses particular challenges to the integration of data within the rest of the ECPO dataset and the sharing and publication of data. We will nonetheless continue to make at least the biographical base data available, and to merge overlaps in the datasets wherever possible while protecting individual rights and interests.

## 7. Outlook

Despite these concerns, the expansion of the data structure and the agent service's implementation has significantly increased the analytical potential of the ECPO. With the addition of locations, events, and inter-agent relations, the agent service gained several new dimensions and now makes it easier to add agent-only data. Its central position within the ECPO infrastructure facilitates name control across database modules and systematic name management. To ensure our identification of agents is transparent and verifiable, we provide links to international authority records wherever possible. At the moment, we provide references to the German Integrated Authority File (GND),<sup>12</sup> OCLC's Virtual International Authority File (VIAF),<sup>13</sup> Wikidata,<sup>14</sup> DBpedia,<sup>15</sup> and the Chinese online encyclopedia *Baidu Baike* (百度百科).<sup>16</sup> Even though we do not import data from Wikidata or Baidu Baike (as these sites do not adhere to scientific standards), tracking the information on persons, organizations, and institutions

---

11 Lars Backstrom, Cynthia Dwork, and Jon Kleinberg, "Wherefore Art Thou R3579x?: Anonymized Social Networks, Hidden Patterns, and Structural Steganography," in *Proceedings of the 16th International Conference on World Wide Web - WWW '07* (ACM Press, 2007), 181–90.

12 The Integrated Authority File (*Gemeinsame Normdatei*, GND), is the central German authority service hosted by the German National Library, <https://www.dnb.de/EN/gnd>.

13 The Virtual International Authority File combines multiple national and international name authority files in a single OCLC-hosted name authority record service.

14 Wikidata is a collaboratively edited knowledge base hosted by the Wikimedia Foundation, <https://www.wikidata.org>.

15 DBpedia aims to allow users to semantically query structured information from Wikipedia resources, <http://dbpedia.org/>.

16 Baidu Baike is a Chinese-language encyclopaedia by the Chinese search engine Baidu; <http://baike.baidu.com/>.

## Network Data in ECPO

stored in these external datasets is nonetheless of value. It enables us to cross-check ECPO entries to assess their accuracy, or to probe which biographical information might be missing from the Baidu Baike entries because they are subject to government censorship. In other words, linking to external datasets opens new paths for exploring biases and gaps in popular and official historical narratives.

In the future, references to external authorities can certainly be expanded, especially towards more domain-specific services, for example to the China Biographical Database (CBDB, Harvard), the “Authorship of Chinese Women’s Periodicals” database (*Funü qikan zuozhe yanjiu pingtai* 婦女期刊作者研究平台) at the Academia Sinica, or the Modern China Biographical Database (MCBD, Lyon).<sup>17</sup> We also plan to develop an API to the agent service to provide machine-readable data for external reuse. We hope that the ECPO will continue to grow in the coming years and that more projects will contribute to the database, whether they are analyzed scans, “just” scans, or agent-related material.<sup>18</sup>

## 8. References

- Arnold, Matthias, and Lena Hessel. “Transforming Data Silos into Knowledge: Early Chinese Periodicals Online (ECPO).” In *E-Science-Tage 2019: Data to Knowledge*, edited by Vincent Heuveline, Fabian Gebhart, and Nina Mohammadianbisheh, 95–109. Heidelberg: heiBOOKS, 2020.
- Backstrom, Lars, Cynthia Dwork, and Jon Kleinberg. “Wherefore Art Thou R3579x?: Anonymized Social Networks, Hidden Patterns, and Structural Steganography.” In *Proceedings of the 16th International Conference on World Wide Web - WWW '07*, 181–90. ACM Press, 2007.
- China Historical GIS*, 2001–2020. <http://chgis.fas.harvard.edu/>.

---

17 Harvard University, “Introduction,” China Biographical Database Project (CBDB), 2020, <https://projects.iq.harvard.edu/cbdb/home>; Institute of Modern History, Academia Sinica, “The Introduction of the Database,” Authorship of Chinese Women’s Periodicals, 2018, <http://mhdb.mh.sinica.edu.tw/ACWP/about.php>; ENP-Project, “Databases.”

18 Templates of data structures for an import in ECPO are available upon request.

- Chinese Entertainment Newspapers*. Heidelberg University, 2006. <http://xiaobao.uni-hd.de>.
- Chinese Women's Magazines in the Late Qing and Early Republican Period (WoMag)*. Heidelberg University, 2009–2020. <https://uni-heidelberg.de/womag>.
- ENP-Project. "Databases." Elites, Networks and Power in Modern China (ENP China), 2020. <https://www.enpchina.eu/bases-de-donnees/>.
- Harvard University. "Introduction." China Biographical Database Project (CBDB), 2020. <https://projects.iq.harvard.edu/cbdb/home>.
- Hockx, Michel, Joan Judge, and Barbara Mittler, eds. *Women and the Periodical Press in China's Long Twentieth Century: A Space of Their Own?* Cambridge: Cambridge University Press, 2018. <https://doi.org/10.1017/9781108304085>.
- Institute of Modern History, Academia Sinica. "The Introduction of the Database." Authorship of Chinese Women's Periodicals, 2018. <http://mhdb.mh.sinica.edu.tw/ACWP/about.php>.
- Pasin, Michele, and John Bradley. "Factoid-Based Prosopography and Computer Ontologies: Towards an Integrated Approach." *Digital Scholarship in the Humanities* 30, no. 1 (April 2015): 86–97.
- Song, Doris, Liying Sun, and Matthias Arnold. "The Birth of a Database of Historical Periodicals: Chinese Women's Magazines in the Late Qing and Early Republican Period." *Tulsa Studies in Women's Literature* 33, no. 2 (2014): 227–37.
- Yeh, Catherine V. "Entertainment Newspapers and the Rise of Mass Culture." *Chinese Entertainment Newspapers*, 2006. <http://projects.zo.uni-heidelberg.de/xiaobao/index.php?p=intro>.