Postprint: Construction of a Linked Data Model for Research Entity Name Standardization

Authors: Zhou Yi, Zhang Jianyong, Liu Zheng, Liu Xiumin

Date: 2023-04-01T00:00:00+00:00

Abstract

[Purpose/Significance] This study aims to investigate the data model for linked data—a key challenge in publishing the research entity name authority data of the National Science and Technology Library (NSTL) as linked data. Research on the data model for research entity name authority data facilitates the sharing, interconnection, and quality improvement of NSTL's research entity data, as well as its integration into the Internet, while also providing a model reference for other institutions to use and publish linked data. [Method/Process] First, we analyze and compare the data models adopted in domestic and international linked data publishing projects, finding that data models in such projects mainly fall into two categories: those centered around Schema.org and those combining multiple standard vocabularies. Based on the characteristics of NSTL's name authority data, we design two forms of linked data models and compare them in terms of their expressive completeness for name authority data and model complexity to select the superior solution. Finally, we conduct an experiment using D2RQ as a tool to publish sample NSTL name authority data as linked data. [Results/Conclusion] The analysis reveals that between the two schemes, the one with Schema.org as the core vocabulary demonstrates superior expressive completeness, lower model complexity, and easier integration into the Internet compared to the scheme combining multiple standard vocabularies, thus making it more suitable as the linked data model for NSTL name authority data.

Full Text

Construction of a Linked Data Model for Research Entity Name Authority Data

Zhou Yi, Zhang Jianyong, Liu Zheng, Liu Xiumin National Science Library, Chinese Academy of Sciences, Beijing 100190

Abstract

[**Purpose/Significance**] This study investigates the key challenge in publishing the National Science and Technology Library's (NSTL) research entity name authority data as linked data: the linked data model itself. Research into the data model for research entity name authority data facilitates the sharing, interconnection, and quality improvement of NSTL's research entity data, enabling its integration into the broader internet while providing a reference model for other institutions seeking to use and publish linked data. [**Method/Process**] First, we analyzed and compared data models employed in linked data publishing projects both domestically and internationally, identifying two primary categories: those centered on Schema.org and those combining multiple standard vocabularies. Based on the characteristics of NSTL's name authority data, we designed two alternative linked data models and compared them in terms of their expressiveness for name authority data and model complexity to select the superior option. Finally, we conducted an experiment using D2RQ to publish sample NSTL name authority data as linked data. [**Result/Conclusion**] Our analysis revealed that the Schema.org-centered vocabulary scheme offers superior expressiveness and lower model complexity compared to the multiple vocabularies combination scheme, making it easier to integrate with the internet and thus more suitable as the linked data model for NSTL's name authority data.

Keywords: research entities; name authority; linked data; data model

---

Research entities constitute the fundamental components of scientific research activities, primarily including research subjects (researchers, institutions, etc.), research activities, research conditions (research methods, scientific literature, journals, etc.), and research outputs (research results, etc.) [1]. The essence of authority control is to achieve concept-based description and matching [2], with the goal of collocating different name forms for the same entity and distinguishing different entities with identical names to achieve semantic disambiguation. Information on research entities such as persons, institutions, publications, and funding plays a crucial role in information services. NSTL has standardized research entity information from its various systems and publisher data, establishing the NSTL Name Authority System covering four types of name authority data: institutions, persons, funding, and journals.

NSTL's name authority data currently follows the *Name Authority Metadata Standard*, designed with principles such as modularity and minimal granularity to enable detailed description of authority objects. However, it uses individual records as descriptive units without forming a description framework or ontology based on real-world entities, thus failing to support dataset reuse and open linking. Additionally, NSTL's name authority data is extracted from bibliographic data, resulting in limited descriptive information and sparse data that requires integration with other linked datasets to improve quality.

Therefore, to fully realize the application value of authority data and improve data quality, it is necessary to study how to transform NSTL's name authority data model into a linked data model and publish it as linked data. As a national resource guarantee base, service integration hub, and service development support center for scientific and technological literature information, publishing NSTL's name authority data as linked data holds significant importance: (1) it shifts data description toward real-world entities and their relationships; (2) it enables reuse by other NSTL systems or third-party institutions; and (3) it facilitates integration with other linked datasets to improve data completeness and quality.

Building a data model is a critical step in the linked data publishing process. How to make data more comprehensible to machines and more easily linked with other linked data depends on a scientific and sustainable data model. Drawing on experiences from linked authority data publishing projects, this paper aims to construct a linked data model suitable for NSTL's name authority data. By transforming NSTL's name authority data model into a linked data model, we achieve linked data publication.

---

2. NSTL's Research Entity Name Authority Data

NSTL aggregates data from its processing systems, historical document databases, researcher information platforms like VIVO, and bibliographic data from nine publishers including WoS and Wiley. From these sources, it extracts research entities such as institutions, persons, funding, and journals to build the NSTL Name Authority System. The system comprises four types of name authority data: over 30,000 journal records, more than 17 million person records, over 3.5 million institution records, and over 3 million funding records.

Based on the *Name Authority Metadata Standard*, NSTL's name authority data is extracted from bibliographic metadata. Table 1 shows the number of elements and attributes describing various object types. For persons, there are elements and attributes for names, contact information, birth/death dates, and affiliations; for institutions: names, contact details, addresses, and historical information; for funding: names, dates, subjects, and descriptions; and for journals: titles, identifiers, publication dates, and issue information. These attributes, distinct from ontology properties, constitute components of elements that define element characteristics. An authority record is essentially a set of records about a single authority object, potentially containing multiple records from different data sources, with one preferred name and others as variant names.

Although the *Name Authority Metadata Standard* describes authority objects in detail, it does not form an entity-based description framework or ontology, failing to meet linked data publishing requirements. Transforming NSTL's name authority data model into a linked data model requires a model that can fully express existing data while maintaining interoperability with other linked datasets.

In building the linked data model, we can learn from domestic and international linked data publishing projects.

---

3. Data Models in Linked Data Publishing Projects

3.1 Overview of Data Models in Linked Data Publishing Projects

OCLC's Virtual International Authority File (VIAF) merges multiple name authority files into a single service, aiming to provide authority information online through matching and linking. VIAF currently includes data from over 40 organizations across more than 30 countries [3]. Its data model has evolved multiple times, gradually shifting focus from concepts and names to entities. Initially treating authority objects as concepts, VIAF later distinguished different entity types including places, organizations, persons, and works. Early versions primarily used SKOS for description, later adding FOAF and RDA. In 2014, VIAF adopted Schema.org as its core vocabulary [5]. Additionally, in 2012, OCLC published WorldCat as linked data using Schema.org and its library extension (SchemaBibEx) [6].

Since 2010, the German National Library has experimented with publishing its authority data as linked data [7]. The Integrated Authority File (GND) consolidates the Corporate Body Authority File (GKD), Name Authority File (PND), Subject Headings Authority File (SWD), and Uniform Title File (EST) [8]. GND aims to create a web-compatible authority file connecting various information resources from German-speaking libraries and cultural institutions, opening data to academic and cultural communities beyond libraries via the semantic web [9]. GND uses the GND ontology as its data model [7], with major entity classes including Person, Conference or Event, Corporate Body, Place or Geographic Name, Subject Heading, and Work, with multiple subclasses under each.

In 2017, Shanghai Library utilized ontology and linked data technologies to transform traditional name authority documents into a new name authority database. This database integrates genealogical, rare collection, archival, and other special collection names from Shanghai Library, compiling information on 840,000 individuals. The database provides not only basic personal information but also key events, social relationships, and searchable, aggregatable, and analyzable works [10]. Shanghai Library's name authority database model is based on IFLA-LRM and BIBFRAME. Shanghai Library built the shlnames ontology for its name authority database, with main classes including Person, Place, Time, Event, Physical Object, and Material (see Figure 1 [Figure 1: see original paper]). The ontology reuses FOAF for describing persons, the Relationship vocabulary for inter-class relationships, and PROV-Ontology for events [11].

VIVO, funded by the NIH and jointly developed by seven institutions including Cornell University, Indiana University, and the University of Florida, aims to build a national scientist network supporting communication and discovery

using semantic web and ontology technologies. VIVO leverages linked data, ontologies, and visualization to achieve semantic interconnection of digital resources, representing an internationally prominent researcher information platform. The VIVO ontology serves as the data model for scientists, with main classes including Concept, Time, Journal, Person, Publication, Funding, and Organization. VIVO reuses 18 standard vocabularies, using FOAF for persons and organizations, BIBO for journals, and custom classes and properties for funding [12].

Additionally, in 2010, the National Library of Hungary published bibliographic and authority data as linked open data, using FOAF for name authority data [13]. In 2014, North Carolina State University Libraries published its Organization Name Authority (ONA) as linked data, reusing FOAF for organizations [14].

3.2 Analysis of Data Models in Linked Data Publishing Projects Data modeling involves conceptual model design followed by formalization. Conceptual model design analyzes entity attributes and relationships in data to abstract conceptual classes, properties, and relationships. Formalization employs standardized terminology to describe abstract concept properties, values, and relationships. Various communities have developed numerous metadata standards for specific purposes, including structural standards like MARC and Dublin Core, content standards like cataloging rules, value standards like classification systems and name authorities, and exchange standards like XML [15]. The "standard vocabularies" referenced in this paper's linked data model construction are structural standards defining data structure and semantics.

In conceptual model design, projects abstract entities and relationships based on their data content. Besides modeling authority objects as main classes, projects also model related information as associated classes when necessary. For example, Shanghai Library's person authority data extracts persons and related descriptive items as conceptual classes, while GND consolidates multiple authority documents (GKD, PND, SWD, EST) with richer extracted classes. Differences exist among projects: (1) they handle identical descriptive items differently—Shanghai Library treats time associated with persons as a class, while GND treats birth dates and publication dates as properties; (2) based on the complexity of their authority data types, some projects create multi-level classification systems with multiple subclasses (e.g., Shanghai Library, German National Library), while others maintain single-level systems (e.g., VIVO). All provide detailed metadata descriptions revealing concepts and relationships, formalizing the conceptual models.

Projects employ two formalization approaches: (1) using one standard vocabulary as the core supplemented by others for accurate description, such as VIAF using Schema.org as its core vocabulary [5]. Schema.org [16], sponsored by Google, Microsoft, Yahoo, and Yandex, provides consistent semantic description vocabulary across broad domains and allows extension under its struc-

ture. It includes over 600 classes and 900 properties covering persons, organizations, places, creative works, time, medicine, products, etc. OCLC recognized Schema.org as a potentially effective tool for describing library information resources, offering opportunities to integrate OCLC's rich bibliographic and authority data across the web [17]. Chinese scholar Jia Junzhi argues Schema.org's high generality, expressiveness, and rich data types offer advantages for describing Chinese person name authority data [18]; (2) establishing custom ontologies as data models that reuse multiple widely-used standard vocabularies or align with existing ones—i.e., combinations of multiple standard vocabularies plus customization. Examples include GND ontology for the German National Library, shlnames ontology for Shanghai Library, and VIVO ontology [12,19]. These projects frequently reuse standard vocabularies such as FOAF, vCard, and BIBO (see Table 2). FOAF is a semantic framework for linking people and information online, providing classes and properties for describing persons, groups, organizations, and documents. vCard focuses on describing persons and organizations, including location information [20]. First proposed in 1995, vCard was later complemented by FOAF (2005) and ORG ontology (2013). FOAF emphasizes relationships between people, agents, things, and social networks, while ORG ontology focuses on organizational structures, roles, and activities. The three ontologies have some overlap but provide useful vocabulary individually and enhanced information when used collaboratively [21]. BIBO is an ontology for describing various publications and related resources [22].

---

4. Modeling Research Entity Authority Data

Drawing on experiences from existing linked data projects and the characteristics of NSTL's name authority data, we constructed a linked data model for NSTL name authority data.

4.1 Conceptual Model Design In conceptual model design, we analyzed abstractable entities and inter-entity relationships in person, institution, funding, and journal authority data to establish an abstract conceptual model, shown in Figure 2 [Figure 2: see original paper]. NSTL's name authority data conceptual model includes five classes—Person, Organization, Funding, Work, and Place—and their relationships. The "Journal" class belongs to the "Work" class as a subclass.

4.2 Conceptual Model Formalization

4.2.1 Principles of Model Formalization For model formalization, W3C's linked data best practices recommend reusing existing standard vocabularies whenever possible to avoid reinventing the wheel and facilitate linked data exchange and sharing [23]. Existing linked data publishing projects follow this principle, reusing standard vocabularies or aligning with them when building

linked data models. In practice, finding a single vocabulary that meets all requirements is difficult, making it necessary to reuse multiple standard vocabularies. Generally, more vocabulary types lead to higher mixing complexity, greater linked data model complexity, and more difficult external data linking. Therefore, we should minimize the number of reused standard vocabularies while ensuring quality. This paper follows these principles in selecting standard vocabularies: (1) prioritize widely-used standard vocabularies in linked data publishing projects to ensure high interoperability; (2) adopt widely-used, well-maintained standard vocabularies to make the linked data model more understandable; (3) use fewer standard vocabulary combinations to reduce model complexity, facilitate external data linking, and lower maintenance costs; (4) for portions not covered by multiple standard vocabulary combinations, define custom classes and properties following best practices [24] as needed.

4.2.2 Comparison of Two Schemes As discussed in Section 3.2, existing linked data project formalization schemes fall into two categories. Which is more suitable for NSTL's name authority data? We designed two linked data model formalization schemes and compared them in terms of expressiveness for NSTL name authority data and model complexity to select the better option.

Referencing common practices in linked data publishing projects, we used frequently-used standard vocabularies, supplementing with others to build two schemes: (1) a Schema.org-centered scheme, referencing VIAF's use of Schema.org as the core vocabulary supplemented by FRAPO for funding description; (2) a multi-vocabulary combination scheme using FOAF, ORG, VIVO, BIBO, vCard, etc., as shown in Table 3 .

We mapped both schemes to NSTL name authority data. NSTL data uses element-attribute combinations—for example, dates are described by the element "date," with date types constrained by the attribute "date-type." "Birth date" is represented as `<date date-type="birth">……</date>`. In building the linked data model, we must consider both elements and attributes.

Three mapping types emerged: (1) one-to-one mapping: one element maps to one property (e.g., element `given-names` corresponds to `schema:givenName` in Schema.org); (2) one-to-many mapping: one element may map to multiple properties due to different attribute constraints (e.g., dates can be birth or death dates—birth date `<date date-type="birth">……</date>` maps to `schema:birthDate`, death date `<date date-type="death">……</date>` maps to `schema:deathDate`); (3) many-to-one mapping: for example, dates described by combined year, month, and day elements. A publication date of January 1, 2019, represented as `<date date-type="pub-date"><day>01</day><month>01</month><year>2019</year><` maps to `schema:datePublished` in Schema.org.

Mapping results show both schemes have elements and attributes they cannot express. We counted and compared these unmapped elements and attributes. For institutions, element mapping is shown in Table 4 and attribute mapping

in Table 5 . For institutional description, Schema.org provides rich properties, with only a few elements and attributes unmapped. In contrast, ORG lacks properties for expressing dates and external links when describing institutional authority data. For persons, Schema.org also provides rich properties, lacking only a few attributes for degrees and specialties. FOAF has fewer properties for occupation and education information, making Schema.org more expressive overall. For funding, FRAPO provides rich properties for currency types, subjects, and keywords compared to VIVO. For journals, BIBO is comparable to Schema.org, with Schema.org having relatively richer title categories, though neither distinguishes issue types such as main issues, sub-issues, or special issues. For address description, Schema.org and vCard are consistent.

Table 5 shows a comparison of unmapped institutional attributes between the two schemes. Statistical comparison of unmapped elements and attributes appears in Figure 3 [Figure 3: see original paper]. The left bar shows unmapped elements and attributes for the Schema.org-centered scheme, the right bar for the multi-vocabulary combination scheme. Overall, the multi-vocabulary combination scheme has more unmapped name authority data elements and attributes than the Schema.org-centered scheme. Therefore, in terms of expressiveness completeness, the Schema.org-centered scheme is superior.

Regarding model complexity, the Schema.org-centered scheme uses Schema.org as the core vocabulary supplemented by the funding vocabulary FRAPO, resulting in lower complexity compared to the four-vocabulary mixture of the multi-vocabulary combination scheme. This makes it easier to link with external data and to update and maintain. In vocabulary usage, Schema.org [16] is developed by major search engines, offering natural friendliness to search engines and making name authority data more discoverable on the web. Schema.org itself has rich data types that can accurately describe persons, organizations, journals, and places in name authority data. Using the same standard vocabulary to describe multiple objects maintains descriptive consistency and facilitates linking with existing linked datasets like VIAF that already use Schema.org as their core vocabulary.

Therefore, NSTL's name authority system should adopt the Schema.org-centered scheme for its linked data model.

4.3 Linked Data Model for NSTL Name Authority Data Using the Schema.org-centered scheme, we constructed the linked data model for NSTL name authority data, shown in Figure 4 [Figure 4: see original paper] (limited space shows only partial properties). In the conceptual model, Person, Organization, Funding, Work, and Place are represented as `Schema:Person`, `Schema:Organization`, `Frapo:Funding`, `Schema:CreativeWork`, and `Schema:PostalAddress`, respectively. `Schema:CreativeWork` represents all types of works, with journals represented as `Schema:Periodical`, a subclass of `Schema:CreativeWork`.

Solid lines between classes reveal inter-class relationships supported by NSTL name authority data, such as affiliation between persons and organizations (`Schema:affiliation`) and location relationships between organizations and addresses (`schema:location`). Lines connecting works with persons, organizations, and funding link authors, institutions, and funders with their outputs.

Dashed lines connected to classes represent data properties. Each entity has a unique URI identifier, with other identifier types represented by the property `Schema:identifier`. Names of all classes are represented by `Schema:name` and `schema:alternateName`, with person name attributes being particularly rich.

Additionally, `Schema:sameAs` in Schema.org can connect local authority data with external datasets such as DBpedia and VIAF.

---

5. Linked Data Publishing Experiment for Research Entities

We selected sample data from NSTL's name authority system, including 2,000 records each for persons, institutions, funding, and addresses. We used the widely-adopted D2RQ open-source software to publish name authority data as linked data.

D2R [25] is an open-source project from the Free University of Berlin that primarily converts relational database data into dynamic virtual RDF (Resource Description Framework) data and provides RDF views for querying relational databases. D2RQ mapping language provides a URI generation template mechanism to guide the creation of real URIs for instantiated resources. This experiment used absolute URI patterns, configuring each table object with the `d2rq:uriPattern` property in the form `JournalPaper/@@JournalPaper.PAPER_{ID}@@`, where the portion before "/" is the table name and the portion between "@@" is the table column.

Using D2RQ mapping language based on the linked data model for NSTL name authority data described in Section 4.3, we completed the semantic mapping from relational data tables to classes, properties, and relationships, creating mapping files from existing authority data to the linked data model, as shown in Figure 5 [Figure 5: see original paper].

Through mapping files, relational database data was transformed and accessed, publishing name authority data as linked data. Figure 6 [Figure 6: see original paper] shows the detail page for an institution after linked data publication—Peking University. Clicking the location field `schema:location` links to the corresponding location, enabling inter-entity connectivity, as shown in Figure 7 [Figure 7: see original paper].

---

6. Discussion and Conclusion

The experiment successfully published NSTL name authority data as linked data, demonstrating authority objects and their relationships in the linked data state. Based on the mapping between the Schema.org-centered scheme and NSTL name authority metadata (Section 4.2.2), we identified several issues in the publication process: (1) No universal data model can fully express all data. Although we reused existing standard vocabularies as much as possible to facilitate connections with external datasets, some data still lacks corresponding expressions in existing standard vocabularies (see Tables 4 and 5), requiring customization. Custom classes or properties increase data model complexity, hindering data reuse and interconnection; (2) Data loss may occur during conversion to linked data. Due to inconsistent granularity between the linked data model and original data model, more fine-grained data in the original model may be lost. For example, in NSTL's metadata, the element "abbreviated source title" (`abbrev-source-title`) has the attribute `abbrev-type` to distinguish different abbreviation types, and the element "date" (`date`) has the attribute `calendar` to distinguish different calendar types, but current standard vocabularies lack equally granular descriptions. This issue can be addressed by unifying calendar types or abbreviation types in the data.

In summary, this paper constructed a linked data model for NSTL name authority data based on analysis of linked data publishing projects. We compared two schemes—multiple standard vocabularies combination versus a single core vocabulary—and found the Schema.org-centered scheme offers better expressiveness, lower complexity, and easier internet integration. Therefore, we selected the Schema.org-centered scheme as NSTL's name authority linked data model. Finally, we used D2RQ to publish sample data as linked data, presenting the results to provide a reference for other institutions' linked authority data modeling.

Researching linked data models for authority data and publishing NSTL's name authority data as linked data facilitates data sharing, interconnection, quality improvement, and internet integration while providing a model reference for other institutions. Future research should explore the application value and prospects of published linked authority data, methods for achieving broader data linking and quality improvement, and how to leverage linked authority data to build richer knowledge services.

References

[1] Zeng Jianxun. Research on constructing scientific research relationship networks based on massive digital resources[J]. Journal of the China Society for Scientific and Technical Information, 2013, 32(9): 929-935. [2] Liu Wei, Zhang Chunjing, Xia Cuijuan. Authority control in the World Wide Web era[J]. Journal of Library Science in China, 2015(3): 2-33. [3] VIAF[EB/OL]. [2019-06-25].

https://www.oclc.org/zh-Hans/viaf.html. [4] Thomas BH, Jeffrey AY. Description of the VIAF (virtual international authority file) dataset[EB/OL]. [2019-06-25]. http://outgoing.typepad.com/outgoing/2015/04/viaf-rdf-changes.html. [5] VIAF. RDF changes[EB/OL]. [2019-06-25]. https://outgoing.typepad.com/outgoing/2015/04/viaf-rdf-changes.html. [6] OCLC adds linked data to WorldCat.org[EB/OL]. [2019-06-25]. http://www.oclc.org/news/releases/2012/201238.en.html. [7] Version history of the linked-data service[EB/OL]. [2019-11-22]. https://www.dnb.de/DE/Professionell/Metadatendienste/Datenbezug/LDS/lds_{versionhistorie}.html?nn=2
[8] GND ontology[EB/OL]. [2019-06-25]. https://d-nb.info/standards/elementset/gnd#CharactersOrMorphem
[9] Gemeinsame normdatei (GND)[EB/OL]. [2019-06-25]. https://lod-cloud.net/dataset/dnb-gemeinsame-normdatei. [10] Xia CJ, Liu W. Name authority control in digital humanities: building a name authority database of Shanghai Library[J]. International Journal of Librarianship, 2018, 3(1): 21-35. [11] Shanghai Library. Name authority database ontology (shlnames)[EB/OL]. [2019-06-25]. http://data.library.sh.cn/ont/ontology/tree?g=http://ont.library.sh.cn/graph/shlnames.
[12] Woods SA. Source ontologies for VIVO[EB/OL]. [2019-06-25]. https://wiki.duraspace.org/display/VIVOD
[13] National Library of Hungary. National széchényi library (national library of Hungary) on the semantic web[EB/OL]. [2019-06-25]. http://nektar.oszk.hu/wiki/Semantic_{web}.
[14] Hanson EM. A beginner's guide to creating library linked data: lessons from ncsu's organization name linked data project[J]. Serials Review, 2014, 40(4): 251-258. [15] Hu Xiaojing. Bibliographic control: from digitization to datafication[J]. Journal of Library Science in China, 2019, 45(3): 49-61. [16] Google, Yahoo, Microsoft, et al. Schema.org[EB/OL]. [2019-10-10]. https://schema.org/. [17] Zhang Xuesong, Tan Hairong, Yao Xiangzhong. SchemaBibEx and its application for online bibliographic resource description[J]. Library Journal, 2016(5): 67-75. [18] Jia Junzhi, Shi Yanqing. Research on linked data transformation of Chinese personal name authority files[J]. Journal of the China Society for Scientific and Technical Information, 2016, 35(7): 696-703. [19] Li Boyang. Research on revealing scientific research relationships based on linked data[D]. Changchun: Northeast Normal University, 2016. [20] Nogales AM, Elena GB. Measuring vocabulary use in the linked data cloud[J]. Online Information Review, 2017, 41(2): 252-271. [21] Iannella R, McKinney J. vCard ontology—for describing people and organizations[EB/OL]. [2019-06-25]. https://www.w3.org/TR/vcard-rdf/. [22] D'Arcus B, Giazzon F. Bibliographic ontology specification[EB/OL]. [2019-06-25]. http://bibliontology.com/. [23] Hyland B, Atemezing G, Terraza VB. Best practices for publishing linked data[EB/OL]. [2019-05-23]. https://www.w3.org/TR/ld-bp/. [24] Wood D, Zaidman M, Ruth M, et al. Linked data: structured data on the web[M]. Jiang Nan, trans. Beijing: Posts & Telecom Press, 2018: 3. [25] Bizer C, Cyganiak R. D2R server—publishing relational databases on the semantic web[EB/OL]. [2019-07-13]. http://richard.cyganiak.de/2008/papers/d2r-server-iswc2006.pdf.

Author Contributions

Zhou Yi: Wrote and revised the paper;
Zhang Jianyong: Responsible for NSTL name authority data and experiments;
Liu Zheng: Developed the paper framework and revised the paper;
Liu Xiumin: Provided technical support for publishing experiments.

Abstract

[**Purpose/significance**] The purpose of this paper is to study the linked data model of publishing the NSTL's research entity name authority data as linked data. After the name authority data is published as linked data, it can be reused as an open linked dataset by other systems or organizations, and also can be better integrated with other linked datasets to improve data quality. In addition, it also provides a model building reference for other organizations to publish authority data as linked data. [**Method/process**] First, this paper analyzed and compared the data models used in the linked data publishing projects at home and abroad. It showed that the data models in the linked data publishing projects were mainly divided into two categories. Then, combined with the characteristics of NSTL name authority data, two forms of linked data models were designed. It compared the two models from the expression level of the NSTL's data and the complexity of the models. The better one was selected. Finally, it used D2RQ as tool to publish the sample data as linked data. [**Result/conclusion**] The analysis found that the model with Schema.org as the core standard vocabulary has better expression completeness and lower model complexity, and is easier to integrate into the internet. So it is more suitable as a linked data model for NSTL's name authority data.

Keywords: research entities; name authority; linked data; data model

Note: Figure translations are in progress. See original paper for figures.