Library Hi Tech

An empirical study on the performance evaluation of scientific data sharing platforms in China

Li Si Yueting Li Xiaozhe Zhuang Wenming Xing Xiaoqin Hua Xin Li Juanjuan Xin

An empirical study on the performance evaluation of scientific data sharing platforms in China

Received 26 September 2014
Revised 18 December 2014
Accepted 6 March 2015

Li Si
Center for the Studies of Information Resources, Wuhan University, Wuhan, China
Yueting Li and Xiaozhe Zhuang
Department of Library Science, School of Information Management, Wuhan University, Wuhan, China
Wenming Xing
Department of Library Science, School of Public Administration, Xiangtan University, Xiangtan, China
Xiaoqin Hua
Zhejiang Institute of Hydraulics & Estuary, Hangzhou, China, and
Xin Li and Juanjuan Xin
Department of Library Science, School of Information Management, Wuhan University, Wuhan, China

Abstract

Purpose – The purpose of this paper is to conduct performance evaluation of eight main scientific data sharing platforms in China and find existing problems, thus providing reference for maximizing the value of scientific data and enhancing scientific research efficiency.

Design/methodology/approach – First, the authors built an evaluation indicator system for the performance of scientific data sharing platforms. Next, the analytic hierarchy process was employed to set indicator weights. Then, the authors use experts grading method to give scored for each indicator and calculated the scoring results of the scientific data sharing platform performance evaluation. Finally, an analysis of the results was conducted.

Findings – The performance evaluation of eight platforms is arranged by descending order by the value of *F*: the Data Sharing Infrastructure of Earth System Science (76.962), the Basic Science Data Sharing Center (76.595), the National Scientific Data Sharing Platform for Population and Health (71.577), the China Earthquake Data Center (66.296), the China Meteorological Data Sharing Service System (65.159), the National Agricultural Scientific Data Sharing Center (55.068), the Chinese Forestry Science Data Center (56.894) and the National Scientific Data Sharing & Service Network on Material Environmental Corrosion (Aging) (52.528). And some existing shortcomings such as the relevant policies and regulation, standards of data description and organization, data availability and the services should be improved.

This paper is one of the research outcomes of a Humanities and Social Sciences Project supported by the Ministry of Education of P.R. China (Project Name: Study on the Guarantee of Scientific Data Integration and Sharing in China, Project No. 13YJA870016).

Library Hi Tech
Vol. 33 No. 2, 2015
pp. 211-229

0737-8831
DOI 10.1108/LHT-09-2014-0093

Originality/value – This paper is mainly discussing about the performance evaluation system covering operation management, data resource, platform function, service efficiency and influence of eight scientific data sharing centers and made comparative analysis. It reflected the reality development of scientific data sharing in China.

Keywords Performance measurement, Data management

Paper type Research paper

1. Introduction

Scientific data refers to the raw and basic data acquired in scientific activities (such as experiment, observation, detection, survey) or in other ways as well as the data sets processed and organized systematically according to different scientific activities' needs. Since 2003, the Chinese government has invested more than two billion RMB ($320 million) to the National Science & Technology Infrastructure Program. Currently, the program has achieved remarkable progress in the integration and sharing of scientific resources, involving large-scale scientific instruments, natural science and technology resources, scientific data and literature. In 2011, the Ministry of Science and Technology and the Ministry of Finance of the People's Republic of China jointly invited experts to assess the first batch of National Science & Technology Infrastructure. In total, 23 platforms obtained qualifications, among which seven focussed on scientific data sharing. The authors selected these seven platforms as the evaluation objects, along with the Basic Science Data Sharing Center (shown in Table I), and conducted an

No.	Name of platform	Construction unit of platform (competent department)	URL
1	National Scientific Data Sharing & Service Network on Material Environmental Corrosion (Aging)	University of Science and Technology Beijing (Ministry of Education)	www.ecorr.org/
2	Chinese Forestry Science Data Center	Chinese Academy of Forestry (State Forestry Administration)	www.cfsdc.org
3	Data Sharing Infrastructure of Earth System Science	Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences (Academy of Sciences)	www.geodata.cn/Portal/?isCookieChecked = true
4	National Scientific Data Sharing Platform for Population and Health	Chinese Academy of Medical Sciences (Ministry of Health)	www.ncmi.cn/1
5	National Agricultural Scientific Data Sharing Center	Agriculture Information Institute of Chinese Academy of Agriculture Science (Ministry of Agriculture)	www.agridata.cn/
6	China Earthquake Data Center	China Earthquake Networks Center (China Earthquake Administration)	http://data.earthquake.cn/index.html
7	China Meteorological Data Sharing Service System	National Meteorological Information Center (China Meteorological Administration)	http://cdc.cma.gov.cn/home.do
8	Basic Science Data Sharing Center	Chinese Academy of Sciences (Ministry of Science and Technology)	www.nsdc.cn/index.html

Table I. The list of eight scientific data sharing platforms empirical study on the performance evaluation of scientific data sharing platforms. The necessary data are obtained by visiting their web sites, consulting the webmasters and reviewing related literature.

2. Literature review

Since the launching of National Scientific Data Sharing Project, the construction of National Science &Technology Infrastructure has been pushing forward steadily, and the scientific data sharing service has gained significant improvement as well. In this case, the research and practice about the performance evaluation of scientific data sharing platforms have attracted a great deal of attention in China.

In 2011, the Ministry of Science and Technology and the Ministry of Finance of the People's Republic of China jointly issued Notice on Carrying out Assessment and Performance Evaluation of National Science &Technology Infrastructure, together with The Performance Evaluation Indicator for the Services of National Science &Technology Infrastructure. The latter focussed on the service quantity, the service efficiency and the sharing function of scientific platforms, as well as the feedback of users, enumerating four first-level indicators (service quantity, service efficiency, operation management and resource integration) and 12 second-level indicators.

The Evaluation Indicator System of Sharing Service, issued by the Basic Science Data Sharing Center, designed six first-level indicators and 20 second-level indicators. The six first-level indicators are listed as follows: construction ability of data resource, technological ability, organization and management, sharing service, effects and feature. Three first-level indicators, involving resource, management and service, and six second-level indicators were put forward for the performance evaluation of Shanghai Infrastructure for Public Research and Development of Scientific Data Sharing System. The Evaluation System Rules of Earthquake Data Sharing Program, which includes five first-level indicators and 27 second-level indicators, was proclaimed by the Department of Earthquake Monitoring and Prediction of China Earthquake Administration. The five first-level indicators refer to resource development, system construction, operation management, sharing performance and influence and future prospect. The Data Center of the National Scientific Data Sharing Platform for Population and Health also framed its evaluation indicator system, which includes three first-level indicators, eight second-level indicators and 16 third-level indicators. The three first-level indicators consist of data center resource, data center standard and data center web site. Dong *et al.* (2007) designed a performance evaluation indicator system for scientific data sharing platforms, which includes four first-level indicators and 18 second-level indicators. The four first-level indicators are institutional guarantee ability, scientific data, external service ability and comprehensive effectiveness. Li and Cui (2009) also designed a system called Quality Assessment and Indicator System Frame of the National Medical Scientific Data Sharing Project, which includes four first-level indicators and 12 second-level indicators. The four first-level indicators are resources construction quality, conformation with sharing standards, platform quality and project management level. Fan and Gong (2009) proposed a system including seven first-level indicators and 33 second-level indicators. The seven first-level indicators refer to system quality, information quality, service quality, data use, users' satisfaction, personal influence, organizational and social influence. Zhao *et al.* (2013) used "Visibility-Availability-Usability" theory assessment model to evaluate scientific and technology resources of different types and regions in China.

These researches mainly discussed the performance evaluation system of one particular platform, however they gave us some useful references to build our indicator system. We designed an indicator system including four first-level indicators, 11 second-level indicators and 26 third-level indicators, and used it to evaluate the performance of the eight scientific data sharing platforms in China.

3. Research methods

The authors used the analytic hierarchy process (AHP) to set indicator weights and the experts grading method to give scores for the indicators. The AHP was proposed by T.L. Saaty, an American operational research expert, in the early 1970s. It is a decision-making model that aids people in making decisions in our complex world. The three-step process includes identifying and organizing decision objectives, criteria, constraints and alternatives into a hierarchy; evaluating pair-wise comparisons between the relevant elements at each level of the hierarchy; and making synthesis by using the solution algorithm of the results of the pair-wise comparisons over all levels. In the end, the algorithm result demonstrates the relative importance of alternative courses of action. The expert scoring method is a quantitative method of qualitative description. Generally speaking, the method consists of three steps. First, based on the specific requirements of evaluation objects, the items which will be evaluated are selected. Second, the evaluation criteria are worked out. Finally, a number of representative experts will be invited to grade the items according to the evaluation criteria with their experience, after that the scores will be collected.

3.1 Construction of evaluation indicator system

Data resource, function, service effect and the users' satisfaction should be valued in the performance evaluation of a scientific data sharing platform. Based on the related research mentioned above, the authors put forward four first-level indicators for scientific data sharing platform performance evaluation, namely, operation management, data resource, platform function, service efficiency and influence. The index of operation management and data resource set target of two second-level indicators, the index of platform function sets target of three second-level indicators, the index of service efficiency and influence sets target of four second-level indicators.

Thus, our indicator system is made up of four first-level indicators, 11 second-level indicators and 26 third-level indicators (shown in Table II).

3.2 Evaluation method

The yaaph0.5.2, a piece of relatively mature AHP software, is utilized to draw hierarchical model structure chart, set judgment matrix and calculate expert group decision. The specific process of evaluation is as follows:

- First, setting the weight of first-level indicators and the weight of second-level indicators according to AHP, and calculating the comprehensive weight of second-level indicators.
- Second, grading third-level indicators. "10" is the highest point and then "9," "8," "7," … decreasing progressively to the lowest point "0" (shown in Table III).

First-level indicator	Second-level indicator	Third-level indicator
A Operation management	A1 Talent team	A11 Service team scale A12 Personnel quality
	A2 Management system	A21 Policy and regulation A22 Organization establishment
B Data resource	B1 Data product	B11 Data set quantity
	B2 Data description and organization	B21 Data description B22 Data organization and integration
C Platform function	C1 Retrieval and browse	C11 Retrieval mode C12 Browse mode C13 Retrieval result C14 Response speed
	C2 Data service	C21 Availability C22 Usability C23 Data updating
	C3 Platform service	C31 Data download C32 Data collection C33 Customized service C34 Effectiveness of link C35 Information release frequency
D Service efficiency and influence	D1 Research efficiency	D11 Influence on National Major Scientific Plan Project D12 Support of the scientific research achievement, patent and paper output
	D2 Citation	D21 Cited time
	D3 Service statistics	D31 Number of registered user D32 Page view D33 Linked time
	D4 Customer satisfaction	D41 Serviceobject satisfaction

Table II. Performance evaluation indicator system of the scientific data sharing platform

- Third, making summation of the scores of third-level indicators under every second-level indicator, then dividing the number of third-level indicators, so as to get the average score (represented as P_{ij}, shown in Table IV), which is the score of every second-level indicator.
- Finally, multiplying the score of every second-level indicator by its comprehensive weight to get the weighted value, then adding the weighted value of every second-level indicator to get the final score of every scientific data sharing platform's performance.

3.3 Evaluation step

The specific steps of using AHP to set indicator weights of scientific data sharing platform's performance evaluation is presented as follows:

- According to the judgment scale, judgment matrix was constructed for all first-level indicators and second-level indicators. Then, the authors invited experts to compare the indicators in the judgment matrix in pairs to judge their relative importance. The criteria scale we used is the nine relatively important scale proposed by Saaty (shown in Table V). The classical 1-9 scale is used to

Table III.
The scores of the third-level indicators of eight scientific data sharing platforms

Third-level indicator	Name of the platform							
	National Scientific Data Sharing & Service Network on Material Environmental Corrosion (Aging)	Chinese Forestry Science Data Center	Data Sharing Infrastructure of Earth System Science	National Scientific Data Sharing Platform for Population and Health	National Agricultural Scientific Data Sharing Center	China Earthquake Data Center	China Meteorological Data Sharing Service System	Basic Science Data Sharing Center
A11 Service team scale	5	10	10	7	5	6	6	8
A12 Personnel quality	6	8	10	6	5	6	6	9
A21 Policy and regulation	7	3	5	10	7	7	5	6
A22 Organization establishment	9	5	7	10	7	4	4	6
B11 Data set quantity	4	7	8	9	7	7	8	10
B21 Data description	6	10	10	9	6	9	4	7
B22 Data organization and integration	6	8	6	8	6	8	8	10
C11 Retrieval mode	6	8	10	4	8	8	10	6
C12 Browse mode	6	6	6	6	6	6	6	10
C13 Retrieval result	5	7	5	5	8	5	9	5
C14 Response speed	9	10	9	9	8	5	7	8
C21 Availability	6	6	4	4	4	10	6	6
C22 Usability	4	5	7	4	5	6	8	9
C23 Data updating	5	2	8	7	3	10	10	10
C31 Data download	5	7	7	5	9	10	7	7

(*continued*)

Table III.

Third-level indicator	Name of the platform: National Scientific Data Sharing & Service Network on Material Environmental Corrosion (Aging)	Chinese Forestry Science Data Center	Data Sharing Infrastructure of Earth System Science	National Scientific Data Sharing Platform for Population and Health	National Agricultural Scientific Data Sharing Center	China Earthquake Data Center	China Meteorological Data Sharing Service System	Basic Science Data Sharing Center
C32 Data collection	2	10	10	10	5	2	2	5
C33 Customized service	2	2	9	8	2	10	9	9
C34 Effectiveness of link	9	10	6	10	10	7	7	9
C35 Information release frequency	6	3	7	8	9	3	3	6
D11 Influence on National Major Scientific Plan Project	5	4	10	7	3	4	3	6
D12 Support of the scientific research achievement, patent and paper output	4	2	6	8	5	4	4	5
D21 Cited time	4	5	7	4	5	6	10	3
D31 Number of registered user	5	2	9	7	4	6	8	7
D32 Page view	3	1	8	3	2	4	7	10
D33 Linked time	2	0	4	5	0	6	1	9
D41 Service object satisfaction	7	6	8	7	6	8	8	9

Table IV.
The P_{ij} of the eight scientific data sharing platforms

First-level indicator	Second-level indicator	P_{ij} National Scientific Data Sharing & Service Network on Material Environmental Corrosion (Aging)	Chinese Forestry Science Data Center	Data Sharing Infrastructure of Earth System Science	National Scientific Data Sharing Platform for Population and Health	National Agricultural Scientific Data Sharing Center	China Earthquake Data Center	China Meteorological Data Sharing Service System	Basic Science Data Sharing Center
A Operation management	A1 Talent team	5.5	9	10	6.5	5	6	6	8.5
	A2 Management system	8	4	6	10	7	5.5	4.5	6
B Data resource	B1 Data product	4	7	8	9	7	7	8	10
	B2 Data description and organization	6	9	8	8.5	6	8.5	6	8.5
C Platform function	C1 Retrieval and browse	6.5	7.8	7.5	6	7.5	6	8	7.3
	C2 Data service	5	4.3	6.3	5	4	8.7	8	8.3
	C3 Platform service	4.8	6.4	7.8	8.2	7	6.4	5.6	7.2
D Service efficiency and influence	D1 Research efficiency	4.5	3	8	7.5	4	4	3.5	5.5
	D2 Citation	4	5	7	4	5	6	10	3
	D3 Service statistics	3.3	1	7	5	2	5.3	5.3	8.7
	D4 Customer satisfaction	7	6	8	7	6	8	8	9

Note: P_{ij} is accurate to one decimal place

Score	Intensity of importance	Explanation
1	Equal importance	Two activities contribute equally to the objective
3	Moderate importance	Experience and judgment slightly favor one activity over another
5	Strong importance	Experience and judgment strongly favor one activity over another
7	Very strong or demonstrated importance	An activity is favored very strongly over another; its dominance demonstrated in practice
9	Extreme importance	The evidence favoring one activity over another is of the highest possible order of affirmation
2, 4, 6, 8	Median between two adjacent judgments	

Table V. Table of scale

confirm the relative importance of indicators which distribute in different matrices by comparing them in pairs. It is a process that makes subjective thinking quantitative. If measurement indicator i scores ij when it compares with measurement indicator j, then j scores ji ($ji = 1/ij$) when it compares with i. For example, if measurement indicator A is strong importance when it compares with measurement indicator B, then A scores 5 when it compares with B and B scores 1/5 when it compares with A. The authors issued 11 questionnaires – Questionnaires about the Weight of Scientific Data Sharing Platform's Performance Evaluation – to relevant experts of scientific data sharing platforms, and acquired 11 effective questionnaires, with an effective return ratio of 100 percent.

- According to the result of questionnaires, we used yaaph0.5.2 to judge matrix and calculate group decision, getting the weight of first-level indicators (W_{1i}). The consistency scale of judgment matrix is 0.0093, which passes consistency test. Arranged in descending order of W_{1i} (shown in Table VI), the weight of the first-level indicators are, respectively, service efficiency and influence (0.3688), platform function (0.3264), data resource (0.1871) and operation management (0.1177).
- On the basis of the same principle, we can get the weight of second-level indicators (W_{2ij}). Then the comprehensive weight of second-level indicators (W_{ij}) was calculated based on W_{1i} and W_{2ij}.

The calculation formula was proposed: $W_{ij} = W_{2ij} \times W_{1i}$. Based on this formula, we got the comprehensive weight of second-level indicators (shown in Table VII). The top

First-level indicator	Operation management	Data resource	Platform function	Service efficiency and influence	W_{1i}
Operation management	1.0000	0.5503	0.3574	0.3684	0.1177
Data resource	1.8171	1.0000	0.4850	0.5246	0.1871
Platform function	2.7982	2.0617	1.0000	0.7418	0.3264
Service efficiency and influence	2.7144	1.9064	1.3480	1.0000	0.3688

Table VI. The weight of first-level indicators of scientific data sharing platform's performance evaluation

Table VII. The comprehensive weight of second-level indicators of scientific data sharing platform's performance evaluation

W_{1i}	W_{2ij}	W_{ij}
A Operation management (0.1177)	A1 Talent team (0.6182)	0.0728
	A2 Management system (0.3818)	0.0449
B Data resource (0.1871)	B1 Data product (0.6135)	0.1148
	B2 Data description and organization (0.3865)	0.0723
C Platform function (0.3264)	C1 Retrieval and browse (0.1441)	0.0470
	C2 Data service (0.4087)	0.1334
	C3 Platform service (0.4472)	*0.1460*
D Service efficiency and influence (0.3688)	D1 Research efficiency (0.4071)	*0.1501*
	D2 Citation (0.1246)	0.0460
	D3 Service statistics (0.1047)	0.0386
	D4 Customer satisfaction (0.3635)	*0.1341*

Note: W_{ij} is accurate to four decimal places

three of W_{ij} among the 11 second-level indicators are separately research efficiency (0.1501), platform service (0.1460) and customer satisfaction (0.1341).

4. Result and analysis

4.1 Calculation of the evaluation result

According to the evaluation method, we calculated the scores of performance evaluation of the platforms, including the score of every second-level indicator *Pij*, the comprehensive weight of each second-level indicator W_{ij} and we assumed the weighted value of score as F_{ij}, the summation of F_{ij} as F. Then two computational formulas of F_{ij} and F were proposed separately as follows:

$$F_{ij} = W_{ij} \times P_{ij} \tag{1}$$

$$F = \sum_{i=1}^{4} \sum_{j=1}^{j(i)} F_{ij} \tag{2}$$

where "4" stands for the number of first-level indicators, $j(i)$ stands for the number of second-level indicators under their first-level indicator.

The final calculation result is shown in Table VIII.

4.2 Analysis of the evaluation result

The evaluation result of first-level indicators is presented below.

Operation management. This indicator covers two items: talent team and management system. The W_{ij} of talent team and management system are 0.0728 and 0.0449, respectively (shown in Table VII).

(1) Talent team. Talent team is one of the key factors of scientific data sharing platform. There are two third-level indicators designed: service team scale and personnel quality. The F_{ij} of eight platforms range from 0.3 to 0.8 (shown in Table VIII). The F_{ij} of the Data Sharing Infrastructure of Earth System Science (0.7280) is the highest. More than 40 universities famous for geosciences in China, The World Data Center, The International Centre for Integrated Mountain Development, University of

Table VIII. Result of the performance evaluation of eight scientific data sharing platforms

First-level indicator	Second-level indicator	F_{ij}: National Scientific Data Sharing & Service Network on Material Environmental Corrosion (Aging)	Chinese Forestry Science Data Center	Data Sharing Infrastructure of Earth System Science	National Scientific Data Sharing Platform for Population and Health	National Agricultural Scientific Data Sharing Center	China Earthquake Data Center	China Meteorological Data Sharing Service System	Basic Science Data Sharing Center
A Operation management	A1 Talent team	0.4004	0.6552	0.7280	0.4732	0.3640	0.4368	0.4368	0.6188
	A2 Management system	0.3592	0.1796	0.2694	0.4490	0.3143	0.2470	0.2021	0.2694
B Data resource	B1 Data product	0.4592	0.8036	0.9184	1.0332	0.8036	0.8036	0.9184	1.1480
	B2 Data description and organization	0.4338	0.6507	0.5784	0.6146	0.4338	0.6146	0.4338	0.6146
C Platform function	C1 Retrieval and browse	0.3055	0.3643	0.3525	0.2820	0.3525	0.2820	0.3760	0.3408
	C2 Data service	0.6670	0.5781	0.8449	0.6670	0.5336	1.1561	1.0672	1.1117
	C3 Platform service	0.7008	0.9344	1.1388	1.1972	0.9928	0.9344	0.8176	1.0512
D Service efficiency and influence	D1 Research efficiency	0.6755	0.4503	1.2008	1.1258	0.6004	0.6004	0.5253	0.8256
	D2 Citation	0.1840	0.2300	0.3220	0.1840	0.2300	0.2760	0.4600	0.1380
	D3 Service statistics	0.1287	0.0386	0.2702	0.1930	0.0772	0.2059	0.2059	0.3345
	D4 Customer satisfaction	0.9387	0.8046	1.0728	0.9387	0.8046	1.0728	1.0728	1.2069
Result	*F* Weighted sum	5.2528	5.6894	7.6962	7.1577	5.5068	6.6296	6.5159	7.6595
	Hundred-mark system	52.528	56.894	76.962	71.577	55.068	66.296	65.159	76.595

Note: F_{ij} is accurate to four decimal places

Maryland in America and many other international organizations participate in its construction and provide the services of this project, which shows its high personnel quality, large scale and internationalized composition. Besides, the F_{ij} of the Chinese Forestry Science Data Center (0.6552) and the F_{ij} of the Basic Science Data Sharing Center (0.6188) are higher than the other platforms. The Chinese Forestry Science Data Center has 143 project personnel in total, including researchers, university faculty members and engineers. The Basic Science Data Sharing Center has 50 service specialists, providing real-time reference and form consultation services through e-mail and telephone. The F_{ij} of the National Agricultural Scientific Data Sharing Center (0.3640) is the lowest, indicating that the scale and quality of its service team should be improved in the near future.

(2) Management system. Management system ensures the efficient operation of scientific data sharing platform. There are two third-level indicators designed: policy and regulation, organization establishment. There is little difference among the F_{ij} of eight platforms, varying between 0.1 and 0.5 (shown in Table VIII). The F_{ij} of the National Scientific Data Sharing Platform for Population and Health (0.4732) is the highest because of its better system of policies and regulations, dealing with scientific data resources, supervision and acceptance, data archiving and management, information safety and so on. In addition, the leadership team, the specialist group, the working team and the project office have been founded, forming the organization and management system of its operation and service, which characterized by multi-sectoral cooperation and multi-layer management. The F_{ij} of the National Scientific Data Sharing & Service Network on Material Environmental Corrosion (Aging) and the F_{ij} of the National Agricultural Scientific Data Sharing Center are 0.3592 and 0.3143 respectively. The former sets up the management system of its network and experimental station as well as some responding departments, such as management department, standard department and data sharing and application department. The latter also formulates management measures for agricultural scientific data sharing and data collection, as well as establishes the leadership team, the advisory group, the technological group, the working team and the project office. The F_{ij} of the Chinese Forestry Science Data Center (0.1796) is the lowest, whose construction of policies and regulations should be strengthened. The F_{ij} of the rest four platforms are between 0.2 and 0.3.

Data resource. The index of data resource is mainly applied to investigate the data quantity, the data description and organization of eight platforms, including two second-level indicators: data product, data description and organization.

(1) Data product. There are four platforms (accounting for 50 percent) whose F_{ij} of data product exceed 0.9184 (the score of third-level indicators is more than eight); three platforms (accounting for 37.5 percent) whose F_{ij} of data product range from 0.6888 to 0.9184 (the score of third-level indicators ranges from 6 to 8); only one platform's F_{ij} is below 0.6888 (the score of third-level indicators is below 6) (shown in Table VIII). The F_{ij} of the Basic Science Data Sharing Center (1.1480) ranks first for its massive data resources with 448 databases/data sets, 104.86 TB in volume (investigated on August 17, 2014). The F_{ij} of the National Scientific Data Sharing & Service Network on Material Environmental Corrosion (Aging) (0.4592) ranks last with only 11 data sets. In general, the data resource construction of the platforms is satisfied except the National Scientific Data Sharing & Service Network on Material Environmental Corrosion (Aging) and the Chinese Forestry Science Data Center.

(2) Data description and organization. There are two third-level indicators designed: data description, data organization and integration. The indicator is mainly used to investigate if the metadata description, the metadata standard, the metadata column, the classification standard and the methods of data organization and integration could be provided on the web sites. There are five platforms (accounting for 62.5 percent) whose F_{ij} of data description and organization is above 0.5784 (the score of third-level indicators is more than 8); while the rest three platforms(accounting for 37.5 percent) whose F_{ij} of data description and organization are 0.4338 (the score of third-level indicators is 6, shown in Table VII). The F_{ij} of the Chinese Forestry Science Data Center (0.6507) is the highest with its metadata description, metadata standard, metadata column and classification standard are offered on the web site. Moreover, there are two modes of data access and acquisition under the metadata column (by browsing and by retrieval), but there is only one mode of data organization (by theme). The F_{ij} (0.4338) of the National Scientific Data Sharing & Service Network on Material Environmental Corrosion (Aging), the National Agricultural Scientific Data Sharing Center and the China Meteorological Data Sharing Service System tie for the last place, owing to the lack of standards of data description and organization.

Platform function. The index of platform function is mainly applied to investigate the data retrieval and service function of the platforms, including three second-level indicators, i.e. retrieval and browse, data service and platform service.

(1) Retrieval and browse. There are four third-level indicators designed: retrieval mode, browse mode, retrieval result and response speed. We investigated the number of retrieval mode and browse mode, the sort of retrieval result, the refine search and the response speed of eight platforms. The response speed is tested by the tool 114la (tested on August 21, 2014, 0:18-0:40 a.m.)[1]. The F_{ij} of the China Meteorological Data Sharing Service System (0.3760) ranks first (shown in Table VIII). It offers four kinds of retrieval modes including simple retrieval, data classification retrieval, subject retrieval and data positioning retrieval; the classification browse mode; and three ways of retrieval result sorting, including recently updated priority, primary node data priority and relevancy priority. And its response speed we tested is 0.144 seconds, which lies in the third place. The F_{ij} of both the National Scientific Data Sharing Platform for Population and Health and the China Earthquake Data Center are the same (0.2820), which is the lowest (shown in Table VIII). The National Scientific Data Sharing Platform for Population and Health does not offer any retrieval mode on its home page. It just provides different retrieval modes on each of its databases' interface.

(2) Data service. There are three third-level indicators designed: availability, usability and data updating, which are mainly used to investigate if it is convenient for users to download, browse data and utilize services and if the data updates frequently (investigated on September 1, 2014). There is obvious difference of F_{ij} among eight platforms (shown in Table VIII). The F_{ij} of the China Earthquake Data Center (1.1561) is the highest with better data availability. There is no requirement of login and register when browsing data or downloading partial data and metadata; higher data usability, well-classified data and rapid updating frequency (daily) also explains its high score. The F_{ij} of the National Agricultural Scientific Data Sharing Center (0.5336) is the lowest for its low data availability. The majority of data on the platform can be browsed only after login and register. Moreover, its updating frequency lies last but one of eight platforms.

(3) Platform service. There are five third-level indicators designed: data download, data collection, customized service, effectiveness of link and information release frequency. The W_{ij} of platform service (0.1460) lies in the second place among 11 second-level indicators (shown in Table VII). This indicator is primarily applied to investigate if the platforms offer data download service, data collection service and customized service, release information in time and whether the platforms' links are valid or not. The effectiveness of links is tested by the tool SEOWHY (tested on September 1, 2014)[2]. The F_{ij} of the National Scientific Data Sharing Platform for Population and Health (1.1972) is higher than others (shown in Table VIII), but there are still 19 data sets (accounting for nearly 7 percent) can be downloaded only by applying to data owners. The platform offers data collection service and customized service of some subject databases, its 83 links are all valid and 24 pieces of information were released within half a year (the second place of eight platforms). The F_{ij} of the National Scientific Data Sharing & Service Network on Material Environmental Corrosion (Aging) (0.7008) is the lowest (shown in Table VIII) for not delivering any data collection service or customized service, and only high priority users of this platform have the right to download its data.

Service efficiency and influence. This indicator is used to examine the scientific research efficiency and social utility of eight platforms' services. Its W_{1i} is the highest among four first-level indicators. Four second-level indicators are included, i.e. research efficiency, citation, service statistics and customer satisfaction.

(1) Research efficiency. The W_{ij} of research efficiency (0.1501) is in the first place among all second-level indicators, which indicates that the key goal of scientific data sharing platform is providing support for science and technology undertakings. To a great extent, the degree of research efficiency validates the operation efficiency and the value of platform. We examined the information issued by eight platforms to analyze their supporting effects to the national major scientific plan projects, scientific research achievements, patents and papers through web site survey (shown in Table IX).

Table IX. Research efficiency of eight scientific data sharing platforms

Scientific data sharing platform	Items of supporting the National Major Scientific Plan Project	Items of supporting the scientific research achievements, patent and paper output
National Scientific Data Sharing & Service Network on Material Environmental Corrosion (Aging)	51	75
Chinese Forestry Science Data Center	26	58
Data Sharing Infrastructure of Earth System Science	1,008	611
National Scientific Data Sharing Platform for Population and Health	12,143	286
National Agricultural Scientific Data Sharing Center	100	43
China Earthquake Data Center	87	63
China Meteorological Data Sharing Service System	79	39
Basic Science Data Sharing Center	102	204

The F_{ij} of the Data Sharing Infrastructure of Earth System Science (1.2008) and the F_{ij} of the National Scientific Data Sharing Platform for Population and Health (1.1258) rank first and second respectively (shown in Table VIII). The former makes an enormous contribution to the "973" Program (on June 4, 1997, the State Science and Education Steering Group decided to organize the implementation of "the National Program on Key Basic Research Project (973 Program)[3]," the purpose of the initiatives was to strengthen basic research in line with national strategic targets) and "863" Program (in 1986, to meet the global challenges of new technology revolution and competition, four Chinese scientists, Wang Daheng, Wang Ganchang, Yang Jiachi and Chen Fangyun, jointly proposed to accelerate China's high-tech development. With strategic vision and resolution, the late Chinese leader Mr Deng Xiaoping personally approved the National High-tech R&D Program, namely the 863 Program)[4], and other projects concerning environmental protection, railway construction, the manned space flight projects, climatological survey, civil engineering and so on. Until now, there are 611 published scientific papers and books, and 12 granted patents benefited from the support of the platform. The National Scientific Data Sharing Platform for Population and Health also offers vigorous support for the national major scientific plan projects, thus greatly promotes its service impact. Additionally, it develops the software for searching research projects and achievements, which brings users significant convenience.

(2) Citation. This indicator mainly aims to investigate the cited time of platforms' names. Considering that the construction of platforms experienced different stages, we chose different expressions such as "sharing pilot project," "sharing center," "sharing network," "sharing service network" and "sharing platform" as keywords to prevent from missing information. The above terms are all used and linked by the Boolean operator "OR" when performing full-text search. We retrieved these terms in the Chinese National Knowledge Infrastructure on September 1, 2014. The retrieval result is ordered by the number of citations and the top three are: the China Meteorological Data Sharing Service System (2010 results), the Data Sharing Infrastructure of Earth System Science (371 results) and the China Earthquake Data Center (191 results).

(3) Service statistics. It is clear that the number of registered users is one of the important indicators. The number of visits and linked times of platforms partly reflect the public attention of them and indicate their potential value. The number of registered users, visits and linked times of eight platforms were counted on March 23, 2013 (shown in Table X). The linked times is tested by the tool "chinaz[5]."

Scientific data sharing platform	Number of registered user	Visits	Linked times
National Scientific Data Sharing & Service Network on Material Environmental Corrosion (Aging)	5,983	1,122,170	6
Chinese Forestry Science Data Center	2,792	181,767	0
Data Sharing Infrastructure of Earth System Science	67,752	10,463,261	22
National Scientific Data Sharing Platform for Population and Health	50,657	1,351,435	24
National Agricultural Scientific Data Sharing Center	4,179	626,598	0
China Earthquake Data Center	6,152	1,563,159	26
China Meteorological Data Sharing Service System	61,965	10,356,001	1
Basic Science Data Sharing Center	50,839	21,812,482	63

Table X. Service statistics of eight scientific data sharing platforms

The F_{ij} of the Basic Science Data Sharing Center (0.3345) and the F_{ij} of the Data Sharing Infrastructure of Earth System Science (0.2702) are in the first and second place (shown in Table VIII). The Data Sharing Infrastructure of Earth System Science, the China Meteorological Data Sharing Service System and the Basic Science Data Sharing Centers enjoy higher user participation by number of registered users and visits. Besides, the linked times of the Basic Science Data Sharing Center (63) is the highest. However, the Chinese Forestry Science Data Center and the National Agricultural Scientific Data Sharing Center are not linked at present (shown in Table X).

(4) Customer satisfaction. The W_{ij} of customer satisfaction (0.1341) occupies the third place among all second-level indicators (shown in Table VII). We set five score criteria for this indicator: quite satisfied (10), satisfied (9), relative satisfied (8), general (6), rarely satisfied (5). There are four platforms whose F_{ij} are more than 1: the Basic Science Data Sharing Center (1.2069), the Data Sharing Infrastructure of Earth System Science (1.0728), the China Earthquake Data Center (1.0728) and the China Meteorological Data Sharing Service System (1.0728). The F of these four platforms are relatively high among eight platforms (shown in Table VIII).

Eventually, the performance of eight platforms is arranged in descending order by the value of F: Data Sharing Infrastructure of Earth System Science (76.962), Basic Science Data Sharing Center (76.595), National Scientific Data Sharing Platform for Population and Health (71.577), China Earthquake Data Center (66.296), China Meteorological Data Sharing Service System (65.159), the National Agricultural Scientific Data Sharing Center (55.068), Chinese Forestry Science Data Center (56.894) and National Scientific Data Sharing & Service Network on Material Environmental Corrosion (Aging) (52.528) (shown in Table VIII).

5. Conclusion and discussion

Through the investigation, it is concluded that there are some problems in the construction of scientific data sharing platforms in China:

(1) The relevant policies and regulations need to be improved. We found that only the National Scientific Data Sharing Platform for Population and Health has 14 laws and 12 regulations guiding its operation and management. Some of the rest seven platforms only have few policies and regulations to deal with data resource sharing and classification. For example, the National Scientific Data Sharing & Service Network on Material Environmental Corrosion (Aging) only formulates the management system of its station network, the Chinese Forestry Science Data Center even does not release any policies and regulations. Sensible policies and regulations should include the rules and regulations of data collection, quality control, operation, maintenance management, client management, safety and emergency measures, in order to offer guidance for the platform's operation during the whole data life cycle.

(2) The standards of data description and organization are deficient. Data are the resource foundation of sharing service of platform, and its standard construction is the basis of data exchange and sharing. However, the National Agricultural Scientific Data Sharing Center, the China Meteorological Data Sharing Service System do not have metadata description standard. Moreover, the National Scientific Data Sharing & Service Network on Material

Environmental Corrosion (Aging), the Data Sharing Infrastructure of Earth System Science, the National Agricultural Scientific Data Sharing Center do not have data classification standard. It is recommended that the standards of data description and organization should include metadata description standard and data classification standard, covering the basic principles and methods of standardization of the metadata, the XML/XSD markup rules of metadata, the data classification coding scheme of scientific data sharing project, the search protocols and extracting protocols of metadata and so forth.

(3) The data availability remains at a relatively low level. Among these platforms, only the users of the China Earthquake Data Center can browse its whole data and download partial data without registering. The users of the rest seven platforms can just browse partial data without registering. For instance, if users of the National Scientific Data Sharing & Service Network on Material Environmental Corrosion (Aging) want to browse the data of a certain topic, they are required to fill in an application form, waiting to be approved. The rules about data download are more rigid for some platforms, the download application must be submitted, and some of the data download services are only available for senior users or core users, while normal users have to register or even fill in application form and wait for approving.

(4) The services provided by these platforms are somewhat undiversified. The data download service, data collection service, customized service and other necessary services should be provided by the platforms. All of our investigated platforms provide data download service, five platforms provide customized service (accounting for 62.5 percent), and three platforms provide data collection service (accounting for 37.5 percent). The Data Sharing Infrastructure of Earth System Science and the National Scientific Data Sharing Platform for Population and Health provide all three kinds of these services, but its customized service limits to the data reservation service. From our survey, it can be seen that the services provided by the platforms are not varied, and the forms of service need to be diversified.

(5) The gaps in service efficiency between the platforms turn out to be wide. The W_{ij} of research efficiency (0.1501) is in the first place among all second-level indicators. Nevertheless, its F_{ij} is generally lower than the F_{ij} of platform service and customer satisfaction, whose W_{ij} are in the second place and third place respectively (shown in Table VIII), which shows that the research efficiency of all scientific data sharing platforms is relatively low. As a result, the influence on scientific research achievements, patents, papers and the support to national major scientific plan projects should be strengthened. As for F_{ij}, it can be seen that only two platforms' F_{ij} are more than 1, and the F_{ij} of the rest five platforms ranges between 0.4 and 0.7. The F_{ij} of the Data Sharing Infrastructure of Earth System Science (1.2008) is the highest, while the F_{ij} of the Chinese Forestry Science Data Center (0.4503) is the lowest and their difference reaches 0.7505. We believe that the construction of scientific data sharing platform is aimed at providing services for national scientific research projects and national major scientific plan projects. Every scientific data sharing platform plays an irreplaceable role on scientific research and the sustainable development of society and economy, as well as provides strong support for national scientific

research projects of all levels and fields. Therefore, the national scientific data sharing platforms in China are in urgent need of maximize the value of possessed scientific data and improve their research efficiency.

Our research discussed the performance evaluation system covering operation management, data resource, platform function, service efficiency and influence of eight scientific data sharing centers and made a comparative analysis. It reflected the reality development of scientific data sharing in China to some extent. However, due to the difficulties in collecting first-hand data, the data sources are limited to web sites and e-mails. We hope that we can carry an on-the-spot investigation to obtain more in-depth data in further studies.

Notes

1. http://tool.114la.com/sitespeed
2. http://tool.seowhy.com/linkdetection/
3. www.most.gov.cn/eng/programmes1/200610/t20061009_36223.htm
4. www.most.gov.cn/eng/programmes1/200610/t20061009_36225.htm
5. http://tool.chinaz.com/

Reference

Dong, C., Zhao, W. and Tu, Y. (2007), "Research on scientific data sharing organization performance evaluation", *Forum on Science and Technology in China*, No. 8, pp. 74-78.

Fan, X.E. and Gong, Q.Y. (2009), "Discussion on assessment of scientific data sharing system", *Chinese Journal of Public Health Engineering*, Vol. 8 No. 3, pp. 179-180.

Li, H.Y. and Cui, M. (2009), "Quality assessment and indicator system frame of the national medical scientific data sharing project", *World Science and Technology/Modernization of Traditional Chinese Medicine*, Vol. 11 No. 11, pp. 578-581.

Zhao, W., Peng, J., Qu B.Q., Dong C., Song, L.R. and Liu, R.D. (2013), "Comparative analysis of open and sharing evaluation of China's scientific and technical resources information on internet", *Digital Library Forum*, No. 12, pp. 58-63.

Further reading

Alexander, M. (2010), "Decision-making using the analytic hierarchy process (AHP) and SAS/IML", available at: http://analytics.ncsu.edu/sesug/2012/SD-04.pdf (accessed September 21, 2014).

Basic Science Data Sharing Center. (2009), "The evaluation indicator system of sharing service", available at: www.docin.com/p-347476574.html (accessed September 21, 2014).

Li, J., Liu, D.H. and Jiang, H. (2009), "Research on international scientific data sharing", *Library Development*, No. 2, pp. 19-21, 25.

Lu, X.C. (2010), *Activation of the Source of Innovation Achievements Entrepreneurial Dreams: Shanghai R & D Public Service Platform Construction Record*, Chemical Industry Press, Beijing.

Saaty, T.L. (1988), "What is the analytic hierarchy process?", *Mathematical Models for Decision Support*, Vol. 48 No. 1, pp. 109-121.

Saaty, T.L. (2008), "Decision making with the analytic hierarchy process", *Int. J. Services Sciences*, Vol. 1 No. 1, pp. 83-98.

The Data Center of National Scientific Data Sharing Platform for Population and Health (2009), "Evaluation indicator system of national medical scientific data sharing network", available at: www.ncmi.cn/UploadFile/f/a/0d09247519c164d410cfbd94265850af.pdf. (accessed August 28, 2014).

The Ministry of Science and Technology (2011), "The operation service performance evaluation indicator of national science & technology infrastructure", available at: www.most.gov.cn/fggw/zfwj/zfwj2011/201108/W020110805329242187033.pdf (accessed September 21, 2014).

Corresponding author
Professor Li Si can be contacted at: lsiwhu@163.com