How to Evaluate and Select a Data Repository for Humanities and Social Science: A Case Study of Fudan University Data Repository for Humanities and Social Science

Shenqin Yin, Jilong Zhang, Menghao Jia, and Jie Hu

Abstract

The evaluation and selection of the Data Repository for Humanities and Social Science (DRHSS) lacks efficient reference models, which affects the quality of DRHSS selection decision making. Based on literature research, field research, and a focus group method, this research proposes an evaluation criteria framework for the DRHSS system, which comprehensively uses qualitative assessment and quantitative assessment methods. Qualitative assessment focuses on comparative analysis of system functions, and a quantitative assessment describes the evaluation criteria framework and its comprehensive weighting score system. We want to apply the evaluation and selection criteria framework to assess the most appropriate system to ensure economical and technical feasibility. This case study looks at Fudan University's investigation of the four major data repository platforms, Dspace, Fedora Commons, Dataverse, Nesstar, and an online analysis software application, SDA. Furthermore, we provide guidance for using the framework along with recommendations and limitations.

Introduction

Quantitative research based on data is one of the basic social science research paradigms as well as one of the most important steps and methods of scientific research (Balnaves and Caputi 2001). As early as 1662, John Graunt's book on population mortality in London was regarded as a staple of quantitative research (Snow 1936). From the artificial statistics of more than three hundred years ago to the era of big data today, vast amounts of data about social science research have been accumulated. The storage, management, distribution, analysis, sharing, and transmission of humanities and social science data are of great importance for researchers, research institutions, colleges, universities, and countries. To-

LIBRARY TRENDS, Vol. 69, No. 1, 2020 ("Digital Humanities and Libraries in China," edited by Lian Ruan and Xingye Du), pp. 125–37.

gether, these all guarantee exchanges and development of the research in this field.

In the past, humanities and social science data were predominantly owned by just a few national institutions and researchers and used only within a few disciplines. In recent years, some institutions and universities in China and other countries have realized the importance of building data repositories to manage data centrally and share data with the public. Some practices were gradually established (Meng and Qian 2013), and some well-known American and European social science data centers have a history of more than seventy years. These data centers have built relatively mature social science data repositories.

We have conducted field research of European and American institutions challenged with managing humanities and social sciences data. The results indicate that many institutions have built humanities and social sciences data repositories focused on the management of system function, metadata, business models, service modes, and research policy. These repositories are similar to the Inter-University Consortium for Political and Social Research (ICPSR) at the University of Michigan, the UK Data Archive (UKDA), and the Harvard Institute for Quantitative Social Science. There are four kinds of European and American humanities and social sciences data management platforms:

- Data repositories built by the universities, like ICPSR. The University of Michigan started to build the platform in the 1990s and continued upgrading it into new versions.
- Data repositories that are secondarily developed from open-source software. Currently, the main open source software, including Dspace, was developed by the Massachusetts Institute of Technology (MIT; Smith et al. 2003); Fedora Commons was developed by Cornell University (Sefton and Lucido 2009); and Dataverse was developed by Harvard University and MIT (King 2007). In recent years, the data sharing centers at the Hong Kong University of Science and Technology and the University of Edinburgh adopted Dspace to build their platforms. DataStaR at Cornell University, Data Conservancy at Johns Hopkins University, and Embedding Institutional Data Curation Services in Research (EIDCSR) at the University of Oxford adopted Fedora to build their data repositories. Dataverse is also widely used in social science research and already has more than twenty institutional users.
- Commercial software, like Nesstar. Nesstar was developed by the Norwegian Centre for Research Data (NSD; Assini 2002). More than one hundred public and academic institutions are Nesstar's customers. There are many famous users such as UKDA, NORC at the University of Chicago, and the Survey Research Data Archive (SRDA) in Taiwan, China.
- Platforms that adopted online commercial analysis software, like SDA.

SDA was developed by the University of California, Berkeley, and is used by many people in social data research centers all over the world.

These data repositories have undergone a long-term iterative process of research and development, use, update, reuse, and upgrade. Metadata and data harvesting exchange protocols complying with international standards and top-down data management policies meet the requirements of data management and sharing in Europe and the United States. For Chinese researchers, some of the Data Repository for Humanities and Social Sciences (DRHSS) does not support simplified Chinese characters or word segmentation, and some does not support online analysis and visualization of Chinese data variables. Direct migration of data platforms in Europe and the United States does not apply to the actual situation and cultural background of Chinese universities. The solution is to make local software based on technology appropriation. How should existing software be evaluated and selected? The process requires a relatively complete and practical DRHSS selection criteria framework.

Literature Review

The evaluation of information systems is already a complex issue (Irani 2002). Researchers have been trying to create or use models to carry out evaluation work. There are many examples of these models, such as Technology, Organization and Environment (TOE) theory (Zhu, Kraemer, and Xu 2002), the Open Archival Information System (OAIS) model (Jeng, He, and Chi 2017), the Technology Acceptance Model (Bueno and Salmeron 2008), Contingency Theory (Wang et al. 2008), the TOP (technical, organizational [or societal], personal [or individual] perspective) Model (Jahanyan, Azar, and Fard 2012), and so on. Piotrowicz and Cuthbertson (2009) introduced sustainability as a new dimension of information systems (IS) evaluation. The framework was developed for supply chain evaluation rather than IT/IS evaluation, which creates shortcomings in its direct application to all scenarios. All of these theoretical frameworks explain some of the issues and/or methodologies that guide them and bring attention to a narrow range of issues.

Some studies have applied theoretical frameworks for evaluating information systems. Özkan, Hackney, and Bilgen (2007) provide a framework for substantive information systems evaluation factors (PRISE) that supports the flexibility and relevance of PRISE as a framework for information systems evaluation. However, they did not attempt to attribute weights to the processes nor derive a combined unique measure of IS effectiveness. Razieh Dehghani and Raman Ramsin (2015) provide a criteria-based evaluation framework for assessing knowledge management system (KMS) development methodologies. Ji, Liu, and Jin (2018) construct a three-dimensional WSR analysis framework to evaluate an energy management

system. However, the evaluation frameworks above have not been applied to all kinds of information system development methodologies.

Some studies have emphasized the importance of the technological infrastructure of information systems (Curty 2016). Fecher, Friesike, and Hebing (2015) refined three factors: architecture, usability, and management software. Jeng and Lyon (2016) thought availability, usability, facilities, and technical standards were challenging, whereas March and Hevner (2007) presented a layered architecture, including the kernel content management layer, the integration and design layer, and the use layer.

Several studies have focused on system design and users' acceptance. Taylor's Value-Added model (1986) was a broad and ambitious effort to provide a unified framework for focusing on user needs and preferences in evaluating and designing information systems. The Model of Technology Appropriation (Carroll 2004; Carroll, Mendoza, and Stern 2005) represents the process of appropriation from adoption through long-term use. It draws attention to the crucial role played by users' actions in completing the design process and examines the implications for the design and implementation of technological innovations. Alam and Campbell (2014) refined the TAC model and stated that the design process would be incomplete without user participation. For managers, trainers, and IS staff involved in the implementation of innovations, the challenge is to encourage and support users' appropriation activities.

Other researchers have focused on different criteria. Goh et al. (2006) developed a checklist consisting of twelve categories of items for DL (digital library) evaluation. The following five broad requirements were used as their evaluation criteria: content management, user interface, user administration, system administration, and other requirements. Rieger (2007) proposed the repository model selection process, which involves several essential stages including stakeholder analysis, needs assessment, service definition, and identification of use cases and governance-related matters. Marill and Luczak (2009) developed a set of "Master Evaluation Criteria" to provide a decision method for the Digital Repository Evaluation and Selection Project of the National Library of Medicine. The Master Evaluation Criteria included functionality, scalability, extensibility, interoperability, ease of deployment, system security, system performance, physical environment, platform support, demonstrated successful deployments, system support, strength of development community, stability of development organization, and a strength of technology roadmap for the future. The NDIIPP ECHO project digital repository evaluation (Marill and Luczak 2009) used an augmented version of the draft audit checklist for certification of trusted digital repositories (audit checklist) to provide a framework for examining how well current popular repository software applications supported the notion of a "trusted digital repository." Leroux, McBride, and Gibson (2011) proposed a set of eleven desirable guidelines

to determine the appropriateness of the Clinical Trial Management System (CTMS) solution. They hoped these guidelines could set a standard and serve as a reference point for designing CTMS. A diverse set of stakeholders have come together to design and jointly endorse a concise and measurable set of principles that we refer to as the FAIR Data Principles (Wilkinson et al. 2016).

However, these approaches are partial to practice. They have not formed a systematic theoretical framework and lack the factors of variance, concrete evaluation methods, and measurement tools. This study puts emphasis on DRHSS's evaluation and selection, considering that system factors are more complex and specific, involving functional matching, technical feasibility, extensibility, economic feasibility, and personnel and institutional arrangements. We developed a criteria framework with a weighted score for evaluating and selecting DRHSS. The proposed framework is validated by applying it to a case study. Institutions can select the methodology that best fits their requirements based on the evaluation results.

Methodology

This study builds a comprehensive system evaluation criteria framework that integrates both qualitative and quantitative assessment methods. The qualitative assessment focuses on the assessment of functional requirements of the system platform, using comparative analysis methods to thoroughly understand the functional characteristics and differences of various software. The specific checklist is shown in the functional requirements section of Table 1 (excluding the scores in parentheses; available at https://doi.org/10.5281/zenodo.3677581). Functional requirements were initially derived through literature research and field research, and subsequent adjustments were made based on expert recommendations.

Quantitative assessment is a weighted system evaluation criteria framework, and the construction steps are as follows: First, through literature research, field research, and focus group interviews, the DRHSS evaluation criteria were obtained, including five first-level dimensions—system function requirements, usability, scalability, operation and maintenance, and market and community. In this case, the first-level dimensions can be subdivided into secondary dimensions, and some important secondary dimensions have three subdimensions, such as metadata standards and online analysis visualization. Second, twenty-two experts in the field of data management were invited to give suggestions and opinions about the DRHSS evaluation criteria, and the framework was comprehensively examined. A working group provided background information to the experts and sought their opinions anonymously. They then analyzed and summarized the expert opinions and gave the statistical results to the experts. The experts revised their opinions according to the feedback re-

sults, and after several rounds of anonymous consultation and feedback, the final analysis conclusion was formed. The Analytic Hierarchy Process was used to calculate the data. The consistency rate (CR) was used as an indicator to measure the consistency of scoring. A CR value not more than 10 percent indicates good consistency. Finally, the weighted scoring evaluation criteria framework was made based on the scores given by the experts to the first and second levels of evaluation factors as shown in Table 1 (Table 1: Weighted Scoring Evaluation Criteria Framework, https://doi.org/10.5281/zenodo.3677581).

In practical applications, the functional requirements occupy 50 percent of the weight, qualitatively comparing and analyzing the functional parts of different software that is suggested, and the checklist is shown in the functional requirements section of Table 1. On the basis of qualitative assessment, the evaluation framework of the DRHSS was used. Functional requirements, scalability, operations and maintenance, market, and community conditions were suggested to be evaluated by the project team, and usability was assessed by users. Then, the total score was obtained by a comprehensive evaluation.

Case Analysis

Fudan University is located in Shanghai, China, and is a comprehensive university that is among the top five in Chinese universities. Fudan University's Institute for Social Research (FISR) was founded in 2011 as a data repository that supported the exchange and sharing of data in humanities and social science (hereinafter referred to as DRHSS). DRHSS aims to integrate and develop all the humanities and social science data resources of Fudan University, providing services including data submission, conservation, management, and sharing. DRHSS also aims to help improve the international academic status and influence of researchers and social science data centers and to promote the exchange and development of humanities and social sciences research. The repository also provides basic support for Fudan University to perform the duties of cultural inheritance, record changes, and serve the country as a "national think tank."

FISR is set up with an academic committee and international academic advisory committee to guide the development of the center's business. FISR consists of a Data Service Department, Social Investigation Department, and Research Department. The Data Service Department is mainly responsible for the research, development, and maintenance of the Data Repository, setting metadata standards and rules for cataloging, as well as propelling the formulation of relevant support policy for scientific data sharing at the university. Researchers mainly come from the Fudan University Library, the School of Social Development and Public Policy, and the School of Computer Science.

In order to build the data repository, FISR established a working group (WG) in 2011. The WG completed the survey on DRHSS supported by several social science research institutions and famous universities at home and abroad in over three years. Based on the survey, the WG started to conduct demand analysis, metadata specification, standards and design scheme setting, technology appropriation, custom development, and testing and deployment. The data repository was officially launched on December 31, 2014, and the WG carried out a series of services and promotional work that achieved good results. This paper focuses on how to apply the system evaluation criteria framework, using both qualitative and quantitative assessments and selecting an appropriate data repository for localization and secondary development.

Phase I: Qualitative Assessment

Comparison of Functions of Four Data Repositories. The WG deployed a test instance of the four major platforms.

In order to conduct selection assessment correctly between several platforms and find the best fit for a data repository, the WG deployed a test instance of the four major platforms, Dspace, Fedora Commons, Dataverse, and Nesstar. A critical mass of actual data was uploaded for testing, and the functions of each platform and online analysis software SDA were all tested together.

In this study, we compared and assessed the platforms from the matching degree of functional requirements, system specification, content management, system management, user interface and retrieving, file saving, and system maintenance. The four platforms all have basic onstage and backstage management functions. The advanced features, which are listed in Table 2, especially need special investigation (Table 2: System Comparison of Advanced Functions, https://doi.org/10.5281/zenodo.3677581).

Comparison of Metadata Standards. DSpace and Fedora Commons adopted the Dublin Core metadata framework and support extended metadata. Dataverse and Nesstar adopted the internationally accepted Data Documentation Initiative (DDI).

Dublin Core has fifteen core element sets, which can be divided into three groups according to the category and scope of its described content. These groups include the description of resources, description of intellectual property rights, and description of external properties. DDI has 101 elements, and according to the generality of social science research resources, DDI divides the descriptions of each resource collection into subject reference information, abstract and scope, data collection and methods, data availability, terms of use, other information, and document description. Each set describes resource objects from different perspectives (DDI Alliance 2013).

Dublin Core and DDI were both established in 1995. While their application fields are different, Dublin Core's emphasis is on personalized,

simple, and easy application, but the description of the basic object is not thorough enough and cannot achieve a high-degree term specificity retrieval. DDI can be applied to the professional scientific research field, and it is of greater integrity, high compatibility, and expansibility. It can both describe macro data and go deep into the micro level of data (Liyun and Hu 2005).

Fudan University has a vast amount of data in various subjects and types. According to the subject and type, we adopt simple description or complex description. Specifically, there are plenty of attributes of most survey data stored in the repository. For example, the data for investigations about changes in the Yangtze River Delta include attributes of coding data, data geographic coverage, geography information unit, data type, data collection scope, sampling process, data weighting, data cleaning, questionnaire recovery rate, and sampling error estimation. Additionally, it needs to provide a link to an explanation for terms of use. The attributes are very important to data searching and navigation, and this kind of data usually needs precise description to elaborate on the unique properties.

In addition to the data itself, the related original documents, derivative publications, and related information subjects within the same topic also need to be navigated. DDI provides a description of these projects' reference information.

Additionally, we found out that the metadata standard DDI could better describe social science data. Besides DDI, Dataverse also supports metadata standards for more than ten other disciplines.

Comparison of Online Data Analysis Function. Social science data contain a lot of survey results and can usually be analyzed online. Therefore, we intensively examined and tested the online analysis functions of Dataverse, Nesstar, and SDA. The results are shown in Table 3 (Comparison of Online Analysis Function, https://doi.org/10.5281/zenodo.3677581).

Taking the statistical data of "model and empirical analysis of the impact of population and consumption on carbon emissions" as an example, Dataverse can show the changes of carbon emissions, carbon emissions intensity, and per capita consumption from 1980 to 2006 by using time series analysis. Nesstar and SDA do not provide similar functions.

Nesstar supports quick view and simple statistical analysis. It is very easy to use and suitable for users who need to analyze secondhand data, but it does not support advanced statistical analysis. SDA has the most comprehensive functions and can support the most complicated analyses. It has a strong and professional statistical analysis function that can meet most users' demands, but its usability needs to be improved. SDA's system is inconvenient for users to check variables, and the analysis results are too complex to read. Furthermore, the data that need to be analyzed must be uploaded to the server of the University of California, Berkeley, which is not effective for data management. Dataverse supports data sharing, automatic format conversion, and advanced statistical analysis and data

visualization. Its usability is better compared to the systems from Nesstar and SDA.

Phase II: Quantitative Assessment

Based on the qualitative assessment of Phase I functional requirements, twelve project team members used the evaluation criteria framework in Table 1 to score the systems. Specially, the usability of the systems was evaluated by fifty invited teachers and students. Dataverse attained the highest score.

Results

The project team spent half a year conducting two rounds of comparison and screening of the four data repositories. Methods of qualitative assessment and quantitative assessment were used comprehensively.

In the first round, the functional requirements account for 50 percent of the overall score. According to the functional requirements list of Table 1, the WG used the qualitative assessment method and conducted a detailed evaluation and comparison. Dspace and Fedora Commons are common institutional platform software, which can meet the general functional requirements of uploading, publishing, storing, displaying, and so on, but they are not tailored for DRHSS research and development. Their metadata description information of research results and humanities and social science data is relatively simple. There is no special online analysis, data visualization, or commenting functions for humanities and social science data. Dspace and Fedora Commons cannot meet the advanced functional requirements. Nesstar does not open the data submission function for researchers and users. This function is only for the background administrator to release data uniformly, and it is suitable for official or institutional exclusive data release; it does not support data citation, automatic conversion of data formats, data template customization, or other functions.

The second round of quantitative assessment included comprehensive consideration of five aspects: system functional requirements, usability, scalability, operation and maintenance, and market and community situation. These aspects were scored by the project team members and users' representatives, and eventually Dataverse ranked highest. Fudan University chose Dataverse and conducted localization and secondary development. A primary advantage of Dataverse is that it is specially developed for the long-term data management and sharing of DRHSS, which can meet the basic and advanced functional requirements of the DRHSS of Fudan University. Other advantages of Dataverse are that the interface is user-friendly; it is open-source software and supports secondary development; it can be customized according to requirements with good scalability; and it is easy to operate and maintain. The numbers of Dataverse institutional users and communities are expanding.

In general, combining qualitative assessment with quantitative assessment is more suitable for DRHSS's evaluation and selection. Qualitative assessments make a detailed analysis of functional requirements, particularly metadata standards and online analysis capabilities, whereas quantitative assessment takes into account multiple major factors in the evaluation and selection of DRHSS by means of quantification and provides intuitive results. Through qualitative evaluation, the project team can better score the quantitative evaluation.

FISR finally confirmed the Dataverse platform was developed by Harvard University as the prototype. We conducted a system translation from an English version to a Chinese version and performed customized development. DRHSS has now been put into operation at http://dvn.fudan.edu.cn. The platform combines the functions of online analysis and data resource management and can be used for data access, storage, transmission, and study communication for researchers. It can help promote data sharing and help researchers to reproduce others' work. In this way, researchers can create, submit, monitor, and share their research data easily.

There are already more than 200 research projects and 645 total datasets. There are scientific databases like the Yangtze residents' consumption and carbon emissions database, all previous censuses of the People's Republic of China by province, China's population, and others. An exclusive directory of "scientific research of liberal arts" was set up on the platform home page, supported by scientific research departments in Fudan. There are three columns in the directory:

- *Faculty:* The platform serves 1,319 faculty who are undertaking research projects, setting exclusive modules such as personal information, teaching plan, courseware, research programs, media reports, social services, academic activities, and research achievements.
- *Scientific Research Project:* A data space for "scientific research of liberal arts" was set up and has been imported to more than 5,000 projects.
- *Research Findings:* A data space for "research findings of liberal arts" was set up and has been imported to more than 45,835 academic theses.

To date, page views for platform homepage number close to five million, and the data have been widely used in teaching, study, research, papers, dissertations, and so on.

For evaluating and selecting DRHSS, Nanjing University and Peking University from China also used this method and finally chose Dataverse as their data repository.

Discussion

This paper has presented a comprehensive evaluation criteria framework model for qualitative and quantitative evaluation of data repositories. It has considered the main factors affecting the selection of DRHSS, further

subdividing the main factors into two dimensions, and subdividing the key two dimensions into three dimensions, which is more comprehensive and systematic and has theoretical significance.

In the qualitative evaluation part, we paid attention to the functional requirements of the system and summarized the functional characteristics of different systems through detailed comparative analysis, which is more intuitive and meticulous. On the basis of qualitative evaluation of functional requirements, quantitative evaluation and scoring are carried out to ensure that project team members have sufficient knowledge and experience of using the system, and scoring is more objective. Quantitative weights are given by experienced experts in this field to ensure that the evaluation results are operable and feasible and have practical significance. The framework model of this study has been applied in the evaluation and selection of Fudan University's DRHSS and achieved good results.

Research and practice in China and other countries show that major social science data platforms in Europe and the United States are more mature. However, there are different scientific management policies and cultural backgrounds in different countries. An overseas platform cannot be directly transplanted into China. The challenges and issues that an institution may encounter during DRHSS system customization and localization include using a Chinese search engine, Chinese word segmentation, and navigation with a Chinese character index as well as online analysis in supporting Chinese characters. The system evaluation criteria framework provides an efficient tool for accessing a data repository.

We recommend that institutions follow these steps of qualitative and quantitative assessment, drawing on their own weighted scoring system. First, in the process of evaluating and selecting DRHSS, the target positioning and functional requirements must be made clear. Second, demand, funds, human resources, funding costs, and disciplines should be taken into consideration when choosing software and a development mode to build DRHSS, especially as the description reveals the degree of resources are different according to the metadata standards in various disciplines. Platforms can be built based on the mature data sharing platform software or their main functions. These uncertain elements need to be adjusted in the framework model according to the actual situation in order to match the needs and capabilities of the organization.

References

Alam, Sultana, and John Campbell. 2014. "Examining Cultural Volunteer Crowdsourcing Technology: An Appropriation Perspective." In *35th International Conference on Information Systems (ICIS 2014), Auckland, NZ, December 14–17, 2014.* https://aisel.aisnet.org/icis2014/proceedings/SocialMedia/4/.

Assini, Pasqualino'titto. 2002. "NESSTAR: A Semantic Web Application for Statistical Data and Metadata." In *Eleventh International World Wide Web Conference: Workshop on*

Real World Applications of RDF and the Semantic. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.536.9314.

Balnaves, Mark, and Peter Caputi. 2001. *Introduction to Quantitative Research Methods: An Investigative Approach*. London: Sage.

Bueno, Salvador, and Jose L. Salmeron. 2008. "Tam-Based Success Modeling in ERP." *Interacting with Computers* 20 (6): 515–23.

Carroll, Jennie. 2004. "Completing Design in Use: Closing the Appropriation Cycle." In *ECIS 2004 Proceedings* 44, 337–47. AIS Electronic Library.

Carroll, Jennie, Antonette Mendoza, and Linda Stern. 2005. "Adoption, Adaptation, Stabilization and Stagnation: Software Appropriation over Time." In *Australasian Conference on Information Systems 2005 Proceedings, Sydney, Australia, November 29–December 2, 2005*, 7. AIS Electronic Library.

Curty, Renata Gonçalves. 2016. "Actors Influencing Research Data Reuse in the Social Sciences: An Exploratory Study." *International Journal of Digital Curation* 11 (1): 96–117.

DDI Alliance. 2013. "Document, Discover and Interoperate." http://www.ddialliance.org/.

Dehghani, Razieh, and Raman Ramsin. 2015. "Methodologies for Developing Knowledge Management Systems: An Evaluation Framework." *Journal of Knowledge Management* 19 (4): 682–710. https://doi.org/10.1108/JKM-10-2014-0438.

Fecher, Benedikt, Sascha Friesike, and Marcel Hebing. 2015. "What Drives Academic Data Sharing?" *PLOS ONE* 10 (2): 1–25.

Goh, Dion Hoe-Lian, Alton Yeow Kuan Chua, Davina Anqi Khoo, Emily Boon Hui Khoo, Eric Bok Tong Mak, and Maple Wen Min Ng. 2006. "A Checklist for Evaluating Open Source Digital Library Software." *Online Information Review* 30 (4): 360–79.

Hong, Zhengguo, and Ying Xiang. 2013. "Construction of University Scientific Data Management Platform Based on Dspace—A Case Study of Scorpion Species and Toxins Database." *Library and Information Service* 6: 39–42.

Irani, Zahir. 2002. "Information Systems Evaluation: Navigating through the Problem Domain." *Information & Management* 40 (1): 11–24.

Jahanyan, Saeed, Adel Azar, and Hasan Danaee Fard. 2012. "Utilising Multi-aspectual Understanding as a Framework for ERP Success Evaluation." *Journal of Enterprise Information Management* 25 (5): 479–504.

Jeng, Wei, Daqing He, and Yu Chi. 2017. "Social Science Data Repositories in Data Deluge: A Case Study of ICPSR's Workflow and Practices." *Electronic Library* 35 (4): 626–49. https://doi.org/10.1108/EL-11-2016-0243.

Jeng, Wei, and Liz Lyon. 2016. "A Report of Data-Intensive Capability, Institutional Support, and Data Management Practices in Social Sciences." *International Journal of Digital Curation* 11 (1): 156–71.

Ji, Boya, Yuming Liu, and Zhanyong Jin. 2018. "An Evaluation of the Design and Construction of Energy Management Platform for Public Buildings Based on WSR System Approach." *Kybernetes* 47 (8): 1549–68. https://doi.org/10.1108/K-07-2017-0265.

King, Gary. 2007. "An Introduction to the Dataverse Network as an Infrastructure for Data Sharing." *Sociological Methods & Research* 36 (2): 173–99.

Leroux, Hugo, Simon McBride, and Simon Gibson. 2011. "On Selecting a Clinical Trial Management System for Large Scale, Multi-Centre, Multi-Modal Clinical Research Study." *Studies in Health Technology and Informatics* 168 (1): 89–95.

Liu, Runda. 2013. "Evaluation on Sharing Websites of Chinese Scientific Data." *China Science News*, June 17. http://news.sciencenet.cn/sbhtmlnews/2013/6/274215.shtm?id=274215.

Liyun, Yang, and Bo Hu. 2005. "Metadata Standard for Science Information Organization-DDI." *New Technology of Library and Information Service* 8: 00002.

March, Salvatore T., and Alan R. Hevner. 2007. "Integrated Decision Support Systems: A Data Warehousing Perspective." *Decision Support Systems* 43 (3): 1031–43.

Marill, Jennifer L., and Edward C. Luczak. 2009. "Evaluation of Digital Repository Software at the National Library of Medicine." *D-Lib Magazine* 15 (5–6).

Meng, Xiangbao, and Peng Qian. 2013. "International Experiences and References of University Social Sciences Data Management: Taking UKDA and ICPSR for Example." *Information and Documentation Services* 34 (2): 77–80.

Özkan, Sevgi, Ray Hackney, and Semih Bilgen. 2007. "Process Based Information Systems Evaluation: Towards the Attributes of 'PRISE.'" *Journal of Enterprise Information Management* 20 (6): 700–725. https://doi.org/10.1108/17410390710830736.

Piotrowicz, Wojciech, and Richard Cuthbertson. 2009. "Sustainability—A New Dimension in Information Systems Evaluation." *Journal of Enterprise Information Management* 22 (5): 492–503. https://doi.org/10.1108/17410390910993509.

Rieger, Oya Y. 2007. "Select for Success." *D-Lib Magazine* 13 (7–8): 1–8.

Sefton, Peter, and Oliver Lucido. 2009. "The Fascinator: A Lightweight, Modular Contribution to the Fedora-Commons World." In *Open Repositories 2009 Proceedings, Atlanta, GA, May 18–21, 2009*. https://eprints.usq.edu.au/5259.

Smith, MacKenzie, Mary Barton, Margret Branschofsky, and Mick Bass. 2003. "DSpace: An Open Source Dynamic Digital Repository." *D-Lib Magazine* 1: 51–65.

Snow, John. 1936. *Snow on Cholera*. London: Humphrey Milford.

Taylor, Robert Saxton. 1986. *Value Added Processes in Information Systems*. Westport, CT: Greenwood.

Wang, Eric T. G., Sheng-Pao Shih, James J. Jiang, and Gary Klein. 2008. "The Consistency among Facilitating Factors and ERP Implementation Success: A Holistic View of Fit." *Journal of Systems and Software* 81 (9): 1609–21. https://doi.org/10.1016/j.jss.2007.11.722.

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 2016. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data* 3: 160018.

Zhu, Kevin, Kenneth Kraemer, and Sean Xu. 2002. "A Cross-Country Study of Electronic Business Adoption Using the Technology-Organization-Environment Framework." In *Proceedings of the 23rd International Conference on Information Systems*, 31. https://www.mendeley.com/catalogue/79241d31-96f2-3c96-a6f3-a151751672b5/.

Shenqin Yin (corresponding author) is a researcher of the Institute for Humanities and Social Science, School of Big Data, Fudan University, China. She serves as a deputy director of Shanghai Big Data Joint Innovation Lab - Science & Research Unit. Her research has been focused on scientific data management and open government data, and she has a number of publications in journals.

Jilong Zhang serves as deputy director of the Fudan University Library. His research focuses on informative campus and scientific data management, and he has a number of publications in journals.

Menghao Jia is a PhD candidate at the University of South Australia, majoring in computer and information science.

Jie Hu is a library assistant at Fudan University Library. Her research focuses on database publishing and scientific data management.