DSI DATA SCIENCE AND INFORMETRICS
VOL.1, NO.4, NOV. 2021

Constructing data literacy evaluation system for social science scholars

Lin Wang[a], Xinyu Huang[b], Tianyu Du[c*]

a. Academy of Science and Education Evaluation, Hangzhou Dianzi University, Hangzhou, China
b. Tongliang District People´s Count, Chongqing, China
c. School of Management, Tianjin Normal University, Tianjin, China

ABSTRACT

To establish a data literacy evaluation system for social science scholars is a part of transformation to data-intensive scientific research paradigm in social science. Based on the literature review of data literacy and survey of social science data management features, this paper analyzed the elements of data literacy of social science scholars. The data literacy of social science scholars mainly consists of data awareness level, data discovery and access ability, data management and organization ability, data processing and analysis ability, data utilization and preservation ability, and data ethics level. Each of these primary indexes has several secondary indexes. It constructed the evaluation system of social science scholars data literacy. The weights of the primary and secondary indexes in the system were calculated by applying AHP. The data literacy evaluation system for social science scholars can provide a reference for assessing and promoting social science scholars' data literacy ability in China.

KEYWORDS

Data literacy; Social science; Scholars; Evaluation index system

1 Introduction

As the rise of data science and data-intensive scientific research paradigm vigorously promotes the development of science (Hanson et al., 2011), researchers are facing the challenge of big data management, such as data organizing and publishing. Data literacy has become an indispensable competence for scientific research and scholarly communication (Meng & Li, 2014). The data literacy of social science scholars is of great concern as well (Peng, 2013). Quantitative research draws more and more attention in social science (Chen & Wu, 2012). Taking economics, a discipline preferring quantitative research, as an example, an increasing number of studies employ data-intensive empirical methods. If the massive data management problems involved in economics research were to be solved effectively, the discipline would make new progress. In general, a huge amount of data resources has been accumulated in social sciences. Making it play a potential value and promoting the developments of disciplines has become a new call. Basically, realizing this goal depends on the improvement of the data literacy of social science scholars. To establish a data literacy evaluation system

* Corresponding author: dutianyu9738@163.com

for social science scholars is a part of transformation to data-intensive scientific research paradigm in social science, which is conducive to social science scholars to improve their data literacy and optimize the scientific research mode. Based on the literature review of data literacy and survey of social science data management features, this paper analyzes the elements of data literacy of social science scholars. It constructs the evaluation system of social science scholars' data literacy.

2 Literature Review

2.1 The Concept of Data Literacy

Generally speaking, data literacy is defined in terms of skills and knowledge (Chen & Qian, 2016). Mandinach et al. (2015) defined data literacy as the ability to collect, analyze and interpret various types of data and transform them into instructional knowledge and practice. Vahey et al. (2012) believed that data literacy is the ability to obtain, analyze, represent, evaluate and interpret data by using appropriate methods and tools. Zhang (2013) proposed that data literacy mainly refers to researchers' ability to collect, process, and analyze data, as well as the conduct code and ethics involved in data production, management, and release. Zhang (2015) regarded data literacy as the data cognition, management, and operation ability, which has dialectical, scientific, and accurate features and conforms to social morality and ethics. Moreover, data awareness is also an important element of data literacy. Finze and Parvate (2012) pointed out that data literacy includes not only charts making and interpretation and statistical tools application, but also the thinking mode cultivation, which requires the subjects to consciously observe, understand and use data. Qin et al. believed that data literacy refers to the scholars' awareness, knowledge, and ability to collect, process, manage, evaluate, and utilize research data. Although the definitions are different to some degree, they take the user capability involved in the data life-cycle process as the core of data literacy.

2.2 Data Literacy and Information Literacy

There have been not a few studies of the relationship between data literacy and information literacy in the academic community. Ridsdale et al. (2015) proposed that data literacy and information literacy overlap. Stephenson and Caravello (2007) and Victor et al. (2013) pointed out that there are similarities between data literacy and information literacy in data or information acquisition, management, analysis, and evaluation; data literacy is a special form of information literacy in practice. Stephenson and Caravello (2007) also emphasized that data literacy is a component of information literacy in social sciences. It can be seen that data literacy is an extension of information literacy. Hunt (2004) proposed that the data literacy concept is different from the traditional information literacy concept. The essential difference between data literacy and information literacy lies in that data literacy primarily involves processing complex data rather than other types of information (Thompson & Edelstein, 2004). Huang and Li (2016) thought that information literacy focuses on information search and use, while data literacy stresses data production and preservation. To sum up, the concept of data literacy is supposed to cover three aspects: Data awareness, subjects' cognition and response to data. Data competence, a series of skills such as data collection, processing, management, analysis, utilization, and preservation. Data ethics, the moral norms of subjects as far as data is concerned.

2.3 Data Literacy Evaluation

By referring to information literacy capability criteria, the data literacy capability framework and evaluation system are constructed according to the features of data literacy. With a scientific data literacy evaluation system, the data literacy level of social science scholars can be objectively and accurately assessed.

Based on the literature review of data literacy, Shen (2015) established an index system of data literacy with three dimensions: data culture, data awareness, and data skills. Long (2015) summarized the data literacy capability categories and constructed the evaluation system of data literacy capability, including six dimensions, such as data awareness and data exchange capability. Based on data management and service experience, Ridsdale et al. (2015) listed the data literacy capabilities required by researchers in the data management process and created a data literacy capability matrix with five dimensions. Carlson et al. (2011) built a data information literacy framework by observing students' behaviors, interviewing teachers, and referencing the information literacy capability standard of Association of College & Research Libraries (ACRL). It mainly includes 12 dimensions like metadata, data management and reuse, and data analysis. On the basis of the Big 6 model and 7 Pillars model of information literacy, Schneider (2013) put forward a theoretical framework of data literacy, including eight primary indexes and 28 secondary indexes. The eight primary indexes are data identification, data processing, data preservation, data protection, data evaluation, data management, and data exchange. Calzada (2013) reviewed the contents related to data management in the information literacy competency standards and integrated them with the research on scientific data management service in libraries to create the core competence framework of data literacy, including data awareness, data management, and other indexes.

Together, these above studies of the data literacy evaluation system mainly focus on the refinement of data capability; and the main objects of the evaluation system are teachers and students. We systematically analyze the established data literacy evaluation systems and lists 21 essential indexes, as shown in Table 1. It will provide a reference for constructing a data literacy evaluation system of social science scholars. Among the 21 indexes, "data awareness" and "data processing and analysis" appear most frequently (six times for each), which indicates that these two indexes are indispensable elements in the data literacy evaluation system. The "data assessment" and "data ethics and morality" follow consequently, both of which appear five times. This implies that these two indexes are also critical to the evaluation system. "data management and organization" and "data discovery and access" appear four times, indicating that they are also of high importance. Moreover, "data type", "data format" and "data utilization and reuse" only appear one time. Besides, the frequencies of the other 12 indexes in the Table 1 vary from 2 to 3 times, meaning that these indexes matter to some extent in the data literacy evaluation system.

Table 1 The indexes of data literacy evaluation system

	Carlson (2011)	Calzada (2013)	Schneider (2013)	Long (2015)	Shen (2015)	Ridsdale (2016)
Data awareness	√	√	√	√	√	√
Data collection					√	√
Data conversion	√				√	√
Data management	√	√	√			√

	Carlson (2011)	Calzada (2013)	Schneider (2013)	Long (2015)	Shen (2015)	Ridsdale (2016)
Data preservation	√		√			
Data analysis	√	√	√	√	√	√
Data quality controlling	√					√
Data visualization	√					√
Data type					√	
Data format	√				√	
Data discovery and access	√	√		√	√	
Metadata	√					√
Data ethics	√	√		√	√	√
Data assessment		√	√	√	√	√
Data utilization and reuse	√				√	
Data sharing					√	√
Data long–term preservation						√
Data security			√		√	√
Data interpretation	√	√			√	
Data citation	√					√
Data exchange			√	√		√

3 The Components and Characteristics of Data Literacy of Social Science Scholars

3.1 Social Science Data and Its Management

Compared with STEM disciplines, social science shows greater pervasiveness, complexity, historicity, and relativity (Jiang & Yang, 2015). American Social Science Data Management Alliance (Data-Pass) claims that social science data mainly include national and government statistical data, social survey data, historical and geographical activity data, etc. (Peng, 2014). Xu and Feng (2018) pointed out that social science data in Chinese universities are mostly concentrated in economics and relevant disciplines, and it mainly manifests as comprehensive and census social data. Such research data aims to provide evidence for government policy-making and is mainly distributed in scientific research teams. Shen (2015) noted that social science data includes not only open data on the Internet, but also experimental data, documentary data and image data.

Social science data management is closely related to data management subjects. Researchers' data management behavior is embodied in a series of activities such as data searching, data using, data publishing, data sharing, and data ethics. He and Chang (2014) held the view that the willingness of researchers to publish data is a crucial activity. Tenopir et al. (2011) stated that the biggest problem of data sharing in scientific research is lacking sharing culture. King (2011) argued that most social science data is valuable and sensitive, so it is not easy for external organizations to access it. In addition, some sensitive data are deleted after relevant studies are completed, which leads the data cannot be copied and reused. Therefore, more scholars ought to regard it as their responsibility to deposit the original and copied data in public archives (such as data archives). In terms of data utiliza-

tion, the emerging issues of data ethics need academia's attention (Chen et al., 2015).

There are differences in data management among different disciplines of social sciences. For example, scholars of economics and sociology show strong data awareness and data needs. They rely on multiple channels to access data. They master the primary descriptive statistical analysis methods and apply regression analysis and correlation analysis, and any other advanced data analysis methods in their research. It shows that these scholars generally have solid capabilities in data processing and analysis. Scholars of these disciplines attach great importance to data and fully utilize data in academic research. They continuously improve their capabilities to acquire new knowledge from data (Shen, 2015). In this situation, social science data management requires highly scientific.

3.2 The Components of Social Science Scholars Data Literacy

3.2.1 The Dimensions of Social Science Scholars Data Literacy

According to the literature review and indexes summary shown in Table 2 and the above analysis, we maintain that the dimensions of data literacy for social science scholars mainly consist of data awareness, data discovery and access, data management and organization, data processing and analysis, data utilization and preservation, and data ethics. They can be used as primary indexes of the data literacy evaluation system for social science scholars.

Table 2 Primary indexes of data literacy

Publication	Primary Index
(Huang, 2016)	Data awareness, data competence, data morality
(Hao, 2016)	Data awareness, data attitude, data knowledge, data competence, data morality and ethics
(Shen, 2015)	Data awareness, data skill, data culture
(Long, 2015)	Data awareness, data assessment, data assess, data analysis and exchange, data morality
(Wu, 2015)	Data foundation, data discovery and access, data understanding and assessment, data management, data utilization
(Zhang, 2013)	Data sensitivity, data application and thinking
(Schneider, 2013)	Data management, data analysis, data evaluation, data protection
(Carlson, 2011)	Data discovery, Database warehouse, data management, data transformation, quality guarantee, data analysis

3.2.2 Refinement of Data Literacy of Social Science Scholars

We summarized the concrete contents of data literacy components of social science scholars to obtain a more comprehensive and multilevel index system of data literacy. Each dimension will be analyzed in detail, respectively, as follow.

Data awareness level refers to understanding data more accurately, improving data insight and sensitivity, and expressing data needs. It is mainly manifested in the following aspects: understanding the value of data amidst research projects, treating and using data generated in the process of scientific research with a rigorous and serious attitude (Long, 2015); finding out the data needed to deal with the current problems from a large amount of data; recognizing the significance and value of various kinds of data for different research projects under different situations (Cheng, 2017); being sensitive to data (Robbins & Robbins, 2016);

identifying the value of the source data and determining when the data is needed (Calzada & Marzal, 2013); recognizing the potential value of research data, show a high-level data sensitivity, deciding the specific range of data required by research activities and clearly express data needs (Shen, 2015). In summary, data awareness includes data value perception, data sensitivity, and data need awareness.

Data discovery and access ability refer to the ability to skillfully retrieve and obtain all kinds of data, accurately interpret the format, type, characteristics of these data, and assess data quality. It mainly includes: mastering data retrieval skills, accurately assessing data sources and data quality according to demands, and then identifying problems and errors during assessment (Ridsdale et al., 2015); understanding the attributes of various data sources, and using data sets to find the data needed in scientific research at hands (Shen, 2015); obtaining relevant data from a variety of channels and having a good understanding of data sources, such as different kinds of databases or information channels, conducting various types of data collection and retrieval, accurately interpret the acquired data, auditing the data and eliminating incorrect or invalid data (Long, 2015); being able to employ data warehouse and accurately assess data quality obtained through other channels accurately, Moreover, importing and transforming data (Carlson et al., 2011). To sum up, data discovery and access ability involve data acquisition, data retrieval, data quality assessment, and data formatting.

Data management and organization ability refer to the ability to fully mine the data relationships, create metadata schemes, and select appropriate metadata schemes to describe data sets. It includes: find out the correlation between data (Shen, 2015); creating metadata scheme and constructing reasonable metadata sets (Ridsdale et al., 2015); classifying data with the help of metadata to use them (Shen, 2015); being capable of extracting and summarizing data (Dai & Li, 2018). Thus, data management and organizational ability include data association, data extraction, and metadata skills.

Data processing and analysis ability refer to the ability to use software to process and analyze raw data, visualize data, and compare old and new data. It includes technology and tools of data visualization (Hu & Wu, 2016); analyzing and processing the raw data with the help of software, for instance, counting and analyzing data by SPSS software (Long, 2015); being familiar with the technology and application tools for data processing in a specific field, and being informed of how to use data and perform repetitive data analysis task through scientific and reasonable work-flow (Carlson et al., 2011); understanding various kinds of data representation tools, such as PowerPoint (Shen, 2015); comparing old and new data effectively (Dai & Li, 2018). In summary, data processing and analysis ability mainly include employing data analysis software, data comparison, and data modeling.

Data utilization and preservation ability refer to the ability to preserve, publish and share data, and use data to serve for decision-making. It includes long and short-term data preservation, data publishing, data sharing, and licensing agreements (Hu & Wu, 2016); using data for decision making (Shen, 2015; Long, 2015) and writing papers or reports by utilizing data and data analysis results. It can be seen that data utilization and preservation ability mainly includes data application, data exchange, and data preservation.

Data ethics level refers to understanding laws and regulations related to data issues, accessing and using data reasonably and legally, and complying with relevant social norms. It includes: learning-related data ethical and legal issues, and abiding by data ethics in research (Ridsdale et al., 2015); being responsible for the data used in the paper, understanding the ethical and moral issues involved in each step of empirical research, i.e., from data collection to data sharing, respecting the data research results of others, and citing data sources when

using (Long, 2015); being knowledgeable about data privacy and intellectual property rights as well as the relevant precautions for data sharing and preservation, learning to quote data and acknowledging the author of original data, and avoiding data misuse and misrepresentation (Carlson et al., 2011). It is found that the data ethics level mainly includes data legal level and data moral level.

The components of data literacy of science scholars are summarized in Table 3. In order to verify the validity of the listed data literacy components, we quantified the occurrence frequency of each data literacy ability in the retrieved literature set. We combined literacy indexes that share similar or same attributes along each dimension in the literature and counted the occurrence times. In this process, the evaluation dimensions of data literacy are taken as statistical units. The results show that the six constituent dimensions of social science scholar data literacy are the literacy or competencies mentioned in most literature, which indicates that the academic circle has accepted these as the essential competencies of data literacy.

Table 3 The components of data literacy

Data Literacy Dimensions	Representative Literature	Occurrence Frequency	Refined components
Data awareness level	Huang (2016) Hao (2016) Shen (2015) Long (2015) Zhang (2013)	66	Data sensitivity
			Data value perception
			Data need awareness
Data discovery and access ability	Shen (2015) Long (2015) Wu (2015) Zhang (2013) Carlson (2011)	108	Data acquisition
			Data retrieval
			Data formatting
			Data quality assessment
Data management and organization ability	Ridsdale (2016) Shen (2015) Wu (2015) Schneider (2013) Carlson (2011)	94	Data association
			Data extraction
			Metadata skills
Data processing and analysis ability	Ridsdale (2016) Shen (2015) Long (2015) Schneider (2013) Carlson (2011)	128	Data analysis software
			Data modeling
			Data comparison
Data utilization and preservation ability	Hao (2016) Shen (2015) Long (2015) Wu (2015) Zhang (2013)	116	Data application
			Data exchange
			Data preservation
Data ethics level	Ridsdale (2016) Huang (2016) Hao (2016) Long (2015) Carlson (2011)	70	Data legal level
			Data moral level

4 Designing Data Literacy Evaluation System for Social Science Scholars

4.1 Preliminary Construction of the Evaluation System

According to the above analysis, social science scholars' data literacy evaluation system includes six primary indexes and eighteen secondary indexes, as shown in Table 4.

Table 4 The primary data literacy evaluation system for social science scholars

Primary Index	Secondary Index
Data awareness level	Data sensitivity
	Data value perception
	Data need awareness
Data discovery and assess ability	Data acquisition
	Data retrieval
	Data formatting
	Data quality assessment
Data management and organization ability	Data association
	Data extraction
	Metadata skills
Data processing and analysis ability	Data analysis software
	Data modeling
	Data comparison
Data utilization and preservation ability	Data application
	Data exchange
	Data preservation
Data ethics level	Data legal level
	Data moral level

4.2 Further Revision of the Evaluation System

In order to ensure the effectiveness of the data literacy evaluation system for social science scholars and optimize it, we consulted experts for suggestions on the primary data literacy evaluation system for social science scholars by questionnaire survey. We then revised the evaluation system according to the survey results. We mainly sought experts' advice on the rationality of data literacy classification, omissions and overlapping of some indexes, the wording appropriateness of the indexes, and the importance of the indexes.

The experts surveyed in this study, a total of eight, were mainly university professors and librarians in social sciences. Those university professors whose serial numbers began with "U" in this study were mainly from Tianjin Normal University and Southwest University of Political Science and Law. Those university librarians of social sciences whose serial numbers begin with "L" in this study are mainly from Nankai University Library and Tianjin Normal University Library. All professors in the survey got master's degrees or above and engaged in

data-intensive academic research. All social science librarians in the survey had rich experience in data management, service, and data literacy training.

The questionnaire is a semi-open questionnaire. In the questionnaire, experts were asked to rank the importance of the indexes (primary and second-level) and suggest the addition, deletion, and modification of indexes. The 5-level Likert Scale was adopted in the survey.

The statistical results of the questionnaire survey show that all experts considered the primary and second-level indexes to be "all-important" or "important", except that 12.5% of them considered the data need awareness to be "averagely important". More than 80% of the indexes score over 4 points, indicating that most experts believed that all the data literacy evaluation indexes of social science scholars in this study are significant.

Based on the expert's opinions, we revised the indexes as follows.

Experts U1 pointed out that data identification should be added under the category of data awareness level as a secondary index. It was described as "the ability to distinguish useful information, filtering useless information". However, it overlaps with the "data value perception" index under the same category. In addition, there exists the secondary index of data quality assessment under the category of data discovery and access ability, where the part of this index is suitable. Therefore, no revision was made.

Expert U2 argued that the secondary index of data formating under the category of data discovery and access ability was not correctly expressed, and it ought to be changed to "acquiring of data in various formats", and the description of the index could be "the ability to effectively acquire data of various formats and transform data formats". It was adopted.

Expert U3 noted that the data need awareness index, referring to "the ability to clarify the data required by of research problem," was inaccurate. It should be revised to "the ability to clarify the target data category of topics being studied, and express data demands explicitly". It was adopted.

Expert U4 suggested that "the ability to distinguish how to preserve different types of data" could be restated as "the ability to preserve different types of data under different categories". It was adopted.

Expert U5 pointed out that the expression "understand" in data moral level index was not quantified or operable enough, so "the ability to understand the moral and ethical issues involved in data collection, utilization and sharing" should be changed to "the ability to list the common moral and ethical issues involved in data collection, utilization, and sharing". It was adopted.

Expert L1 maintained that the description of the secondary index of metadata skills under the category of data management and organization was incomplete, so the "metadata skills" could be modified to "metadata creation", and "the ability to create metadata schemes" ought to be added in the description. It was adopted.

Expert L2 claimed that the description of the secondary index of data exchange under the category of data utilization and preservation capability was incomplete. The data exchange and sharing should include the research results and the data itself. It was adopted.

Expert L3 thought that the time limit of data preservation should be made clear in the concept of data preservation. It was adopted.

Table 5 shows the revised evaluation index system and its description of data literacy for social science scholars.

Table 5 Revised data literacy evaluation system for social science scholars

Primary Index	Secondary Index	Index Description
Data awareness level	Data sensitivity	High sensitivity to important research data
	Data value perception level	Ability to recognize the data value
	Data need awareness	Ability to clarify the target data category of topics being studied and express data needs explicitly
Data discovery and access ability	Data acquisition	Access to data from multiple channels
	Data retrieval	Ability to retrieve data by multiple methods
	Acquiring of data in various formats	Ability to effectively acquire data of various formats and transform data formats
	Data quality assessment	Ability to assess the accuracy, timeliness and legitimacy of data
Data management and organization ability	Data association	Ability to discover and demonstrate relationships between data
	Data extraction	Ability to extract and summarize data
	Metadata creation skills	Ability to create metadata schemes, understand the basic principles of metadata and interpret and describe metadata
Data processing and analysis ability	Data analysis software	Ability to process and analyze raw data with software
	Data modeling	Ability to analyze and represent data, and conduct data modeling with visualization tools
	Data comparison	Ability to compare data analysis results with other research data sets and write data analysis reports
Data utilization and preservation ability	Data application	Ability to apply data analysis results to decision making and solve research problems
	Data exchange	Ability to exchange and share data–based research results and data sets with people in various academic situations
	Data preservation	Ability to preserve various types of data under different categories to realize the data value and facilitate data reuse, and clarify the preservation time limit
Data ethics level	Data legal level	Ability to use laws and regulations to ensure the safety and legality of data
	Data moral level	Ability to list the common moral and ethical issues involved in data collection, utilization, sharing, acknowledging others´ research data and citing data sources when using them

4.3 Index Weight Calculation with Analytic Hierarchy Process

4.3.1 Questionnaire Design and Distribution

After analyzing the common methods used to calculate the weights of the indexes, this study chose the analytic hierarchy process (AHP). Huang et al. (2019) applied AHP to Calculate the weights of first and second-level indexes in the evaluation system of college students' information literacy in new information environment. Experts determined the weights of the primary and secondary indexes through a questionnaire survey. We still chose the same group of experts who received our interview before and asked them to finish the survey. These experts whose serial numbers begin with "Q" in the following paragraphs were familiar with the indexes of data literacy so that they could make reasonable and accurate rankings.

4.3.2 Primary Index Weights Calculation

The calculation of index weights by AHP is mainly divided into five steps: establishing a hierarchical structure, constructing a judgment matrix, calculating the index weights, testing the sorting consistency of a single level, and determining the weight value. In this paper, an original matrix was selected, and the normalized column averaging method was adopted to illustrate the calculation process.

The data awareness level, data discovery and access ability, data management and organization ability, data processing and analysis ability, data utilization and preservation ability, data ethics level were presented by A, B, C, D, E, F, respectively. The judgment matrix of expert Q1 is presented in Table 6.

Table 6 Expert Q1 judgment matrix

	A	B	C	D	E	F
A	1	1	1	1	1	5
B	1	1	1	1	3	5
C	1	1	1	1/3	1	5
D	1	1	3	1	1	5
E	1	1/3	1	1	1	5
F	1/5	1/5	1/5	1/5	1/5	1

It can be denoted as :

$$E=\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 5 \\ 1 & 1 & 1 & 1 & 3 & 5 \\ 1 & 1 & 1 & 1/3 & 1 & 5 \\ 1 & 1 & 3 & 1 & 1 & 5 \\ 1 & 1/3 & 1 & 1 & 1 & 5 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 & 1 \end{pmatrix}$$

Calculation steps were as follows:

We transformed the matrix E to the normalized matrix E- through normalization processing.

$$E- = \begin{pmatrix} 0.1923 & 0.2206 & 0.1389 & 0.2206 & 0.1389 & 0.1923 \\ 0.1923 & 0.2206 & 0.1389 & 0.2206 & 0.4167 & 0.1923 \\ 0.1923 & 0.2206 & 0.1389 & 0.0735 & 0.1389 & 0.1923 \\ 0.1923 & 0.2206 & 0.4167 & 0.2206 & 0.1389 & 0.1923 \\ 0.1923 & 0.0735 & 0.1389 & 0.2206 & 0.1389 & 0.1923 \\ 0.0385 & 0.0441 & 0.0278 & 0.0441 & 0.0278 & 0.0385 \end{pmatrix}$$

Normalized Weight "W" was calculated based on the above matrix:

W= [0.1839 0.2302 0.1594 0.2302 0.1594 0.0368]

As for testing Consistency Ratio (CR) of a matrix, when CR ≤ 0.1, it is considered acceptable, which implies that the weights have been reasonably distributed. Otherwise, the judgment matrix needs to be reconstructed until it meets the criterion. The test formula is: CR=CI/RI. CI is Consistency Index, and its calculation formula is : $CI= \frac{入 max-n}{n-1}$, 入 max represents the principal eigenvalue, and "n" represents the size of the matrix. RI represents *Random Index*. Referring to Table 7 for more details:

Table 7 RI standard indexes

N	1	2	3	4	5	6	7	8	9
RI	0	0	0.58	0.90	1.12	1.24	1.32	1.41	1.45

The value of the 入 max of matrix E is 6.2790, and hence the value of CI is 0.0558. The CR value can be calculated by the formula CR = CI / RI, and it comes out of 0.045 (lower than 0.10). Therefore, the judgment matrix is effective, and the weights are reasonably distributed. We continued to calculate the weights and CR values of other experts' judgment matrices. The items with CR above 0.1 need to be eliminated in AHP. Then the mean value ought to be calculated for the remaining data, from which the final weights of the primary indexes are obtained. We found that the CR value of the expert Q8 matrix is 0.148, which exceeds 0.10. Thus the corresponding data is excluded.

We also applied AHP to calculate the weights and CR values of secondary indexes as well. We found that the CR value of the expert Q5 matrix for data awareness was 0.254, and that of the expert Q3 matrix for data discovery and access ability was 0.254. Both of them were more than 0.10. Thus the corresponding data is eliminated.

4.3.3 Total Hierarchy Permutation

In summary, after removing Q8, Q3 and Q5, the mean value of the primary index expert judgment matrix is W1=0.1913, W2=0.1994, W3=0.1547, W4=0.2073, W5=0.1502, W6=0.0970 respectively, that is, weights of the primary indexes of data awareness level, data discovery and access ability, data management and organization ability, data processing and analysis ability, data utilization and preservation ability, data ethics level are 0.1820, 0.1945, 0.1578, 0.2058, 0.155, 0.1005, respectively. The average weights of expert matrix for secondary indexes under the primary index of data awareness level are W1=0.3911,W2=0.3793,W3=0.2296 respectively; the corresponding weights of data discovery and access ability are W1=0.3126, W2=0.3162, W3=0.0817, W4=0.2858; for data management and organization ability, W1=0.2685, W2=0.4057, W3=0.3257; for data processing and analysis ability, W1=0.3940, W2=0.2876, W3=0.3183; for data utilization and preservation ability, W1=0.4989,W2=0.1923,W3=0.3088; for data ethics level, W1=0.5000,W2=0.5000.

The consistency test of overall hierarchical sequencing was carried out again, and it turned out that CR = 0.005, less than 0.10. The consistency test result proves that the weight distribution was reasonable.

4.4 The data Literacy Evaluation System for Social Science Scholars

We established the data literacy evaluation system for social science scholars through the above analysis, as shown in Table 8.

Table 8 Data literacy evaluation system for social science scholars

Primary Indexes	Relative Weight	Secondary Indexes	Relative Weight	Final Weight
Data awareness level	0.1913	Data sensitivity level	0.3911	0.0748
		Data value perception level	0.3793	0.0726
		Data need awareness	0.2296	0.0439
Data discovery and access ability	0.1994	Data acquisition	0.3162	0.0631
		Data retrieval	0.3162	0.0631
		Access of data in various formats	0.0817	0.0163
		Data quality assessment	0.2858	0.0570
Data management and organization ability	0.1547	Data association	0.2685	0.0415
		Data extraction	0.4057	0.0628
		Metadata creation skills	0.3257	0.0504
Data processing and analysis ability	0.2073	Data analysis software	0.3940	0.0817
		Data modeling	0.2876	0.0596
		Data comparison	0.3183	0.0660
Data utilization and preservation ability	0.1502	Data application	0.4989	0.0749
		Data exchange	0.1923	0.0289
		Data preservation	0.3088	0.0464
Data ethics level	0.0970	Data legal level	0.5000	0.0485
		Data moral level	0.5000	0.0485
Total	1.0000			

5 Conclusion

Based on systematically reviewing the literature on data literacy and survey of social science scholars' data management behavior characteristics, this paper analyzed the components of data literacy for social science scholars. It established the data literacy evaluation system for social science scholars. According to experts' feedback on each index's rationality, importance, and description accuracy, the evaluation system was revised and improved. It applied the analytic hierarchy process (AHP) to calculate the index weights. The data literacy evaluation system for social science scholars can provide a reference for assessing and promoting social science scholars' data literacy ability in China. Relevant departments can also apply the evaluation system to the data literacy education and training and further optimize the evaluation system through education practice.

Acknowledgment

This research is supported by Chinese National Social Science Key Grant Funding "Big Data-Driven Cloud Platform Construction and Intelligent Service of Science and Education Evaluation".

References

Calzada, Prado, J. & Marzal, M. N. (2013). Incorporating data literacy into information literacy programs: Core competencies and contents. *Libri, 63* (2), 123–134.

Carlson, J., Fosmire, M., Miller, C. C. et al. (2011). Determining data information literacy needs: a study of students and research faculty. *Portal Libraries & the Academy, 11* (3), 257–271.

Chen, H., Zhao, N., & Wang, W. (2015). Basic problems of integrating big data into humanities and social sciences research. *Academic Forum, 38* (12), 106–110. (in Chinese).

Chen, M., & Qian, P. (2016). Interdisciplinary, multi–subject, multi–perspective—A survey of data literacy research. *Library Construction, 2016* (10), 50–56. (in Chinese).

Chen, Y., & Wu, X. (2012). "Replication research": A new trend in quantitative analysis of social science. *Evaluation and Management, 2012* (4), 47–47. (in Chinese).

Cheng, L. (2017). *A study of the status quo of postgraduates´ data literacy in china—Taking Wuhan university for example* [M.A. Thesis]. Wuhan University. (in Chinese).

Dai, L., & Liu, Y. (2018). Evaluation of college students´ information literacy from the perspective of information ecology. *Journal of Modern Information, 38* (12), 42–49. (in Chinese).

Finzer, W., & Parvate, V. (2021). Who will teach them about data? the responsibility mathematics of and statistics educators to support the integration of data analysis across all subjects. Retrieved June 27, from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.214.9735&rep=rep1&type=pdf.

Hanson, B., Sugden, A., & Alberts, B. (2011). Making data maximally available. *Science, 331* (6018), 649.

He, L., & Chang, Y. (2014). *Research progress of scientific data publishing at home and abroad. Library and Information Service, 58* (5), 104–110. (in Chinese).

Hu, Ran., & Wu, M. (2016). Research on data literacy capability embedded in research workflow and data life cycle. *Library and Information, 2016* (4), 125–137. (in Chinese).

Huang, R., & Li, B. (2016). Data literacy education: Development of information literacy education in big data era. *Document, Information & Knowledge, 2016 (*1), 21–29.(in Chinese).

Huang, X., Peng, J., & Zhang, M. (2019). Information literacy Construction of evaluation index system for college students in new environment. *Research on Library Science, 2019* (24), 25–33. (in Chinese).

Hunt, K. (2004). The challenges of integrating data literacy into the curriculum in an undergraduate. *International Association for Social Science Information Service and Technology (IASSIST), 2004* (28), 12–15.

Jiang, L., & Yang, H. (2015). Analysis on the reform of humanities and social sciences achievement evaluation in big data era. *Information and Documentation Services, 36* (3), 92–97. (in Chinese).

King, G. (2011). Ensuring the data–rich future of the social sciences. *Science, 331* (6018), 719–721.

Long, Q. (2015). Construction of data literacy ability index system and investigation and analysis of data literacy ability of college teachers and students. *Library, 2015* (12), 51–56. (in Chinese).

Mandinach, E. B, Friedman, J. M., & Gummer, E. S. (2015). How can schools of education help to build educators´ capacity to use data? A systemic view of the issue. *Teachers College Record, 117* (4), 1–50.

Meng, X., & Li, A. (2014). Research on scientific data literacy education in foreign university libraries. *Journal of Academic Libraries, 32* (3), 11–16.

Peng, J. (2013). Practice and enlightenment of humanities and social sciences data management in north America. *Journal of Academic Libraries, 31* (6), 33–37.

Peng, J. (2014). The development and reference of American Social Science Data Management Alliance (DataPass). *Library and Information Service, 2014* (10), 117–121. (in Chinese).

Ridsdale, C., Rothwell, J., Smit, M. et al.(2021). Strategies and best practices for data literacy education. Retrieved June 27, 2021, from http://www.mikesmit.com/wp–content/papercite–data/pdf/data_literacy.pdf .

Robbins, N. B. & Robbins, J. (2021). Quantitative literacy across the curriculum: Improving graphs in college textbooks. Retrieved June 27, 2021, from http://www.perceptualedge.com/articles/visual_business_intelligence/quantitative_literacy_across_curriculum.pdf

Schneider, R. (2013). Research data literacy. European Conference on Information Literacy. *Springer International Publishing*, 134–140.

Shen, T. (2015). Analysis on the usage characteristics of scientific data in the field of humanities and social sciences: An empirical study based on Chinese social science sample papers. *Journal of Academic Libraries, 33* (3), 101–107.

Shen, T. (2015). Data literacy and its impact on scientific data management. *Library Tribune, 2015* (1), 68–73.

Stephenson, E., & Caravello, P. S. (2007). Incorporating data literacy into undergraduate information literacy programs in the social sciences: A pilot project. *Reference Services Review, 35* (4), 525–540.

Tenopir, C., Allard, S., Douglass, K., et al. (2011). Data sharing by scientists: practices and perceptions. *Plos One, 6* (6), 21101.

Thompson, K. A., & Edelstein, D. M. (2004). A reference model for providing statistical consulting services in an academic library setting. *Iassist Quarterly, 2004* (2–3), 35.

Vahey, P., Rafanan, K., Patton, C. et al. (2012), A cross–disciplinary approach to teaching data literacy and proportionality. *Educational Studies in Mathematics, 81* (2), 179–205.

Victor, P. J., Otto, J., & Mutschler, C. (2013). Assessment of library instruction on undergraduate student success in a documents–based research course: The benefits of a librarian, archivist, and faculty collaboration. *Collaborative Librarianship, 5* (3), 154–176.

Xu, M., & Feng, X. (2018). A comparative study of social science data management and service in Chinese and foreign universities. *Library Work and Study, 2018* (2), 28–35. (in Chinese).

Zhang, J. (2013). Data literacy education in big data era. *Science, 65* (4), 29–32. (in Chinese).

Zhang, Y. (2015). Research on library scientific data management from the perspective of user data literacy education. *Library and Information, 2015* (4), 139–141. (in Chinese).