



## Name Authority Control in Digital Humanities: Building a Name Authority Database at Shanghai Library

Cuijuan Xia and Wei Liu

### Abstract:

Name authority control is an important tool for libraries. However, traditional library data are not built in the way that digital humanities research requires, which makes it difficult for digital humanities researchers to use them directly. This study is to address this problem through using the Linked Data approach to build knowledge bases in transforming and normalizing name authority data into the format that can be easily deployed by digital humanities research. A name authority database was built on various sources and formed the content infrastructure to provide Linked Open Data services, which enables sophisticated searches and uses of document resources knowledge base with multiple types of documents and multimedia, instead of digital collections with only a keyword search function. The process of design and development as well as the way through which resources are interlinked are described in detail in this paper.

### To cite this article:

Xia, C. & Liu, W. (2018). Name Authority Control in Digital Humanities: Building a Name Authority Database at Shanghai Library. *International Journal of Librarianship*, 3(1), 21-35. doi: <https://doi.org/10.23974/ijol.2018.vol3.1.68>

### To submit your article to this journal:

Go to <http://ojs.calaijol.org/index.php/ijol/about/submissions>

# **Name Authority Control in Digital Humanities: Building a Name Authority Database at Shanghai Library**

Cuijuan Xia and Wei Liu  
Shanghai Library, Shanghai, China

## **ABSTRACT**

Name authority control is an important tool for libraries. However, traditional library data are not built in the way that digital humanities research requires, which makes it difficult for digital humanities researchers to use them directly. This study is to address this problem through using the Linked Data approach to build knowledge bases in transforming and normalizing name authority data into the format that can be easily deployed by digital humanities research. A name authority database was built on various sources and formed the content infrastructure to provide Linked Open Data services, which enables sophisticated searches and uses of document resources knowledge base with multiple types of documents and multimedia, instead of digital collections with only a keyword search function. The process of design and development as well as the way through which resources are interlinked are described in detail in this paper.

**Keywords:** name authority, authority control, digital humanities, genealogy, visualization, linked data, web service

## **INTRODUCTION**

With the advent of digital humanities, humanities research has been through a “paradigm shift” (Kuhn, 1970). Human beings may be the most important research subject in history, humanity, and social studies. A person's date of birth, death, where the events happened, and other related information about the person such as employment, social activities, and intellectual works are the fundamental building blocks of society and history. For example, social network analysis—a research method in digital humanities—has been utilized and has revealed that cultural centers shifted from Rome to Paris, then to Los Angeles and New York, from B.C. 600 to 2012. This was

discovered through the spatiotemporal birth and death information of 150,000 notable individuals spanning that time period (Schich et al., 2014). This research relies on name authority databases to identify individuals.

However, the question of how to uniquely identify a person is not new. Name authority control is the effort to distinguish one person from another, even they have the same name. With the development of linked data and semantic web, a person is often treated as an entity and the name of this person and other related information are the attributes of the entity.

In digital humanities, it becomes even more important to maintain a name authority list to differentiate a person from another and aggregate related information about one person. Temporal and spatial information are two essential dimensions relating to a person. In addition, a person's biographical information, including awards received, achievements, employment, works, and social relationships can be a starting point for a humanities research.

Researchers at Shanghai Library aim to leverage traditional name authority files and explore expanding name authority into digital humanities by using Linked Open Data (LOD) to construct a new name authority database that amalgamates names from genealogy, rare book collections, archives, and other special collections of the Shanghai Library. The new name authority database is the fundamental block of the digital humanities infrastructure of Shanghai Library that provides services to researchers and ensures its openness, consistency, and high efficiency.

## LITERATURE REVIEW

Name authority control is one of the core functions of libraries. Library of Congress (LC) Name Authority File<sup>1</sup>, The German National Library (DNB)<sup>2</sup>, Norwegian National library<sup>3</sup>, and OCLC (Online Computer Library Center) Virtual International Authority File (VIAF)<sup>4</sup> are examples of name authority control. In particular, VIAF combines the name authority files from different sources into a single name authority service, and expands “the concept of universal bibliographic control by (1) allowing national and regional variations in authorized form to coexist; and (2) supporting needs for variations in preferred language, script and spelling” (OCLC, 2012, para. 5).

The name authority files (e.g., LC Authorities, DNB, and VIAF) contain a large number of individual persons and corporate bodies. More importantly, the “[l]ibrary data tends to be of very high quality, being collected, revised and maintained by trained professionals” (Hannemann & Kett, 2010, p. 2). The name authority files are valuable sources that offer a backbone for the Semantic Web (Neubert & Tochtermann, 2012).

---

<sup>1</sup> <http://id.loc.gov/download/>

<sup>2</sup> [http://www.dnb.de/EN/Standardisierung/GND/gnd\\_node.html](http://www.dnb.de/EN/Standardisierung/GND/gnd_node.html)

<sup>3</sup> <http://data.bibsys.no/autreg/index.html>

<sup>4</sup> <https://www.oclc.org/en/viaf.html>

In digital humanities, many scholarly projects (e.g., FIHRIST<sup>5</sup> -- Union Catalogue of Manuscripts from the Islamicate World) have relied on VIAF for authority control (Smith-Yoshimura, & Michelson, 2013). For libraries, the main purposes of name authority control are to facilitate information organization and retrieval. Humanities scholars have different views on “preferred name” that could “exclude some historical or cultural perspectives” and furthermore they need to “know the provenance of each form of name” (Smith-Yoshimura, & Michelson, 2013, paras. 7–8). Moreover, digital humanities scholars have requirements for the types of information they compile and organize. The requirements are often associated with scholars' research interests. They collect more granular information than what is in name authority files. For example, the China Biographical Database (CBDB) Project contains not only names but other biographical information, including social associations, career data and kin relations (Harvard University, Academia Sinica, & Peking University, 2018).

Regarding content organization, name authority files focus on one person with different names and establish “see also” relations and the relations between different names and corresponding works. Events, employment, experiences are very sketchy or missing in name authority files. For example, in the LC Name Authority File, one person's information only consists of a formal name, alternative names, date of birth, date of death, and works. In digital humanities databases, the coverage of one person is broader. For example, in the Database of Names of Biographies<sup>6</sup> created by the Institute of History and Philology, Academia Sinica (“台湾中研院历史语言研究所”), one person's information consists of two parts: 1) basic information about a person including name, alternative names with provenance information, date of birth, date of death, place of birth, expertise, and other fields, and 2) detailed biographical information including employment, key events related to the person and descriptive information from other biographical sources<sup>7</sup>.

The way of establishing a name authority file is incremental. Cataloguers add entries into the name authority files following a set of rules. In digital humanities, the databases are often created and maintained by researchers in the fields. The databases are more specific.

Technology-wise, early name authority files were coded in MACHine-Readable Cataloging (MARC) format and highly coupled with integrated library systems in libraries. With the development of linked open data, national libraries from various countries started to publish name authority files as linked open data. The approach is similar to mapping authority file entries (MARC fields) into the Simple Knowledge Organization System (SKOS) predicates and stores in Resource Description Framework (RDF) statements (Papadakis, Kyprianos, & Stefanidakis, 2015). Each person's information is encoded in RDF statements that are more suitable for publishing on the Internet. In this way, each person has a unique Hypertext Transfer Protocol (HTTP) Universal Resource Identifier (URI) so that it can dereference on the Internet. For example, <http://viaf.org/viaf/71426854> is the HTTP URI for a notable Chinese scholar named Hu Shi (“胡

---

<sup>5</sup> <https://www.fihrist.org.uk/about>

<sup>6</sup> [http://archive.ihp.sinica.edu.tw/ttsweb/html\\_name/index.php](http://archive.ihp.sinica.edu.tw/ttsweb/html_name/index.php)

<sup>7</sup> [http://archive.ihp.sinica.edu.tw/ttsweb/html\\_name/build.php](http://archive.ihp.sinica.edu.tw/ttsweb/html_name/build.php)

适”). In comparison, the databases in digital humanities often store information in relational databases. For example, the China Biographical Database uses Microsoft Access.

With libraries' deeper participation in digital humanities, name authority files in libraries are not enough to support the research requirements of digital humanities. Thus, the convergence of name authority control and digital humanities is inevitable and can benefit both. Meanwhile, the need to build cyberinfrastructure (American Council of Learned Societies, 2006) has a significant impact on libraries and digital humanities. Name authority files such as VIAF can be the fundamental blocks that support cyberinfrastructure for digital humanities (Smith-Yoshimura, & Michelson, 2013).

The aim of the name authority database at Shanghai Library is to further explore how to integrate name authority files with digital humanities databases and take advantage of new models such as the International Federation of Library Association (IFLA) library reference model (Riva, Bœuf, & Žumer, 2017) and BIBFRAME (Bibliographic Framework) 2.0 (Library of Congress, 2016) for modelling data in the name authority database. In these models, a person is defined as an entity that is an individual human being and identified by a unique URI. The person, or entity, is restricted to real individuals who live or are assumed to have lived. The different names of one person are attributed to values of the entity, along with other attributes such as birth and death dates, place of birth and death, expertise and so on. The person entity is associated with a Work or Instance through roles such as author, editor, artist, photographer, composer, and illustrator.

## **BUILDING A NAME AUTHORITY DATABASE AT SHANGHAI LIBRARY**

The name authority database at Shanghai Library can be enriched and augmented with the genealogy databases, manuscripts, archives, and rare books collections of Shanghai Library. Names included in the name authority database are mainly collected from bibliographical databases of Shanghai Library including names of authors and some historically and culturally important people related to the authors. Supplementary data is from Wikipedia, name list, and the CBDB open database.

The name authority database is built on Linked Open Data. Linked data is a method and technique to publish data on the semantic web and is related to a knowledge graph. In Linked Data, the name of an entity (such as a real person) can be expressed by a string, and the name is just an attribute of a real-world object. Properties are the conceptualization of relations between a concept and features of the concept. HTTP URIs are used as unique identifiers in Linked Data. One of the functions of HTTP URIs is to uniquely identify an object that can be utilized in name authority control for the identification of individuals.

In reality, a name is not a unique property for an object but can be distinct in combination

with other properties of the object. For example, a person has a name, date of birth and death, birthplace, kinship relationships, social relationships, etc. All that information can uniquely identify a person. In Linked Data, information is expressed in a set of RDF triples that provide support to the uniqueness of an object.

An HTTP URI can be referenced in the semantic web and linked to a structured description of the referent of the HTTP URI. This feature can be utilized by the name authority database that provides data services through the Internet. As part of the infrastructure of the digital humanities platform, the name authority database provides not only basic information on a person, but key events, social relationships, and works that can be displayed, searched, aggregated and analyzed.

### *Data Model*

The main task of data modeling in the construction of the name authority database is to conceptualize a “person” and the attributes of the concept, such as date of birth and death, place of birth, nationality, and relationships between related concepts, including relationships between one person and other persons, a person and other related entities such as documentations, organizations, places, time, events, and objects. Person, Place, Time, Event, Physical Objects and Materials and their relations are conceptualized in a model denoted in Figure 1. In this model, Person, Place, Time, Event, Physical Objects and Materials are concepts drawn for material resources that exist in the real world. The concepts connect with each other through attributes and relations.

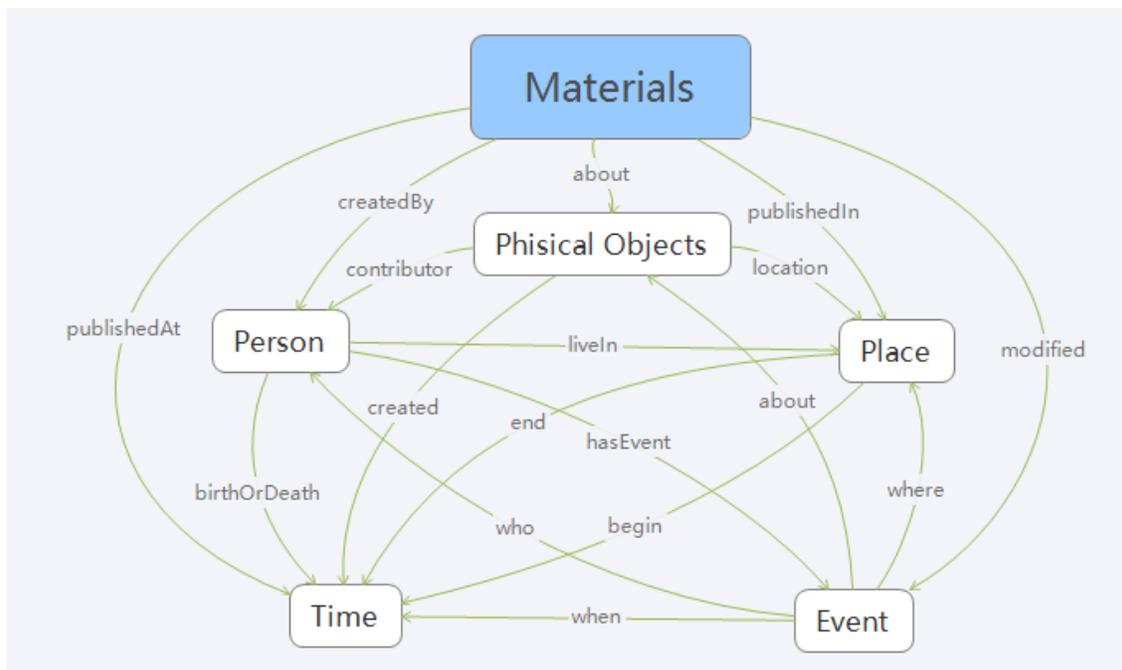


Figure 1. An abstract data model of a name authority (Xia, 2017, p. 51)

With the model, concepts, attributes, and relations can be formalized, which is the process of developing an ontology. Figure 2 shows the person ontology model of the name authority in

Shanghai Library. In Figure 2, each concept is denoted by a rounded rectangle. Each relation is illustrated by a directed arrow. The design of a person ontology needs to address multiple names of a person, relations between one person and others, important events of a person such as birth, death, marriage, and employment.

Regarding multiple names of a person, foaf:name is for formal names and the shl:Name class is for alternative names. In addition, types of names are denoted by a property (shl:nameType) of shl:Name. The class of shl:Name is repeatable to include all names used by a person. Properties of the class are for describing things related to a specific name. For example, the pseudonym name Xiaofan (“小凡”) of a notable Chinese writer MaoDun (“茅盾”) was used the first time in Shen bao yue kan (“《申报月刊》”) in 1934. The class shl:Name is for the pseudonym name Xiaofan (“小凡”) and foaf:Name is for the predominant name MaoDun.

The relationships between one person and another can be kinship and social relationships. In CBDB, there are more than 100 descriptions for relationships between two individuals. It is straightforward to define a property for a relationship. The abstract data model of name authority was formalized into a person ontology. In the ontology, a Relationship class is developed and different relationships are distinguished by a property relationType. For important relationships, relationships in the RELATIONSHIP vocabulary<sup>8</sup> are re-used, such as friend (rel:friendOf), spouse (rel:spouseOf), and child/parent (rel:childOf) relationships.

The Event class is for important events that can happen to a human being and has time, place and other properties to describe general events. Regarding special events, sub-class can be used to extend the Event class. For example, an employment event can be named as OfficialEvent as a sub-class of Event, the two properties of PROV Ontology<sup>9</sup> (i.e., prov:startedAtTime and prov:endedAtTime) are reused with a self-defined property officialPosition for describing the position of an employment event. Furthermore, the positions are from a controlled vocabulary.

---

<sup>8</sup> <http://vocab.org/relationship/>

<sup>9</sup> <https://www.w3.org/TR/2013/REC-prov-o-20130430/>

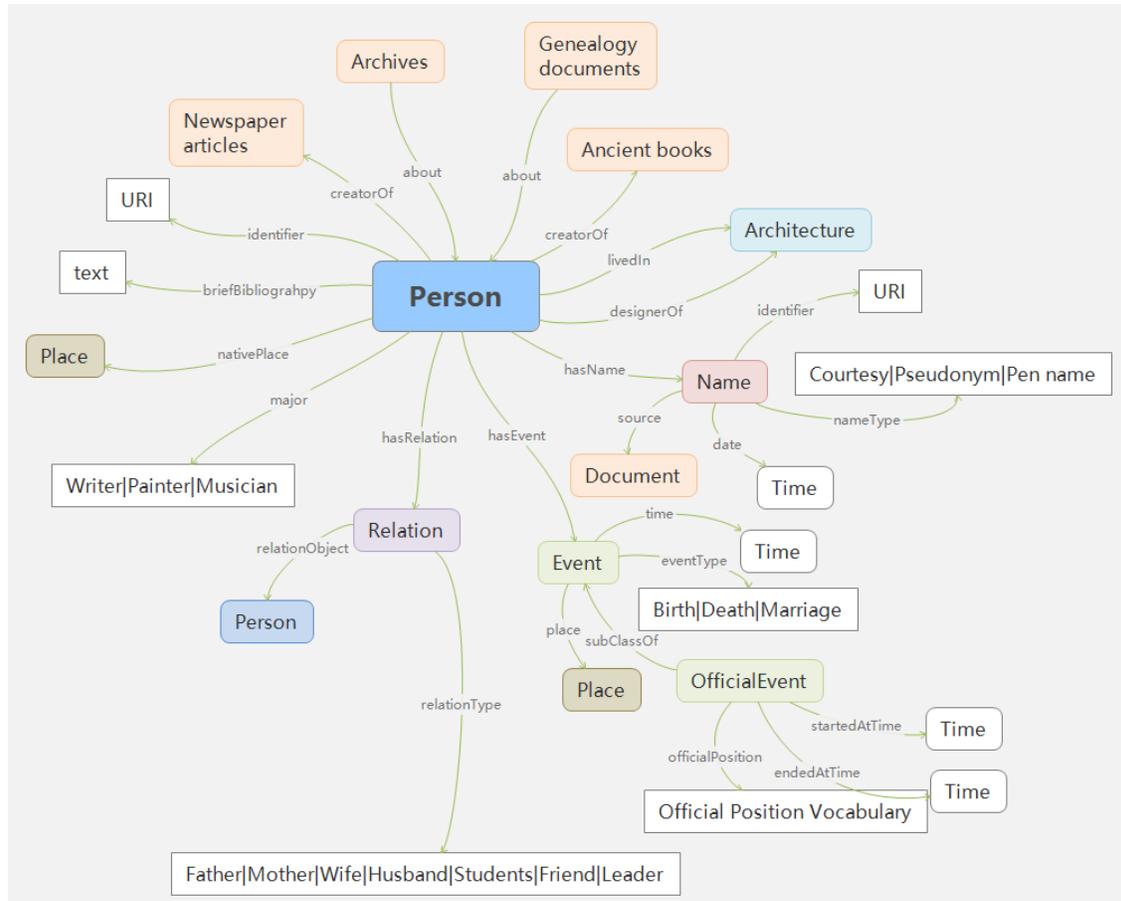


Figure 2. The Name authority ontology model

### *Data collection, clean, and augment*

The aim of data collection is to gather metadata about authors (often including name and dynasty) and to enrich information on a person based on the structure of the person ontology. Metadata about authors are the core of the name authority database at Shanghai Library. The metadata consists of an author's name and time period (e.g., dynasty), other brief information as well as some other basic information about the authors. Unfortunately, metadata are not sufficient for name authority control in digital humanities. More data needs to be retrieved from the Celebrity encyclopedia (“人物百科”), name dictionary (“人名辞典”), and the professionals and experts database (“专业人物数据库”) to supplement the metadata.

The primary goal of data cleansing is to integrate data from different sources to distinguish different persons with the same name and merge the same person with different names. A traditional approach used in the library relies on a human being through manual processes to augment entries of name authority databases. In the age of big data, this type of process cannot solely rely on human involvement. Hybrid approaches can be used to speed up the process through leveraging automatic tools and knowledge bases and validation by a human being.

For example, the author of a rare book titled “Yu yang shan ren jing hua lu” (“渔洋山人精华录”) has two entries “[Qing Dynasty] Wang shizhen” (“【清】王士禛”) and “[Qing Dynasty]

Wang shizhen” (“【清】王士禛”). Detailed information retrieved for the name authority database in Shanghai Library shows the following:

Original text:

“王士禛（1634年9月17日—1711年6月26日），原名王士禛，字子真，一字贻上，号阮亭，又号渔洋山人，世称王渔洋，谥文简。”

Translation:

“Wang shizhen (“王士禛”)(1634-9-15 – 1711-6-26), original name is “Wang shizhen” (“王士禛”), courtesy name is 子真, another courtesy name is 贻上, pseudonym is 阮亭, another pseudonym is 渔洋山人, known for 王渔洋, and posthumous name is 文简. ”

It is straightforward for a human to determine that these are the two names of the same person. If such a task is performed automatically, one of the approaches is to segment the description into smaller units and to extract useful information such as works, birth place, date of birth and death, original name, courtesy name (“字”), pseudonym (“号”), etc. Then an algorithm is used to compare how close the two names are related. In this example, the two names are associated with the same work, “Yu yang shan ren jing hua lu” (“渔洋山人精华录”). Also, the date of birth and death (i.e., September 17, 1634, and June 26, 1711) fall into the same period of the Qing dynasty. Therefore, it is safe to determine that they are the two names of the same person. A data cleansing application developed by Shanghai Library can perform such a task. If the application cannot determine whether the two names are of the same person, it lists the two names that can be validated by a human.

The goal of data merge is to augment the same individual's information from different sources. Regarding one property of a person, there may be multiple values (e.g., one value from each source). The values may be the different. The principal in the name authority database is to keep a unique value such as date of birth and death, birthplace, and gender if possible. In the process, the authority and reliability of each source are examined, then the best data available is picked. Properties often have multiple values such as name (e.g., ancient Chinese often has multiple names), employment and important activities that are kept with links to the sources.

Take “Wang shizhen” for example, after the data cleansing process, the two names “【清】王士禛” and “【清】王士禛” for the name authority database are the names of the same person. The person is uniquely identified by one HTTP URI, and the names are included in this person's name authority entry. “Wang shizhen” (“王士禛”) is the formal name denoted by foaf:name, “Wang shizhen” (“王士禛”) is a type of "original name" denoted by shl:Name, “Yu yang shanren” (“渔洋山人”) is denoted by the type of “pseudonym (“号”)” using shl:Name, “zi zhen(“子真”) is a type of courtesy name (“字”) using shl:Name. The property of brief biography can accommodate data from different sources such as data from Shanghai Library and the National Library of China.

### Name authority services

The name authority database based on linked data is a part of the knowledge base of the digital humanity platform at Shanghai Library. The main purpose of this knowledge base is to provide services to other knowledge bases over the Internet through content negotiation, a SPARQL (SPARQL Protocol and RDF Query Language) Endpoint, Restful Application Programming Interfaces (API), and development kits.

Content negotiation is to respond to clients' requests based on the type of request (e.g., HTML/text or RDF/XML) that are sent from the client side. In the name authority database, each has a unique HTTP URI. In the document knowledge base of the digital humanities platform at Shanghai Library, the authors or persons in the content can be denoted by HTTP URIs for the person. Not only the alternative names, but also different pseudo names of the same person, can link to the same HTTP URI. Therefore, it becomes possible to access all of the works of the person from the HTTP URI of this person. Meanwhile, the work itself is an HTTP URI entity. The relation between works and authors are relations between entities, not the relation between strings. In this way, different works link to the author via HTTP URI and the author can link to different works through a HTTP URI as well.

Figure 3 shows that from the HTTP URI of a person, all works of this person can be discovered regardless of the type of the document.

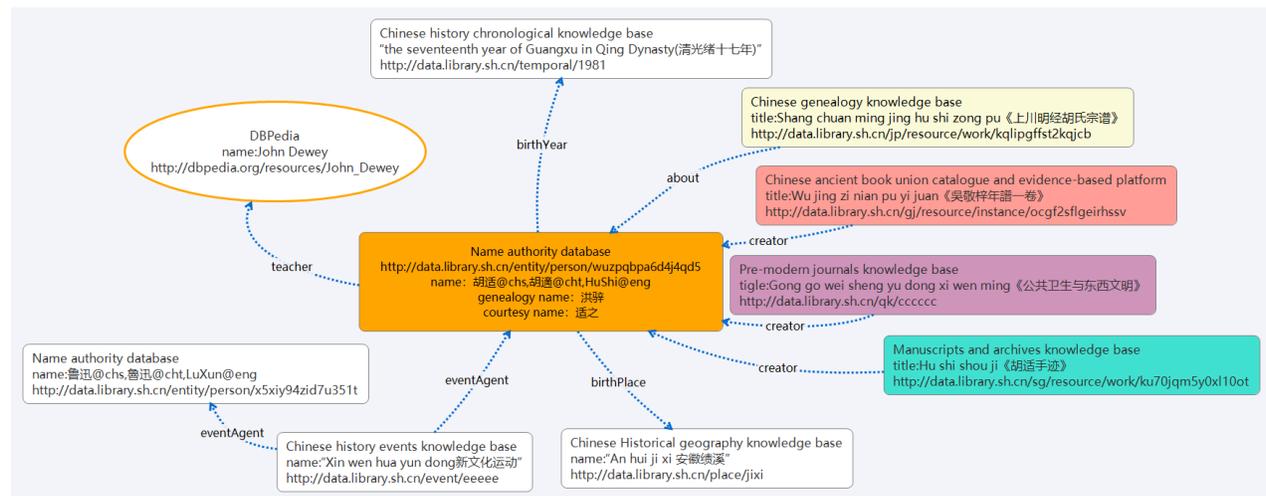


Figure 3. Relations among a scholar (i.e., Hu Shi), place, time, event, object and work

The data interface is openly available to provide detailed information about individuals, which includes Restful APIs and a SPARQL Endpoint. This interface is built on a light-weight web service framework that supports HTTP protocol. API calls often consist of various parameters and form into URLs. Such interface does not rely on specific systems or platforms as long as systems or platforms support the technology infrastructure of the Internet. This interface could serve not only Shanghai Library but also other libraries, the third-party organizations, or individual researchers. Name authority control can be realized in applications by using this data interface. Taking "Ba jin" ("巴金") as an example, the following URL is formatted as one Restful API with

a parameter “Ba jin” (“巴金”):

<http://data1.library.sh.cn/sg/persons/巴金>

This request can return all information about “Ba jin” (“巴金”) in a JavaScript Object Notation for the Linked Data (JSON-LD) format.

```
{
  //HTTP URI of Ba Jin
  "@id": "http://data.library.sh.cn/entity/person/6pnc6xdap3cfkcup",
  "@type": "http://www.library.sh.cn/ontology/Person",
  "identifier": "A000001",
  "friendOf": [ // friends of Ba jin
    "http://data.library.sh.cn/entity/person/ej89g3k7m1eyvfot",
    "http://data.library.sh.cn/entity/person/khmt77c5kjeutwa",
    ... ],
  "birthday": "1904",
  "briefBiography": "著名作家,文学翻译家,曾任.....",
  "createdWork": "《第四病室》; 《随想录》",
  "ethnicity": "http://data.library.sh.cn/entity/ethnicity/han",
  "name": [ //all types of names of Bajin
    "http://data.library.sh.cn/entity/nameOther/0bz4zhl13kxmsczq",
    "http://data.library.sh.cn/entity/nameOther/fd3skfeilxepcvd2",
    "http://data.library.sh.cn/entity/nameOther/442de7v00aqyrvmi",
    ... ],
  "nationality": "http://data.library.sh.cn/entity/nationality/china",
  "nativePlace": "http://data.library.sh.cn/entity/place/n9gd2vjf73f59ve6",
  "officialEvent": [ //employment of Ba jin
    "http://data.library.sh.cn/entity/officialPosition/wg5a3cbayol114mk",
    "http://data.library.sh.cn/entity/officialPosition/otnroqrhove20qvm",
    ... ],
  "speciality": "作家",
  "gender": "男",
  "http://xmlns.com/foaf/0.1/name": [ {
    "@language": "cht",
    "@value": "巴金" },
  ],
  "@context": { //namespace of every property and class used in JSON-LD
  }
}
```

## DISCUSSION

A typical entry in the name authority database consists of date of birth and death, employment history, and all sorts of relations such as the kinship or social relations of a person. All of this data can be referenced or used as facet, classification, and personalized recommendation functions in applications. In addition, the data can be leveraged for distinguishing two types of situations (i.e., the same name but for different people, the same person with different names).

### *Improve precision and recall*

Without name authority control, search engines can only use string matching algorithms to find results. With regards to searching by names, if the input is "Mao Dun" (“茅盾”), the results can only consist of records containing the string "Mao Dun" (“茅盾”). If the input is "Shen Yanbing" (“沈雁冰”), the results can only consist of records containing the string "Shen Yanbing" (“沈雁冰”). With name authority control, if the input is "Mao Dun" (“茅盾”), the system can combine "Mao Dun" (“茅盾”) with "Shen Yanbing" (“沈雁冰”) and return results containing the two names since they are the two names of the same person.

In the name authority database at Shanghai Library, an exact search function supporting name authority control has been implemented. The search function is based on concept matching instead of string matching. The system provides SPARQL Endpoint data service supporting a search by names. The SPARQL Endpoint data service searches persons' names first and then returns corresponding HTTP URIs. Through name HTTP URIs, other related entities such as works can be retrieved accordingly. For example, regardless of whether users input "Mao Dun" (“茅盾”) or “Shen Yanbing” (“沈雁冰”) or “Shen Dehong” (“沈德鸿”) (the three different names of the same person), there is a unique HTTP URI of the person entity that is returned. Through the HTTP URI, the bibliographical database returns all of the works for which Mao dun (“茅盾”) is responsible. This way, both precision and recall are improved.

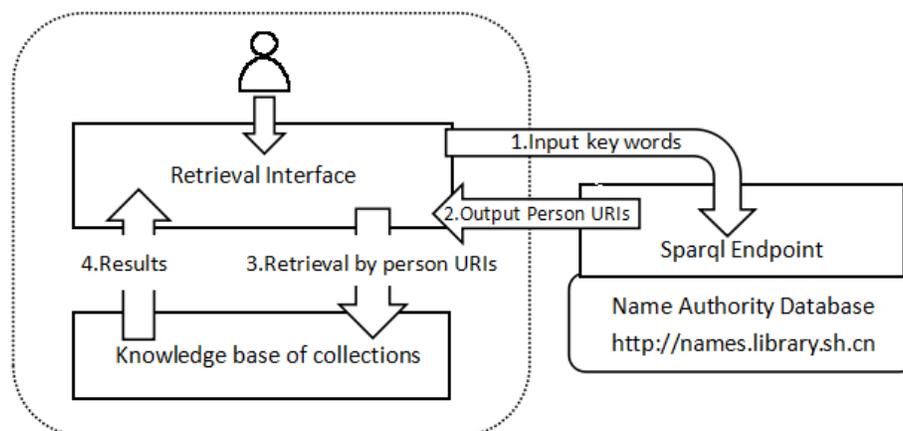


Figure 4. Searches based on name authority control

### Dynamic display of relations between visualization between entities

In recent years, researchers at Shanghai Library started to experiment with shifting the focus on content from work and reorganized metadata to the full, digitized content of historical archives. Effort was made to build internal relations based on the content of historical archives, and further link people, places, times, events, and physical objects in the digital humanities platform.

Based on the abstract model in Figure 1, relations can be constructed between the resource databases of libraries. Since all works are identified by HTTP URIs, this can break silos within and beyond the libraries that own the works. A person can be an author who created works, a creator who created personal archives, notes and photos, and an actor to act in films. All of these things created by individuals have nothing to do with the organization that collected them or how many organizations collected them. From the entity of a person, all things related to that person can be aggregated. This can further reveal people who are related to the creator, events that the person was involved in, and how they are related to time and space.

The genealogy database and “Sheng Xuanhuai” (“盛宣怀”) archives<sup>10</sup> of Shanghai Library are linked together through HTTP URIs in the name authority database. Figure 5 shows that when users visited the “Sheng Xuanhuai” (“盛宣怀”) archives, the system returned an HTML page containing all persons who had corresponded with him at a specific time period, which was visualized in directed graphs. At the same time, when visiting his HTTP URI, the system would search in the genealogy database and related genealogy documents.



Figure 5. Connections between genealogy and archives through relationships of Sheng Xuanhuai

<sup>10</sup> <http://sd.library.sh.cn/sd/home/index>

In the “Sheng Xuanhuai” (“盛宣怀”) archives, social relationships can be revealed through the linkages between archival documents. There are more than 110,000 correspondences between “Sheng Xuanhuai” (“盛宣怀”) and other important persons at that time, including politicians and businessmen. The number of individuals corresponding, when they corresponded, and where the correspondences happened all reveal certain contexts of the correspondence. Figure 6 shows the corresponding relations of “Sheng Xuanhuai” (“盛宣怀”) with others. Each dot represents one person and this person’s detailed information can be found when clicking on the dot. The directed line between two person means they corresponded. The system displays the number of times there were correspondences when a mouse is hovered over a directed line. Figure 6 contains a full view of the corresponding relationships and shows that “Sheng Xuanhuai” (“盛宣怀”) –the left orange donut - communicated with “Sheng Kang” (“盛康”) - the right orange donut - more than any other individual.

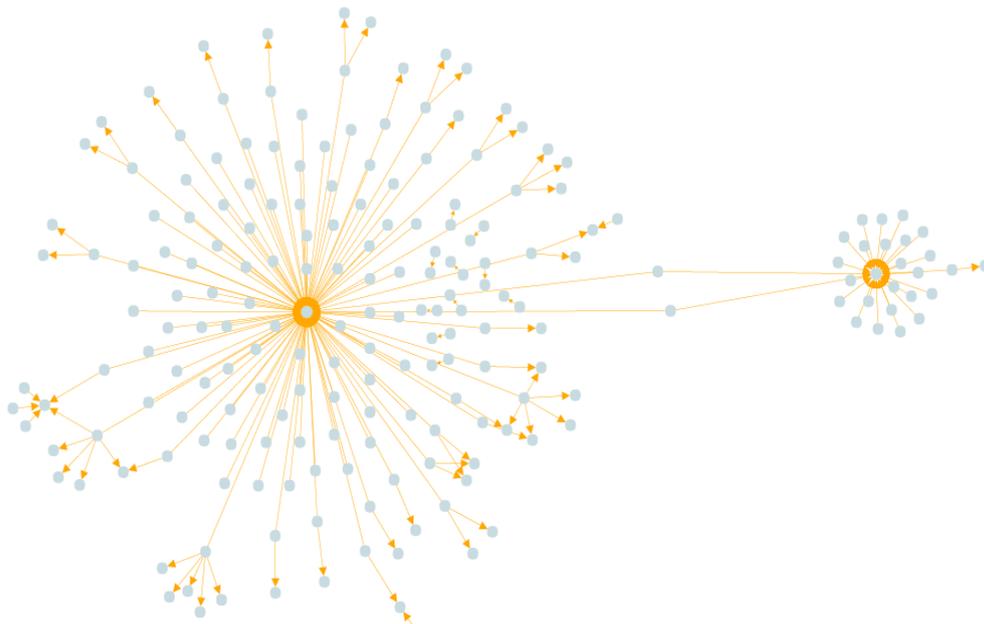


Figure 6. A full communication map from the Sheng Xuanhuai archives

In the platform, visualization tools directly help people visualize a related relationship.

## CONCLUSION

The name authority database in Shanghai Library has advantages over traditional library name authority files and biographical databases in the digital humanities. It leverages ontology and linked data technologies and establishes a linked open data platform that provides linked open data services for name authority control. It has used FOAF, Schema.org, GeoNames schema and integrated name authority files from the Library, Celebrity encyclopedia (人物百科), name dictionary (人名辞典), professional and domain-specific database of celebrities (专业人物数

据库) , and other data sources. As a result, it consists of 840,000 individuals' information. It provides services for genealogy, manuscripts and rare books collections in Shanghai Library. Meanwhile, it provides HTTP URIs, restful APIs, and a SPARQL Endpoint to serve other libraries and researchers.

However, the name authority database in Shanghai Library needs to be augmented with more data. For example, the place of birth information can only be found in 150,000 entries. The date of birth and death information can only be found in 250,000 entries. Name duplication is another problem that needs to be addressed. For example, there are about 70,000 entries that need to be cleared and merged. A person's ontology needs to be extended and further adjusted to support different services such as the integration of historical geographical information systems.

## ACKNOWLEDGEMENT

Thanks Qing Zou for providing valuable guidance during the writing of the paper. He provided insights for the innovative meaning of the paper and worked a lot to translate Chinese into English.

## References

- American Council of Learned Societies. (2006). *Our cultural Commonwealth: The report of the American Council of Learned Societies Commission on Cyberinfrastructure for the Humanities and Social Sciences*. Retrieved from [http://www.acls.org/uploadedFiles/Publications/Programs/Our\\_Cultural\\_Commonwealth.pdf](http://www.acls.org/uploadedFiles/Publications/Programs/Our_Cultural_Commonwealth.pdf)
- Hannemann, J., & Kett, J. (2010). World Library and Information Congress: 79th IFLA General Conference and Assembly. In *Proceedings of the world library and information congress* (pp. 1–11). Gothenburge, Sweden.
- Harvard University, Academia Sinica, & Peking University. (2018). Chinese Biographical Database. Retrieved April 19, 2018, from <https://projects.iq.harvard.edu/cbdb/how-cbdb-collects-data-and-why-you-do-not-always-find-what-you-want>
- Kuhn, T. S. (1970). *The structure of scientific revolutions*. Chicago: University of Chicago Press.
- Library of Congress. (2016). Overview of the BIBFRAME 2.0 Model [webpage]. Retrieved April 11, 2018, from <http://www.loc.gov/bibframe/docs/bibframe2-model.html>
- Neubert, J., & Tochtermann, K. (2012). Linked Library Data: Offering a Backbone for the Semantic Web. In D. Lukose, A. R. Ahmad, & A. Suliman (Eds.), *Knowledge Technology* (Vol. 295, pp. 37–45). Berlin, Heidelberg: Springer Berlin Heidelberg. [https://doi.org/10.1007/978-3-642-32826-8\\_4](https://doi.org/10.1007/978-3-642-32826-8_4)
- OCLC. (2012). VIAF (The Virtual International Authority File). Retrieved April 11, 2018, from <https://www.oclc.org/research/activities/viaf.html>
- Papadakis, I., Kyprianos, K., & Stefanidakis, M. (2015). Linked Data URIs and Libraries: The Story So Far. *D-Lib Magazine*, 21(5/6). <https://doi.org/10.1045/may2015-papadakis>

- Riva, P., Bœuf, P. L., & Žumer, M. (2017). IFLA Library Reference Model A Conceptual Model for Bibliographic Information. *International Federation of Library Associations and Institutions*, 101.
- Schich, M., Song, C., Ahn, Y.-Y., Mirsky, A., Martino, M., Barabasi, A.-L., & Helbing, D. (2014). A network framework of cultural history. *Science*, 345(6196), 558–562. <https://doi.org/10.1126/science.1240064>
- Smith-Yoshimura, K., & Michelson, D. (2013, March 27). Irreconcilable differences? Name authority control & humanities scholarship – Hanging Together. Retrieved April 19, 2018, from <http://hangingtogether.org/?p=2621>
- Xia C. (2017). Building a Digital Humanities Platform by Using Linked Open Data Services. *Journal of Library and Information Science*, 43(1), 47–70.

---

### About the authors

Cuijuan Xia is a senior engineer of Shanghai Library System and Network Center and the director of the Digital Humanities projects. She is expert in humanities data modeling, ontology design and system architecture design. She has more than 40 papers published.

Wei Liu (aka. Keven Liu) is the Deputy Director of Shanghai Library and Institute of Scientific and Technological Information of Shanghai. He is an adjunct professor of Fudan University and Shanghai University in Shanghai, China. He took part in many major digital library projects in China since 1995. He is very active in developing digital humanities infrastructures in recent years.