Heliyon 10 (2024) e35034

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

What are the key factors influencing scientific data sharing? A combined application of grounded theory and fuzzy-DEMATEL approach

Zhongyang Xu [a], Lingyu Liu [b,*], Zhiqian Meng [c]

[a] School of Information Management, Nanjing University, Nanjing, 210023, China
[b] Library and Information Service Center, Lishui University, Lishui, 323000, China
[c] School of Business Administration, University of Science and Technology Liaoning, Anshan, 114051, China

ARTICLE INFO

Keywords:
Scientific data
Scientific data sharing
Influencing factor
Grounded theory
Fuzzy-DEMATEL approach

ABSTRACT

Scientific data sharing (SDS) has become essential for scientific progress, technological innovation and socioeconomic development. Identifying the key influencing factors of SDS can effectively promote SDS programmes and give full play to the critical role of scientific data. This study used grounded theory and information ecology theory to construct an SDS influencing factor model that encompassed five dimensions and 28 influencing factors and followed the fuzzy decision-making trial and evaluation laboratory (fuzzy-DEMATEL) approach to measure and analyse the degree of influence of each influencing factor and identify the key factors. The results show that (1) there are interactions and mutual interactions between the various influencing factors of SDS, which can form a complex network system. (2) 16 influencing factors, such as data-sharing policies, data-sharing regulations and data-sharing standards, comprise the key influencing factors in SDS. (3) The optimisation path of SDS is 'Scientific Researchers' → 'Scientific Data' → 'Policy Environment' → 'Research Organisations → 'Information Technologies'. In this regard, we proposed the following management suggestions to promote the development of SDS programmes in China: focusing on researchers' subjective willingness to share, enhancing the integrated governance of scientific data, fulfilling the role of policy support and guidance, strengthening the support of research organisations and improving SDS platforms with information technology.

1. Introduction

Scientific data, also known as 'scientific research data' or 'scientific and technological data', refers to the data generated in the process of scientific research, scientific research activities and scientific research management, including data in the fields of natural science and engineering technology, such as experimental, observation and testing data, as well as data in the fields of social science, such as survey, statistical and census data [1–3]. Since human society entered the 'data intelligence era', with the rapid development of big data, blockchain, artificial intelligence and other high-tech information technologies, the development of modern science has gradually presented holistic, integrated and complex features. Relying only on the theory, method and technology of a single discipline makes it difficult to address the complex frontier problems in scientific research, which also makes scientific data a crucial issue in

* Corresponding author.
E-mail address: lly@lsu.edu.cn (L. Liu).

https://doi.org/10.1016/j.heliyon.2024.e35034
Received 16 June 2023; Received in revised form 2 July 2024; Accepted 22 July 2024
Available online 22 July 2024

scientific research. This also positions scientific data as an indispensable means of production in scientific research activities [4]. In this context, governments have begun to build and improve the scientific data governance system, promote scientific management and the sharing of scientific data and form interdisciplinary, cross-industry and cross-field scientific research cooperation to encourage the development of science and technology [5]. For example, the U.S. National Center for Biotechnology Information established the DNA sequence database GenBank (www.ncbi.nlm.nih.gov/genbank/), and the University of Cambridge in the United Kingdom established the Cambridge Crystallographic Data Centre (www.ccdc. cam.ac.uk).

In recent years, the Chinese government has, similarly, attached substantial importance to, and strongly supported, the open sharing of scientific data. In April 2018, the General Office of the State Council of China issued the Measures for the Management of Scientific Data [1], marking the Chinese government's formal elevation of scientific data sharing (SDS) to the level of national strategy. Since then, China has established a series of national SDS platforms, such as the National Basic Science Data Sharing Service Platform (http://www.nsdata.cn/), the Agricultural Science Data Sharing Center (http://www.agridata.cn/) and the National Forestry Science Data Sharing Service Platform (http://www2. forestdata.cn/). However, existing studies have shown that although the construction of SDS programmes in China has made considerable progress, problems still persist, such as low levels of data-sharing efficiency [6], difficulties in defining data property rights [7], insufficient support policies [8] and the risk of privacy leakage [9] in the construction and use of SDS platforms. The fundamental reason underlying such issues is that SDS is a systematic project involving numerous influencing factors, including research policy, the research environment, researchers and scientific data. The complex interplay between the influencing factors makes it challenging to grasp the key successful factors (hereinafter, the influencing factors refer to the successful factors). Therefore, herein, we take 'exploring the key influencing factors of SDS' as the research objective and adopt a combination of grounded theory and the fuzzy-DEMATEL method to measure and identify the key influencing factors of SDS from a systematic perspective.

We conduct this study to address the following three research questions.

(1) What factors can influence SDS?
(2) What are the key influencing factors on SDS?
(3) What measures can be taken to promote the development of China's SDS programme?

The specific steps taken in this study are shown in Fig. 1.

The remainder of this study is organised as follows. In Section 2, we examine recent research results on the influencing factors of SDS (hereinafter 'SDS influencing factors') found by information science and library science scholars. In Section 3, we elaborate on the research methods and specific steps in constructing the SDS influencing factor model and identifying key influencing factors. In Section 4, we present the model obtained from the research, interpret the specific meaning of each influencing factor and describe the results of the model calculation. In Section 5, we conduct an in-depth discussion and analysis of the results, provide some management suggestions for facilitating the development of SDS programmes in China and outline the limitations of the study and recommend future research directions. Finally, in Section 6, we elaborate on our conclusions and highlight the study's theoretical and practical contributions.

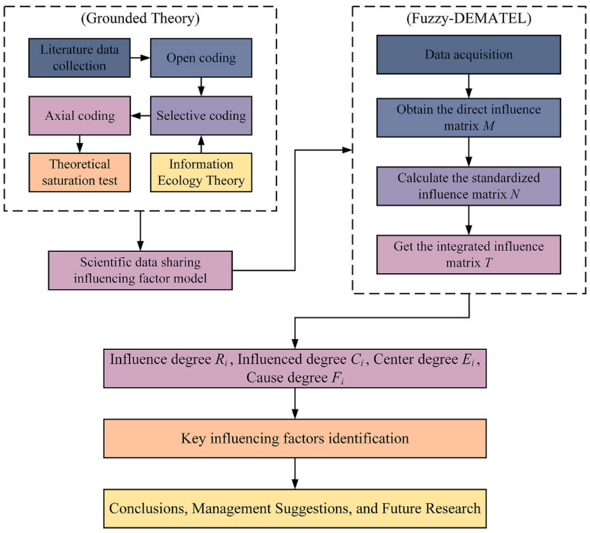

Fig. 1. Flowchart of research steps.

2. Literature review

2.1. Current status of research on SDS influencing factors

Substantial research on SDS influencing factors is currently available, making it a research hotspot in information science and library science. The existing literature mainly focuses on analysing and exploring SDS influencing factors from different perspectives, such as those of researchers, research institutions, data-sharing platforms, technologies and policies.

On the one hand, researchers are the main actors in SDS. Some scholars have explored the SDS influencing factors from the perspective of researchers using various research methods. For example, Dorta Gonzalez et al. [10] used multiple linear regression (MLR) to analyse the relationship between SDS frequency among researchers and the degree of importance of data citation, the degree of recognition of data-sharing activities, the degree of credit received for data sharing and the degree of incentive for data citation; the results showed that researchers' desire for recognition and praise are the primary triggers for SDS. Kim and Nah [11] combined the theory of planned behaviour (TPB) and a structural equation model (SEM) to analyse the specific mechanisms by which individual factors pertaining to researchers, such as data reuse experience, perceived benefit, perceived risk, perceived effort and perceived attitude, influence SDS. In addition, Moltgen et al. [12], Raffaghelli and Manca [13], Borgman et al. [14] and Bearth et al. [15] showed that factors such as personal background, data literacy and researchers' trust in others likewise have some influence on their SDS behaviour.

On the other hand, the organisational structure to which researchers belong will also impact their SDS behaviour. Therefore, some scholars have tried to use various methods to verify the specific impact of organisational factors on SDS. For example, Kim [16] and Kim and Stanton [17] used institutional theory and the TPB to construct a theoretical model and employed MLR and SEM, respectively, to verify that organisational culture and pressure can influence researchers' SDS behaviour. Searle et al. [18] conducted a case study with the librarians of Griffith University as the survey object; their results showed that the organisational structure has a particular impact on SDS and that clarifying the functional division of personnel in an organisation can promote the development of SDS services effectively in the organisation. In addition, Chawinga and Zinn [19], Gonzalez et al. [20] and Abdullahi and Noorhidawati [21] used theoretical analyses, literature reviews and partial least squares analyses to confirm that factors such as organisational management, the organisational climate and organisational incentives of research organisations can, similarly, influence SDS to some extent.

In addition, some scholars have investigated the platforms, technologies and policies related to SDS in detail to extract the SDS influencing factors. For example, Devriendt et al. [22], Sales and Sayao [23] and Si et al. [6] studied SDS platforms and data-sharing technologies. The aforementioned researchers observed that SDS platforms are an essential type of data infrastructure in the context of open science and that the construction of the platforms, the functional services of the platforms, data description, data organisation standards and data warehouse availability all have a direct or indirect impact on the open sharing of scientific data. Li et al. [24], Kim [25] and Zhu [26] analysed SDS policies and noted that policy and normative factors, such as open access policies for scientific data, data-sharing incentives, technical standards for metadata and specifications for data warehousing and data reuse, also impact SDS.

2.2. Review of the current state of research

Clearly, scholars in the fields of information science and library science have explored SDS influencing factors from multiple perspectives, including those of researchers, research organisations, SDS platforms, data-sharing technologies and data-sharing policies, using a variety of methods. Such investigations have provided rich results, providing an excellent theoretical foundation for the present research.

From the research content, SDS is evidently a systematic project that covers many different subjects, and its influencing factors are also characterised by multilevel and multi-type complex systems, which are not only large in number but also interact with the influencing factors. However, some gaps remain in the research on SDS influencing factors. The existing research still utilises a single perspective, such as personnel, organisation or platform, and few studies employ a systemic perspective. To comprehensively separate the SDS influencing factors and explore the key influencing factors, we should conduct research from a systemic standpoint.

From the research methods, theoretical analysis, content analysis, meta-analysis, MLR and SEM are the mainstream methods used in the research on SDS influencing factors and provide excellent reference values for our study. However, at the same time, although these techniques can test the role and influence mechanisms of specific SDS influencing factors effectively, these methods do not analyse and explore the interactions and degree of influence among these influencing factors in depth.

Therefore, we use a combination of qualitative and quantitative research methods to explore SDS influencing factors from a holistic perspective. First, we adopt a qualitative research method, grounded theory, to construct a reasonable SDS influencing factor model and then adopt a quantitative research method, the fuzzy-DEMATEL approach in systems science, to measure the degree of influence of each influencing factor and identify the key factors among them. Finally, we provide management suggestions for promoting SDS and theoretical references for the management and open sharing of scientific data in China.

3. Methodology

3.1. Model construction method

The construction of an SDS influencing factor model is crucial for this research. Here, we used grounded theory to construct the SDS influencing factor model and verified the validity of the model. Grounded theory is a bottom-up qualitative research method for building substantive theory proposed by Columbia University scholars B. G. Glaser and A. L. Strauss A [27]. This method first refines the concepts and categories reflecting the relevant phenomena by systematically collecting, summarising and analysing the original data, obtains the core categories by continuously revising and improving the correlation between the categories and finally builds a substantive theory that reflects the essence of the phenomena by exploring the logic and inner connections between the core categories [27]. The current research paradigm of analysing literature data with the help of grounded theory to construct theoretical models is, likewise, widespread [28,29]. In this study, the specific processes used to conduct the research using grounded theory were as follows: literature data collection, open coding, axial coding, selective coding and a theoretical saturation test.

3.2. Model calculation method

Here, we use the fuzzy-DEMATEL approach as the model calculation method to identify key SDS influencing factors. DEMATEL is a systems science approach proposed by the American scholars A. Gabus and E. Fontela in 1971; this approach treats complex real-world problems as directed graphs with weights and then uses graph theory, matrix, mathematical theory and other methods to analyse and process them (Fig. 2) [30]. In recent years, the DEMATEL approach has been widely used to identify key influencing factors in some large-scale projects, e.g. smart city construction [31], smart urban rail transit development [32] and healthcare big data assetisation [33]. As mentioned previously, SDS is a systematic project with an enormous scope. Therefore, we believe that the DEMATEL method is also applicable to our study, as it can identify the key SDS influencing factors effectively and provide important theoretical references for the construction of SDS projects. However, although the traditional DEMATEL approach can effectively determine the degree of influence and the mechanism of action between the influencing factors in a complex system, the approach determines the degree of influence between the factors by constructing a direct influence matrix. The source of data for the direct influence matrix is the scoring of the expert group, which makes the data susceptible to the influence of the subjective will of the experts [34]. Therefore, to minimise the research error caused by the subjective influence of the experts, we referred to the research results of Barghi and Sikari [35] and Du and Li [36] and combined the traditional DEMATEL and fuzzy set theory to construct the fuzzy-DEMATEL approach. Meanwhile, we apply the fuzzy-DEMATEL approach to our study, using the triangular fuzzy quantitative expert group's subjective scores. We believe that such an approach can reduce research errors effectively, thereby improving the accuracy of the results.

4. Results

4.1. Model construction results

4.1.1. Literature data collection

The China Academic Journal Network Publishing Database (CAJD) is the world's largest continuously and dynamically updated full-text database of Chinese academic journals. It currently contains 7385 domestic academic journals covering various fields such as the natural sciences, engineering technology, agriculture, philosophy, medicine, humanities and the social sciences, with a total of more than 22.25 million full-text documents. To obtain the full-text data conveniently and construct the SDS influencing factors model as comprehensively as possible, we chose CAJD as the literature data source. As the research object, we selected full-text data from information science and library science journals related to SDS in the past 10 years. We set the search conditions as follows: (Topic = scientific data sharing OR research data sharing OR science and technology data sharing), set the literature type as the China Social Science Citation Index, the timespan as 1 January 2013–31 December 2022 and the search time as January 1, 2023. Finally, we obtained 450 papers; after excluding invalid documents such as those unrelated to the information science and library science disciplines, we narrowed it down to 322 target papers and downloaded their full text.

4.1.2. Open coding

A 'category' is a structural concept with the same properties and the common property on which things need to be categorised [27]. 'Open coding' is the process of labelling, conceptualising and categorising primary documents to form categories [27]. We created a research team of two PhD students and two MSc students studying SDS on January 4, 2023 and conducted open coding between 5 January and January 25, 2023. The research team randomly selected 215 of the 322 target documents (2/3) for data coding and then saved the remaining 107 papers (1/3) for the theoretical saturation test. Subsequently, all the team members read the abstracts and full texts of the 215 target documents word by word, focusing on the theme of 'SDS influencing factors', extracted the statements with the highest relevance in the literature and made an initial conceptualisation of the influencing factors involved therein, thereby forming

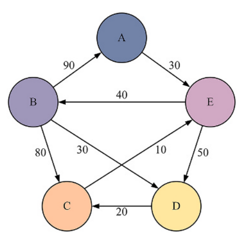

Fig. 2. DEMATEL approach.

1411 initial concepts. Finally, the team members conducted a continuous comparative and inductive analysis of all initial concepts to obtain 207 categories. The open coding process is shown in Table 1.

4.1.3. Axial coding

'Axial coding' is the process of summarising and analysing the inner logicality of all the obtained categories with higher precision and depth based on open coding, finally forming the main categories [27]. The research team performed the axial coding of the obtained 207 categories from 30 January to February 6, 2023 and finally arrived at 28 main categories such as data-sharing policies, data-sharing regulations, data-sharing standards and data-sharing technologies. The axial coding process is shown in Table 2.

4.1.4. Selective coding

'Selective coding' is a deeper comparison, analysis and abstraction of the main categories compared to the primary axial coding. Through selective coding, a new theoretical framework is constructed by extracting the core categories with high generality [27]. The research team conducted selective coding of the above 28 main categories from 7 February to February 11, 2023 and discussed them in depth by continuously comparing and analysing the internal logic and interactions between the main categories and combining them with the 'four elements theory' of information ecology theory [37]. Following two rounds of discussion and consensus, we obtained five core categories: policy environment, information technologies, research organisations, scientific researchers and scientific data. The selective coding process is shown in Table 3.

4.1.5. Theoretical saturation test

A theoretical model constructed from existing data is considered to have reached 'theoretical saturation' when no new categories can be obtained after introducing new data [27]. In this study, the remaining 107 papers (1/3) were used to test for theoretical saturation, and the remaining documents were coded to pass the test if no new categories emerged and to fail if new categories emerged. The research team performed open coding of the remaining studies between 12 February and February 20, 2023, and the results showed that the categories were included in the 207 categories obtained from '4.1.2 Open coding', and no new categories were found. Therefore, it can be concluded that the theoretical model constructed by this grounded theory passes the theoretical saturation test. Referring to Table 3, we built the SDS influencing factors model covering five dimensions and 28 influencing factors (Fig. 3).

4.2. Model calculation results

4.2.1. Data acquisition and fuzzification

To scientifically assess the specific degree of influence among the SDS influencing factors, we first designed a scoring questionnaire based on the 28 influencing factors in the SDS influencing factor model (Fig. 3), as shown in Table 4.

Second, we invited eight experts to form an expert scoring panel. The panel members included two researchers who had applied for provincial or national level funds, two university teachers based in the field of information science and library science, two PhD students in the SDS research field and two enterprise personnel with experience in SDS platform project construction. The personal information of each expert is shown in Table 5.

Next, we asked the experts to compare the influencing factors individually without communicating with each other and rate them according to the degree of influence perceived. The scoring criteria were as follows: 0 = No influence, 1 = Weak influence, 2 = Certain influence, 3 = Strong influence and 4 = Extremely strong influence. Subsequently, we collected eight questionnaire rating forms.

Finally, we fuzzified the expert scores, as shown in Table 6 [38]. The transformed expert scores are triangular fuzzy numbers $F_{ijt} = \left(f_{ij}, g_{ij}, u_{ij}\right), 1 \leq t \leq 8$, and the formula represents the fuzzification of the score of any expert with factor *i* to factor *j*.

4.2.2. Data deblurring process

We used the triangular fuzzy numbers of the scores from eight experts to obtain the direct influence matrix *M*. The steps are presented as follows.

Table 1
Open coding process (partial).

Original documentation statements	Conceptualisation	Categories
'The significant effect of subjective specification on willingness to share scientific data suggests that the influence of others can significantly affect the willingness of researchers to share scientific data'. 'In the process of sharing scientific data, researchers are also very concerned about enhancing their reputation and recognising and respecting others'.	Subjective specification, Reputation, Recognition, Respect	Subjective specification, Reputation benefits, Psychological benefits
'The data resource layer is the foundation of the scientific data-sharing platform'. 'Identify scientific data with Digital Object Identifier (DOI) and develop a relatively uniform metadata standard'. 'The quality of data reflects the value of data and guarantees the level of knowledge services'. 'Integrity protection of scientific data is also critical in data integration'.	Data resource layer, DOI, Metadata standards, Data quality, Data integrity	Metadata standards, Data integrity, Data quality, Data standards
... ...	... ...	... ...

Table 2
Axial coding process.

Main categories	Categories (partial)
Data-sharing policies	National policies and regulations, Local government policies and regulations, etc.
Data-sharing regulations	Data property rights protection regulations, Data transfer regulations, Data use licence regulations, etc.
Data-sharing standards	Data processing standards, Data citation standards, Metadata standards, etc.
Data-sharing technologies	Data mining technologies, Data visualisation technologies, Intelligent retrieval technologies, etc.
Platform-building technologies	Institutional knowledge base, Data guardianship technologies, Data repository technologies, etc.
Platform optimisation technologies	Function optimisation, Information optimisation, Interface optimisation, Design optimisation, etc.
Data security technologies	Authentication and authorisation technologies, Data encryption technologies, Data shielding technology, etc.
Data management mechanism	Access mechanism, Preservation mechanism, Management mechanism, Sharing mechanism, etc.
Organisational structure	Organisational internal structure, Division of labour within the organisation, Organisational external relations, etc.
Organisational climate	Organisational cooperation atmosphere, Organisational communication atmosphere, Organisational sharing atmosphere, etc.
Organisational motivation	Material motivation, Spiritual motivation and Organisational motivation.
Organisational culture	Organisational values, Organisational culture, Organisational shared culture, etc.
Organisational support	Technical support, Educational support, Publicity support, Moral support and Spiritual support.
Organisational management	Process management, Model management, Efficiency management, Risk management and Property management.
Personal characteristics	Age, Gender, Education, Major, Industry, etc.
Personal literacy	Information literacy, Data literacy, Algorithmic literacy, Management literacy, etc.
Shared willingness	Personal willingness, Participation willingness, Shared willingness, Demand willingness, etc.
Subjective specification	Occupational pressure, Organisational pressure, Industry pressure, Social pressure, etc.
Perceived risk	Data infringement, Data distortion, Data monopoly, Data security, Data misuse, etc.
Perceived efforts	Time costs, Economic costs, Experience costs, Competition costs, Opportunity costs, etc.
Perceived benefits	Economic benefits, Reputational benefits, Psychological benefits, Expectation benefits, Academic benefits, etc.
Perceived attitudes	Shared attitudes, Institutional attitudes, Environmental attitudes, etc.
Data quality	Accuracy, Consistency, Authenticity, Completeness, Normality, etc.
Data descriptions	Data name, Data size, Data content, Logical structure, etc.
Data values	Economic values, Academic values, Social values, etc.
Data types	Personal data, Experimental data, Web data, Literature data, etc.
Data formats	Structured data, Semi-structured data and Unstructured data.
Data sensitivity	Accessibility, Specificity, Confidentiality, Sensitivity, etc.

Table 3
Selective coding process.

Core categories	Main categories
Policy Environment	Data-sharing policies, Data-sharing regulations, Data-sharing standards
Information Technologies	Data-sharing technologies, Platform-building technologies, Platform optimisation technologies, Data security technologies, Data management mechanism
Research Organisations	Organisational structure, Organisational climate, Organisational motivation, Organisational culture, Organisational support, Organisational management
Scientific Researchers	Personal characteristics, Personal literacy, Shared willingness, Subjective specification, Perceived risk, Perceived efforts, Perceived benefits, Perceived attitudes
Scientific Data	Data quality, Data descriptions, Data values, Data types, Data formats, Data sensitivity

(1) Standardisation of triangular fuzzy number matrix. We standardised the triangular fuzzy numbers of each expert according to equations (1)–(3) to reduce the range of fuzziness, thereby reducing the error caused by the individual subjective differences among experts.

$$xf_{ij}^{t} = \frac{f_{ij}^{t} - \min f_{ij}^{t}}{\Delta_{\min}^{\max}}, 1 \le t \le 8 \tag{1}$$

$$xg_{ij}^{t} = \frac{g_{ij}^{t} - \min g_{ij}^{t}}{\Delta_{\min}^{\max}}, 1 \le t \le 8 \tag{2}$$

$$xu_{ij}^{t} = \frac{u_{ij}^{t} - \min u_{ij}^{t}}{\Delta_{\min}^{\max}}, 1 \le t \le 8 \tag{3}$$

Notes: $\Delta_{\min}^{\max} = \max u_{ij}^{t} - \min f_{ij}^{t}, 1 \le t \le 8.$

(2) Reduce the fuzzy number, and calculate the standard values. According to equations (4)–(7), we determined the left standard value xfs_{ij}^{t}, the right standard value xus_{ij}^{t} and the total standard value w_{ij}^{t}. We show the fuzzification and defuzzification processes of the data by taking the scoring data of 'Expert 1' as an example, as shown in Fig. 4.

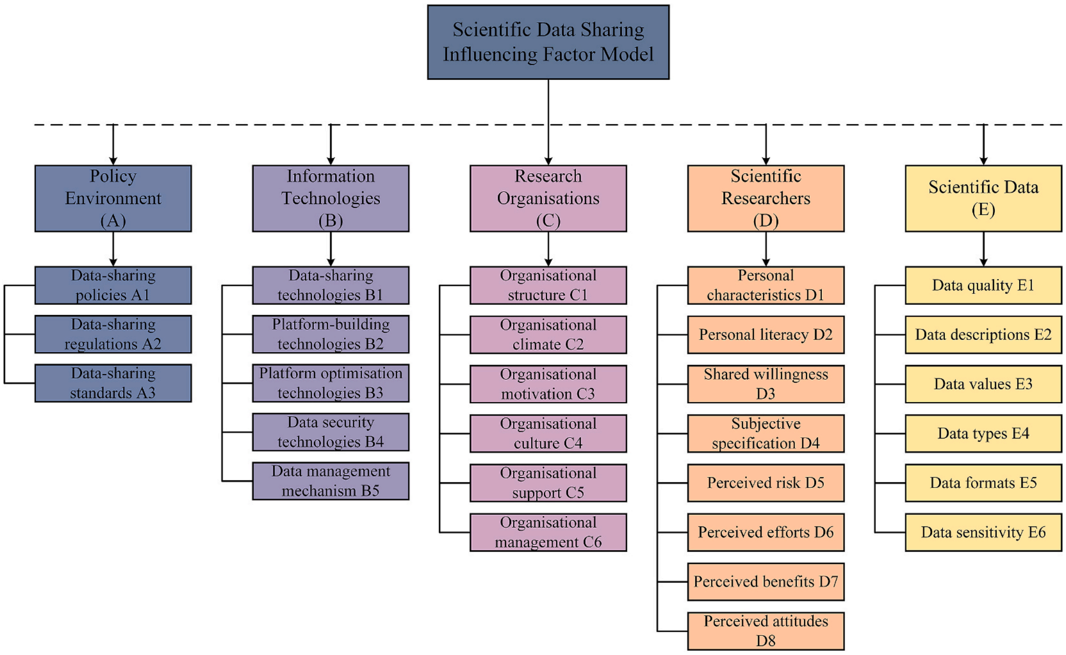

Fig. 3. The SDS influencing factor model.

Table 4
Scoring questionnaire (partial).

	Data-sharing policies A1	Data-sharing regulations A2	… …	Data sensitivity E6
Data-sharing policies A1	0			
Data-sharing regulations A2		0		
… …			… …	
Data sensitivity E6				0

Table 5
Description of experts' personal information.

No.	Occupation	Education	Title	Research direction	Fund experience	Project experience
1	Researchers	Doctor	Associate Researcher	Scientific Data Sharing	Yes	No
2	Researchers	Doctor	Associate Researcher	Scientific Data Governance	Yes	Yes
3	University teacher	Doctor	Associate Professor	Informetrics	Yes	No
4	University teacher	Master	Professor	Information Laws	No	No
5	Enterprise personnel	Master	Senior Engineer	Database Technology	No	Yes
6	Enterprise personnel	Master	Engineer	Database Technology	No	Yes
7	PhD student	Doctor	None	Scientific Data Sharing	Yes	No
8	PhD student	Doctor	None	Scientific Data Sharing	Yes	No

Table 6
Triangular fuzzy numbers for expert scores.

Language variables	Score	Corresponding triangular fuzzy number
No influence	0	(0, 0, 0.25)
Weak influence	1	(0, 0.25, 0.5)
Certain influence	2	(0.25, 0.5, 0.75)
Strong influence	3	(0.5, 0.75, 1)
Extremely strong influence	4	(0.75, 1, 1)

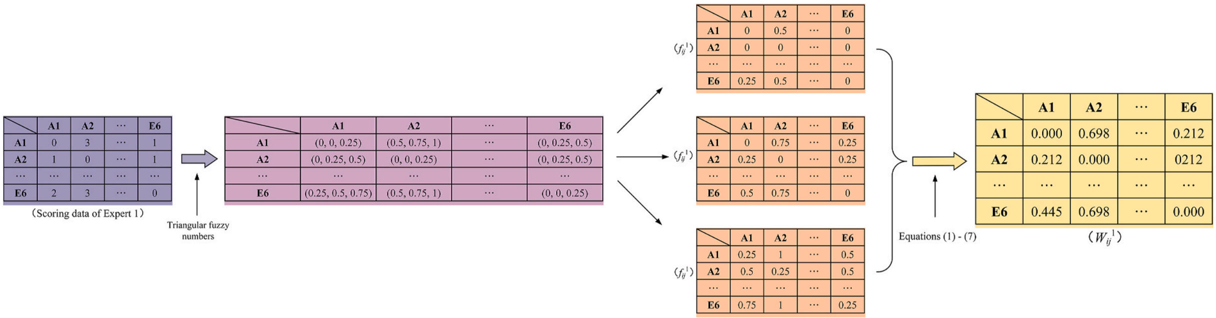

Fig. 4. Fuzzification and defuzzification of data.

$$xfs_{ij}^{t} = \frac{xg_{ij}^{t}}{1 + xg_{ij}^{t} - xf_{ij}^{t}}, 1 \le t \le 8 \tag{4}$$

$$xus_{ij}^{t} = \frac{xu_{ij}^{t}}{1 + xu_{ij}^{t} - xg_{ij}^{t}}, 1 \le t \le 8 \tag{5}$$

$$x_{ij}^{t} = \frac{xfs_{ij}^{t} \cdot \left(1 - xfs_{ij}^{t}\right) + xus_{ij}^{t} \cdot xus_{ij}^{t}}{1 - xfs_{ij}^{t} + xus_{ij}^{t}}, 1 \le t \le 8 \tag{6}$$

$$w_{ij}^{t} = \min f_{ij}^{t} + x_{ij}^{t} \cdot \Delta_{\min}^{\max}, 1 \le t \le 8 \tag{7}$$

(3) Calculate the overall standardised influence degree value. Based on the total standard value w_{ij}^{t} and equation (8), we calculated the overall standardised influence degree values of the eight experts m_{ij} and obtained the direct influence matrix $M = (m_{ij})_{28 \times 28}$ (Fig. 5).

$M=$

	A1	A2	A3	B1	B1	B3	B4	B5	C1	C2	C3	C4	C5	C6	D1	D2	D3	D4	D5	D6	D7	D8	E1	E2	E3	E4	E5	E6
A1	0.000	0.698	0.698	0.666	0.666	0.666	0.635	0.698	0.212	0.698	0.445	0.445	0.509	0.666	0.241	0.477	0.731	0.445	0.666	0.445	0.666	0.445	0.698	0.212	0.448	0.212	0.212	0.212
A2	0.212	0.000	0.416	0.416	0.477	0.477	0.416	0.445	0.445	0.445	0.445	0.445	0.698	0.933	0.212	0.445	0.698	0.698	0.698	0.212	0.698	0.416	0.698	0.387	0.606	0.358	0.212	0.212
A3	0.212	0.056	0.000	0.445	0.445	0.445	0.445	0.445	0.000	0.000	0.000	0.000	0.000	0.445	0.000	0.212	0.212	0.445	0.445	0.698	0.445	0.212	0.967	0.967	0.666	0.698	0.698	0.967
B1	0.445	0.445	0.933	0.000	0.698	0.445	0.698	0.445	0.212	0.000	0.000	0.000	0.212	0.000	0.000	0.212	0.445	0.212	0.212	0.000	0.212	0.212	0.445	0.445	0.445	0.445	0.212	0.445
B2	0.445	0.445	0.445	0.698	0.000	0.967	0.967	0.212	0.212	0.445	0.212	0.000	0.212	0.212	0.000	0.445	0.445	0.212	0.212	0.445	0.212	0.212	0.445	0.212	0.445	0.212	0.212	0.698
B3	0.445	0.445	0.445	0.212	0.698	0.000	0.698	0.212	0.212	0.445	0.241	0.000	0.212	0.241	0.000	0.445	0.445	0.212	0.445	0.445	0.445	0.445	0.445	0.445	0.212	0.212	0.212	0.445
B4	0.666	0.445	0.902	0.666	0.899	0.899	0.000	0.698	0.212	0.212	0.212	0.000	0.185	0.212	0.000	0.445	0.698	0.416	0.445	0.212	0.445	0.574	0.445	0.445	0.445	0.212	0.212	0.698
B5	0.212	0.967	0.212	0.212	0.212	0.212	0.212	0.000	0.000	0.445	0.698	0.212	0.212	0.698	0.000	0.445	0.445	0.212	0.212	0.212	0.212	0.445	0.445	0.212	0.445	0.445	0.698	0.445
C1	0.000	0.000	0.000	0.000	0.212	0.212	0.212	0.445	0.000	0.445	0.445	0.212	0.212	0.698	0.445	0.445	0.212	0.445	0.212	0.212	0.212	0.212	0.445	0.212	0.212	0.212	0.212	0.212
C2	0.445	0.241	0.212	0.000	0.000	0.000	0.000	0.000	0.241	0.000	0.445	0.445	0.445	0.445	0.698	0.445	0.902	0.698	0.698	0.212	0.698	0.445	0.212	0.212	0.445	0.212	0.212	0.212
C3	0.445	0.000	0.212	0.000	0.000	0.000	0.000	0.000	0.241	0.698	0.000	0.698	0.635	0.666	0.477	0.445	0.731	0.445	0.445	0.241	0.967	0.698	0.698	0.445	0.698	0.212	0.212	0.445
C4	0.445	0.933	0.445	0.000	0.000	0.000	0.000	0.000	0.056	0.933	0.698	0.000	0.445	0.698	0.212	0.212	0.212	0.445	0.445	0.212	0.445	0.445	0.212	0.212	0.212	0.000	0.212	0.212
C5	0.212	0.212	0.212	0.212	0.445	0.445	0.445	0.445	0.000	0.445	0.445	0.212	0.000	0.445	0.056	0.212	0.445	0.445	0.445	0.212	0.445	0.698	0.445	0.212	0.445	0.000	0.000	0.212
C6	0.212	0.445	0.212	0.000	0.212	0.212	0.000	0.666	0.445	0.899	0.899	0.698	0.666	0.000	0.445	0.445	0.445	0.445	0.445	0.445	0.445	0.445	0.698	0.212	0.902	0.212	0.445	0.445
D1	0.000	0.000	0.000	0.000	0.053	0.053	0.026	0.000	0.666	0.241	0.082	0.416	0.000	0.082	0.000	0.933	0.641	0.000	0.270	0.000	0.270	0.212	0.445	0.445	0.445	0.445	0.445	0.445
D2	0.212	0.000	0.000	0.212	0.212	0.212	0.212	0.000	0.698	0.967	0.270	0.445	0.445	0.212	0.000	0.000	0.967	0.698	0.698	0.445	0.477	0.445	0.445	0.445	0.445	0.445	0.445	0.212
D3	0.666	0.302	0.212	0.445	0.445	0.445	0.445	0.212	0.000	0.698	0.445	0.212	0.212	0.212	0.000	0.698	0.000	0.445	0.445	0.445	0.666	0.445	0.445	0.445	0.445	0.212	0.212	0.212
D4	0.445	0.000	0.000	0.000	0.212	0.212	0.000	0.000	0.445	0.698	0.445	0.212	0.212	0.445	0.000	0.000	0.731	0.000	0.445	0.445	0.445	0.445	0.212	0.212	0.212	0.212	0.212	0.445
D5	0.000	0.212	0.212	0.212	0.212	0.212	0.445	0.212	0.000	0.445	0.866	0.445	0.212	0.698	0.000	0.000	0.899	0.445	0.000	0.445	0.000	0.212	0.212	0.212	0.445	0.212	0.212	0.698
D6	0.000	0.000	0.212	0.000	0.000	0.000	0.000	0.000	0.000	0.212	0.212	0.212	0.445	0.212	0.000	0.000	0.445	0.000	0.698	0.000	0.212	0.212	0.212	0.212	0.212	0.445	0.445	0.445
D7	0.000	0.445	0.000	0.000	0.445	0.445	0.445	0.212	0.000	0.477	0.899	0.445	0.212	0.698	0.000	0.000	0.899	0.000	0.000	0.212	0.000	0.212	0.212	0.445	0.445	0.212	0.212	0.698
D8	0.212	0.212	0.212	0.000	0.000	0.000	0.000	0.000	0.000	0.212	0.212	0.212	0.212	0.212	0.000	0.000	0.445	0.445	0.698	0.445	0.698	0.000	0.212	0.212	0.212	0.185	0.185	0.445
E1	0.445	0.212	0.445	0.445	0.445	0.445	0.445	0.445	0.000	0.000	0.212	0.212	0.445	0.212	0.000	0.698	0.698	0.212	0.445	0.445	0.445	0.212	0.000	0.212	0.967	0.212	0.212	0.445
E2	0.445	0.212	0.212	0.445	0.000	0.000	0.445	0.000	0.000	0.000	0.000	0.000	0.212	0.000	0.000	0.212	0.212	0.212	0.212	0.212	0.212	0.212	0.445	0.000	0.445	0.185	0.185	0.000
E3	0.698	0.445	0.212	0.445	0.445	0.445	0.445	0.445	0.000	0.000	0.698	0.212	0.445	0.445	0.000	0.698	0.731	0.212	0.445	0.212	0.698	0.212	0.698	0.212	0.000	0.212	0.212	0.212
E4	0.445	0.212	0.445	0.212	0.212	0.212	0.212	0.416	0.000	0.000	0.000	0.000	0.000	0.212	0.000	0.000	0.000	0.000	0.212	0.445	0.000	0.212	0.212	0.212	0.212	0.000	0.185	0.212
E5	0.445	0.212	0.445	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.212	0.000	0.000	0.000	0.000	0.212	0.445	0.000	0.212	0.212	0.212	0.212	0.698	0.000	0.212
E6	0.445	0.666	0.445	0.212	0.445	0.445	0.967	0.445	0.000	0.000	0.000	0.000	0.212	0.212	0.000	0.000	0.212	0.212	0.445	0.698	0.212	0.212	0.212	0.212	0.967	0.212	0.212	0.000

Fig. 5. Direct influence matrix *M*.

$$m_{ij} = \frac{\sum_{t=1}^{t} w_{ij}^{t}}{t}, 1 \leq t \leq 8 \tag{8}$$

4.2.3. Matrix standardisation process

In this study, we obtained the integrated influence matrix T by matrix normalisation. The steps are presented as follows. First, according to equation (9), we used the row and maximum method to process the direct influence matrix M to obtain the normalised influence matrix N. Second, we substituted the normalised matrix N into equation (10) for calculation, where I was a $28\times$ 28 unit matrix. Finally, we obtained the integrated influence matrix $T = (t_{ij})_{28\times 28}$ (Fig. 6). It is worth noting that the integrated influence matrix T is an intermediate process matrix, which is mainly used to calculate the values of indicators such as influence degree, influenced degree, centre degree and cause degree of all the variables.

$$N = \frac{1}{\max \sum_{j=1}^{28} n_{ij}} \cdot M, 1 \leq i \leq 28 \tag{9}$$

$$T = \lim_{n\to\infty} \left(I + N^1 + N^2 + \cdots + N^n\right) = N \cdot (I - N)^{-1} \tag{10}$$

4.2.4. Presentation of calculation results

In accordance with equations (11)–(14), we calculated the values and rankings of influence degree R_i, influenced degree C_i, centre degree E_i and cause degree F_i for each influence factor and constructed the 'centre degree–cause degree' causality diagram with the centre degree E_i as the x-axis and the cause degree F_i as the y-axis, as shown in Table 7 and Fig. 7.

$$R_i = \sum_{j=1}^{28} t_{ij} \tag{11}$$

$$C_i = \sum_{i=1}^{28} t_{ij} \tag{12}$$

$$E_i = R_i + C_i \tag{13}$$

T=	A1	A2	A3	B1	B1	B3	B4	B5	C1	C2	C3	C4	C5	C6	D1	D2	D3	D4	D5	D6	D7	D8	E1	E2	E3	E4	E5	E6
A1	0.070	0.115	0.113	0.097	0.112	0.113	0.115	0.105	0.045	0.127	0.110	0.082	0.099	0.127	0.038	0.100	0.162	0.100	0.134	0.102	0.134	0.104	0.136	0.081	0.128	0.071	0.070	0.098
A2	0.079	0.059	0.085	0.074	0.092	0.092	0.092	0.082	0.059	0.104	0.105	0.079	0.107	0.138	0.035	0.091	0.149	0.111	0.127	0.080	0.128	0.095	0.128	0.085	0.130	0.074	0.064	0.089
A3	0.067	0.052	0.048	0.069	0.078	0.078	0.085	0.072	0.018	0.046	0.050	0.031	0.043	0.082	0.010	0.058	0.088	0.073	0.090	0.102	0.087	0.062	0.127	0.112	0.116	0.090	0.088	0.125
B1	0.077	0.074	0.109	0.038	0.094	0.078	0.099	0.070	0.032	0.043	0.044	0.026	0.051	0.048	0.009	0.056	0.096	0.056	0.068	0.047	0.067	0.058	0.088	0.074	0.092	0.067	0.050	0.084
B2	0.085	0.080	0.083	0.089	0.053	0.119	0.123	0.059	0.037	0.084	0.067	0.033	0.060	0.070	0.013	0.079	0.110	0.064	0.079	0.084	0.078	0.067	0.095	0.063	0.101	0.056	0.054	0.109
B3	0.079	0.075	0.076	0.052	0.094	0.047	0.099	0.054	0.035	0.082	0.067	0.033	0.057	0.070	0.013	0.074	0.106	0.061	0.091	0.081	0.089	0.079	0.090	0.075	0.081	0.053	0.052	0.087
B4	0.108	0.090	0.121	0.094	0.123	0.123	0.068	0.099	0.040	0.079	0.078	0.040	0.066	0.082	0.015	0.087	0.141	0.087	0.106	0.079	0.105	0.101	0.108	0.089	0.114	0.064	0.063	0.121
B5	0.064	0.111	0.058	0.048	0.057	0.057	0.060	0.037	0.021	0.084	0.101	0.051	0.060	0.104	0.016	0.075	0.104	0.062	0.075	0.064	0.075	0.081	0.091	0.058	0.098	0.069	0.086	0.084
C1	0.034	0.031	0.028	0.022	0.043	0.043	0.045	0.057	0.017	0.072	0.071	0.042	0.047	0.087	0.044	0.065	0.070	0.064	0.058	0.050	0.058	0.051	0.073	0.045	0.063	0.042	0.043	0.054
C2	0.075	0.057	0.051	0.029	0.038	0.039	0.041	0.033	0.038	0.055	0.085	0.068	0.072	0.084	0.064	0.074	0.136	0.093	0.105	0.061	0.106	0.078	0.070	0.056	0.092	0.049	0.049	0.066
C3	0.083	0.049	0.058	0.034	0.045	0.045	0.049	0.040	0.039	0.110	0.062	0.091	0.093	0.107	0.051	0.080	0.135	0.082	0.097	0.070	0.134	0.103	0.112	0.078	0.121	0.054	0.054	0.089
C4	0.072	0.103	0.067	0.028	0.036	0.036	0.038	0.033	0.025	0.117	0.100	0.037	0.074	0.102	0.032	0.055	0.086	0.077	0.088	0.059	0.089	0.077	0.070	0.055	0.076	0.033	0.048	0.064
C5	0.059	0.056	0.054	0.046	0.071	0.072	0.074	0.065	0.018	0.080	0.081	0.047	0.040	0.081	0.017	0.056	0.101	0.074	0.086	0.059	0.087	0.094	0.085	0.055	0.091	0.033	0.034	0.066
C6	0.074	0.086	0.064	0.038	0.064	0.064	0.054	0.090	0.057	0.130	0.131	0.095	0.102	0.070	0.052	0.088	0.125	0.089	0.106	0.090	0.107	0.093	0.121	0.067	0.143	0.060	0.077	0.096
D1	0.032	0.027	0.026	0.021	0.030	0.030	0.032	0.023	0.061	0.052	0.039	0.052	0.027	0.040	0.009	0.095	0.092	0.030	0.057	0.032	0.056	0.045	0.068	0.059	0.073	0.056	0.055	0.063
D2	0.064	0.043	0.041	0.047	0.056	0.056	0.060	0.036	0.068	0.122	0.074	0.067	0.075	0.070	0.016	0.043	0.143	0.096	0.109	0.081	0.093	0.081	0.087	0.074	0.095	0.067	0.067	0.068
D3	0.098	0.069	0.062	0.068	0.078	0.079	0.082	0.054	0.022	0.106	0.087	0.053	0.062	0.073	0.015	0.095	0.083	0.081	0.095	0.083	0.111	0.084	0.093	0.078	0.100	0.055	0.054	0.074
D4	0.067	0.033	0.031	0.023	0.045	0.046	0.034	0.027	0.045	0.091	0.073	0.043	0.049	0.073	0.013	0.033	0.107	0.035	0.077	0.069	0.077	0.069	0.058	0.046	0.064	0.042	0.042	0.071
D5	0.046	0.055	0.054	0.044	0.052	0.053	0.072	0.048	0.018	0.080	0.108	0.063	0.055	0.097	0.014	0.040	0.128	0.073	0.054	0.075	0.055	0.060	0.068	0.054	0.091	0.047	0.048	0.097
D6	0.024	0.022	0.036	0.016	0.020	0.020	0.023	0.018	0.007	0.039	0.041	0.032	0.052	0.042	0.006	0.019	0.066	0.022	0.078	0.026	0.042	0.039	0.043	0.036	0.048	0.050	0.049	0.058
D7	0.046	0.073	0.039	0.031	0.070	0.070	0.074	0.049	0.018	0.082	0.110	0.063	0.056	0.097	0.015	0.042	0.128	0.042	0.054	0.059	0.056	0.061	0.069	0.070	0.092	0.047	0.047	0.096
D8	0.043	0.041	0.039	0.019	0.025	0.025	0.028	0.022	0.010	0.047	0.049	0.037	0.041	0.049	0.008	0.023	0.077	0.058	0.084	0.062	0.084	0.029	0.049	0.041	0.055	0.035	0.035	0.065
E1	0.082	0.063	0.077	0.070	0.080	0.080	0.084	0.072	0.020	0.054	0.069	0.049	0.075	0.070	0.011	0.095	0.126	0.062	0.092	0.082	0.092	0.065	0.062	0.061	0.134	0.054	0.054	0.088
E2	0.057	0.038	0.040	0.052	0.026	0.025	0.057	0.022	0.010	0.025	0.026	0.016	0.037	0.027	0.005	0.038	0.054	0.038	0.046	0.040	0.045	0.039	0.062	0.023	0.064	0.033	0.032	0.029
E3	0.103	0.082	0.065	0.071	0.083	0.083	0.086	0.075	0.023	0.062	0.108	0.055	0.081	0.092	0.015	0.099	0.136	0.067	0.097	0.070	0.116	0.071	0.116	0.065	0.074	0.056	0.056	0.077
E4	0.054	0.038	0.055	0.033	0.037	0.037	0.040	0.049	0.008	0.022	0.024	0.015	0.020	0.040	0.005	0.021	0.034	0.021	0.044	0.056	0.027	0.037	0.045	0.036	0.047	0.020	0.033	0.043
E5	0.049	0.031	0.049	0.014	0.016	0.016	0.018	0.016	0.006	0.016	0.018	0.012	0.016	0.035	0.004	0.015	0.025	0.016	0.038	0.052	0.020	0.032	0.038	0.031	0.040	0.065	0.015	0.036
E6	0.076	0.089	0.075	0.051	0.076	0.076	0.114	0.070	0.017	0.044	0.047	0.029	0.054	0.064	0.010	0.041	0.082	0.055	0.085	0.093	0.068	0.059	0.071	0.055	0.126	0.050	0.049	0.052

Fig. 6. Integrated influence matrix T.

Table 7
Influence degree, influenced degree, centre degree, cause degree and ranking.

Factors	Influence degree		Influenced degree		Centre degree		Cause degree	
	R_i	Ranking	C_i	Ranking	E_i	Ranking	F_i	Ranking
A1	2.887	1	1.867	12	4.754	2	1.020	1
A2	2.634	2	1.744	15	4.378	5	0.890	2
A3	2.047	9	1.704	18	3.751	14	0.344	10
B1	1.796	15	1.316	26	3.112	21	0.480	6
B2	2.095	7	1.693	20	3.788	13	0.402	9
B3	1.952	12	1.703	19	3.654	16	0.249	13
B4	2.488	3	1.846	13	4.335	7	0.642	4
B5	1.951	13	1.476	23	3.427	18	0.475	7
C1	1.416	22	0.814	27	2.230	26	0.603	5
C2	1.866	14	2.055	8	3.921	12	−0.189	16
C3	2.162	6	2.022	9	4.184	8	0.140	14
C4	1.777	18	1.342	25	3.119	20	0.435	8
C5	1.782	16	1.672	21	3.453	17	0.110	15
C6	2.432	4	2.122	7	4.553	4	0.310	11
D1	1.280	23	0.557	28	1.836	28	0.723	3
D2	2.001	11	1.732	16	3.733	15	0.269	12
D3	2.092	8	2.889	1	4.981	1	−0.797	27
D4	1.486	21	1.793	14	3.278	19	−0.307	18
D5	1.747	20	2.319	4	4.066	9	−0.571	23
D6	0.972	26	1.909	11	2.881	23	−0.936	28
D7	1.756	19	2.284	5	4.040	10	−0.529	21
D8	1.179	24	1.914	10	3.092	22	−0.735	26
E1	2.025	10	2.319	3	4.345	6	−0.294	17
E2	1.002	25	1.722	17	2.725	24	−0.720	24
E3	2.183	5	2.548	2	4.731	3	−0.365	19
E4	0.941	27	1.491	22	2.432	25	−0.550	22
E5	0.738	28	1.466	24	2.204	27	−0.728	25
E6	1.777	17	2.147	6	3.924	11	−0.370	20

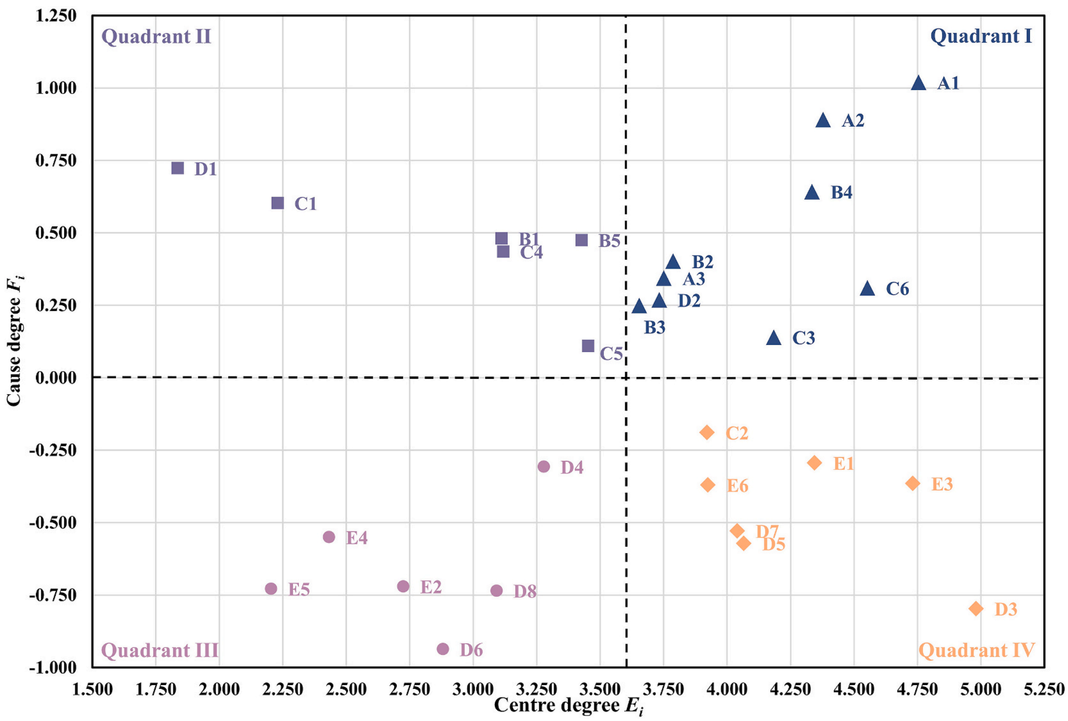

Fig. 7. "Centre degree - cause degree" causality diagram.

$$F_i = R_i - C_i \tag{14}$$

5. Discussion

5.1. Discussion of results

5.1.1. Centre degree analysis

The centre degree E_i is the sum of the influence degree R_i and the influenced degree C_i. It is the level of importance of influence factor i within the influence factor network. The greater the centre degree E_i of influence factor i is, the greater the degree of direct influence of influencing factor i on the SDS is. From Table 7 and Fig. 7, the top 10 influencing factors in terms of centre degree are 'Shared willingness (D3)', 'Data-sharing policies (A1)', 'Data values (E3)', 'Organisational management (C6)', 'Data-sharing regulations (A2)', 'Data quality (E1)', 'Data security technologies (B4)', 'Organisational motivation (C3)', 'Perceived risk (D5)' and 'Perceived benefits (D7)'. Their centre degrees are 4.981, 4.754, 4.731, 4.553, 4.378, 4.345, 4.335, 4.184, 4.066 and 4.040, respectively.

In terms of dimensions, among the above 10 influencing factors, two belong to 'Policy Environment (A)', one belongs to 'Information Technologies (B)', two belong to 'Research Organisations (C)', three belong to 'Scientific Researchers (D)' and two belong to 'Scientific Data (E)'. This also illustrates the substantial direct influence of internal factors such as researchers' willingness to share, perceived risks and perceived benefits as a prerequisite and necessary condition for SDS. Therefore, we believe that the construction and improvement of China's SDS system can be promoted effectively only if scientific researchers genuinely recognise SDS's value for scientific research, scientific innovation and even social progress. However, external factors such as the organisation to which a researcher belongs, the policy environment and the development of data technology also have an impact. This is because external factors have a high degree of influence on researchers' perceptions and judgements of the risks, costs and benefits associated with SDS, which can impact their willingness to share scientific data.

In terms of the magnitude of centre degree E_i, the influence factor that ranked first is 'Shared willingness (D3)', with a centre degree of 4.981, thereby showing a substantial divergence from the other influencing factors. Therefore, we believe that enhancing the willingness to share of researchers as much as possible is the key to promoting SDS. Moreover, five influencing factors, namely, 'Data-sharing policies (A1)', 'Data values (E3)', 'Organisational management (C6)', 'Data-sharing regulations (A2)' and 'Data quality (E1)', also have a high centre degree. All have a centre degree above 4.300, reflecting the importance of national policies, local regulations, department regulations and scientific data. Therefore, it is necessary to promote SDS and build a perfect SDS system from the national, local, departmental and data levels to form top-down cooperation.

5.1.2. Cause degree analysis

The cause degree F_i refers to the difference between the degree of influence R_i and the degree of influenced C_i, which means the degree of the contribution of influence i to the composition of the whole network.

(1) Influencing factor i with a cause degree $F_i > 0$ is a cause factor. The greater the cause degree F_i is, the more it can influence other influencing factors. From Table 7 and Fig. 7, the top five cause factors are 'Data-sharing policies (A1)', 'Data-sharing regulations (A2)', 'Personal characteristics (D1)', 'Data security technologies (B4)' and 'Organisational structure (C1)'. Their cause degrees are 1.020, 0.890, 0.732, 0.642 and 0.603, respectively. Notably, 'Data-sharing policies (A1)' and 'Data-sharing regulations (A2)' both belong to 'Policy Environment (A)' and have high cause and influence degrees. Their cause degrees are 1.020 and 0.890, and their influence degrees are 2.887 and 2.634, respectively, and they are ranked in the top two, much higher than the other influenceing factors. Meanwhile, combined with the direct influence matrix M, we find that the above two influencing factors have some degree of influence on all 27 factors except themselves, and their influence rate is as high as 100 %. It can be seen that the policy environment, such as national policies, local policies, industry systems and department regulations, can have a strong indirect influence on the comprehensive promotion of SDS. The main reason for this is that the policies, regulations and rules issued by government departments at all levels are not only the fundamental guarantee for promoting SDS programmes but also the 'motivation' and 'baton', which can play an essential role in guiding scientific researchers' willingness to share and improving the quality and value of scientific data.

(2) Influencing factor i with a cause degree $F_i < 0$ belongs to the result factor. The smaller the cause degree F_i is, the more it can be influenced by other influencing factors. From Table 7 and Fig. 6, the top five result factors are 'Perceived efforts (D6)', 'Shared willingness (D3)', 'Perceived attitudes (D8)', 'Data types (E5)' and Data descriptions (E2)'; their cause degrees are −0.936, −0.797, −0.735, −0.728 and −0.720, respectively. In terms of dimensions, among the above five influencing factors, three belong to 'Scientific Researchers (D)' and two belong to 'Scientific Data (E)'. It can be seen that internal factors such as researchers' willingness, attitude and the specifics of scientific data are highly influenced by external factors such as the policy environment, information technologies and organisations. Therefore, in the process of building an SDS system, Chinese government can try to promote SDS projects jointly by improving policies, upgrading technologies, optimising research organisations and other external factors, to enhance the degree of sharing, the scope of sharing and the practical path of SDS.

5.1.3. Key influencing factor identification

Referring to Table 7, we calculated the average of the total centre degree and total cause degree of all influencing factors. We divided the 'centre degree–cause degree' causality diagram (Fig. 7) into four quadrants according to the influencing factors.

(1) The influencing factors in Quadrant I are called ‘driving factors’, which include nine influencing factors, such as ‘Data-sharing policies (A1)’, ‘Data-sharing regulations (A2)’ and ‘Organisational management (C6)’. These factors have the most notable influence on SDS and can drive the formation of the entire SDS system effectively.
(2) The influencing factors in Quadrant II are called ‘supporting factors’, which include six influencing factors, such as ‘Data management mechanism (B5)’, ‘Organisational culture (C4)’ and ‘Organisational support (C5)’. Their influence is relatively weak, and these factors exert only a supporting influence in forming the SDS system.
(3) The influencing factors in Quadrant III are called ‘independent factors’, which include six influencing factors, such as ‘Subjective specification (D4)’, ‘Perceived attitudes (D8)’ and ‘Perceived efforts (D6)’. They have a high influenced degree and are the more important influenced factors in the entire system.
(4) The influencing factors in Quadrant IV are called ‘core problematic factors’, which include seven influencing factors, such as ‘Shared willingness (D3)’, ‘Data values (E3)’ and ‘Data quality (E1)’. These key factors are the most vulnerable to other factors in the entire system and have a substantial direct influence on the system.

To identify the key influencing factors, we combined the research results of Jiao et al. [39], He et al. [40], Li et al. [41] and Li et al. [42] with the essential characteristics of influencing factors in each quadrant. Subsequently, we formed a unified opinion after three rounds of discussion and determined the ‘driving factors’ and ‘core problematic factors’ to be the key influencing factors. Thus, we obtained 16 key influencing factors for SDS. The key influencing factors are shown in Table 8.

Based on the number of key influencing factors under each dimension and the average of centre degrees in Table 8, we can conclude that the optimisation path of China’s SDS project is ‘Scientific Researchers’ → ‘Scientific Data’ → ‘Policy Environment’ → ‘Research Organisations’ → ‘Information Technologies’. Because the project’s resources such as money, staff and time, are limited, referring to the above optimisation path can assist an SDS project in achieving better results. In addition, we provide some specific suggestions in the next section.

5.2. Management suggestions

5.2.1. Focus on researchers’ subjective willingness to share

Researchers are the producers and users of scientific data and are the main actors in SDS. Meanwhile, there are four key influencing factors in the dimension of ‘Scientific Researchers (D)’. Therefore, in promoting the development of SDS programmes in China, the Chinese government and research organisations should first focus on the researchers’ needs and enhance their subjective willingness to share scientific data. Specifically, they should (1) provide personal literacy training courses. Managers of research organisations can regularly organise training courses on information literacy, data literacy, privacy literacy, algorithmic literacy, etc., to help researchers improve their literacy in all aspects and to clarify the significance of SDS, the operation process and the expected benefits, thereby enhancing researchers’ subjective willingness to share. (2) Investigate the demand for incentives from researchers. Feedback and suggestions from researchers can effectively optimise the current reward system for SDS and enhance the perceived benefits for researchers. The government and research organisation managers can conduct research on researchers of different ages, disciplines and positions through online questionnaires, message boards and random emails. By tracking the feedback of researchers regularly, the government and research organisation managers can better understand the needs of researchers for a reward system and continuously improve and innovate the reward system related to SDS. (3) Carry out science popularisation activities on the risks of SDS. The government and research organisation managers can actively carry out SDS-related risk popularisation activities to help researchers better understand the potential risks and risk avoidance strategies in the SDS process and the controllable degree of the risks to reduce their perceived risks of SDS.

5.2.2. Enhancing the integrated governance of scientific data

There are three key influencing factors in the ‘Scientific Data (E)’ dimension, and all have a strong influence, with policy environment, data technology, organisation and researchers having a specific impact on them. While enhancing the willingness of researchers to share, the Chinese government and the managers of SDS projects should also actively improve the comprehensive governance of scientific data, optimise the governance and control strategies of scientific data and improve the quality, value and sensitivity of data to promote the development of SDS projects effectively. Specifically, they should (1) establish a scientific data quality evaluation system. Managers of SDS projects should establish a corresponding scientific data quality evaluation system according to the situation of a given project and carry out comprehensive monitoring, assessment and improvement of data quality in multiple dimensions, such as data completeness, operational timeliness, operational accuracy, operational stability and operational timeliness, to ensure a steady improvement in data quality. (2) Accelerate the development of data governance-related standards. The Chinese government and managers of SDS programmes should accelerate the development of data governance standards such as data authentication, citation and publication. The formulation of standards will not only lay a good foundation for data publication, data citation and data metrics but also give full play to scientific data’s potential economic, social and strategic value. (3) Introduce a dynamic data desensitisation system. SDS project managers can consider introducing the data dynamic desensitisation system developed by IBM, Oracle, Delphix and other companies into a given project to realise real-time desensitisation of all data queries and data calling results in the process of SDS, thereby reducing the sensitivity of data and the risk of privacy leakage effectively.

5.2.3. Fulfilling the role of policy support and guidance

The three influencing factors in the ‘Policy Environment (A)’ dimension are all key influencing factors, each with a wide range of

Table 8
Key factors influencing SDS.

Dimension	Key influencing factor	Number of factors	Average of centre degree
Policy Environment (A)	Data-sharing policies A1, Data-sharing regulations A2, Data-sharing standards A3	3	4.294
Information Technologies (B)	Platform-building technologies B2, Platform optimisation technologies B3, Data security technologies B4	3	3.926
Research Organisations (C)	Organisational climate C2, Organisational motivation C3, Organisational management C6	3	4.219
Scientific Researchers (D)	Personal literacy D2, Shared willingness D3, Perceived risk D5, Perceived benefits D7	4	4.205
Scientific Data (E)	Data quality E1, Data values E3, Data sensitivity E6	3	4.333

influence. Therefore, the Chinese government should emphasise optimising the data-sharing policy environment, establishing a top-down multilevel SDS policy system and playing a supportive and guiding role in policy. Specifically, the focus should rest on (1) improving macro-policies regarding SDS. Government agencies at all levels can introduce or improve macro-policies or laws related to SDS. For example, government agencies can introduce policies to clarify and delineate the responsibilities, rights and obligations of different types of institutions, such as higher education institutions, scientific research institutes, funding agencies and publishing agencies, in the process of SDS to promote the active participation of organisations and researchers in SDS. (2) Introduce supporting systems for SDS. All kinds of organisations can also introduce a series of supporting systems and measures under the guidance of macro-policies and follow their conditions. For example, the scientific research organisations can introduce data publication, incentive policies, intellectual property rights, the primary responsibility body, data security and other aspects of the regulations. Through the organic combination of the macro-policy and micro-system and the strict implementation of the policy system, the willingness of researchers to share can be enhanced effectively. (3) Formulate reasonable standards for SDS. The Chinese government should not neglect the standards and norms of SDS. The governmental science and technology authorities should formulate corresponding SDS standards for scientific data of different disciplines, types and purposes as well as optimise and improve the quality, value and sensitivity of data through data standardisation to lay a good foundation for the construction of an SDS platform and the implementation of an incentive system for SDS.

5.2.4. Strengthening the support of research organisations

There are three key influencing factors in the 'Research Organisations (C)' dimension, which substantially influence the key influencing factors in the 'Scientific Researchers (D)' dimension. It can be seen that although researchers are the main agents of SDS, research organisations can still have a substantial impact on SDS by influencing researchers. Therefore, strengthening the support of research organisations plays a vital role in promoting SDS projects. Specifically, it is essential to (1) create an excellent organisational climate. From the perspective of organisational behaviour and psychology, the organisational atmosphere can imperceptibly influence the behaviour and habits of the people in that organisation [43]. Therefore, managers of scientific research organisations can regularly organise team meetings or exchange meetings on SDS to create an excellent organisational atmosphere and help researchers establish and enhance SDS awareness. (2) Formulate diversified incentive policies. An organisation's incentive policy can influence researchers' perceived benefits and willingness to share. Therefore, managers of scientific research organisations should optimise the organisation's benefit distribution mechanism and formulate rich and diversified incentive policies. For example, when scientific data are published, research organisations should regard the impact of data as equivalent to the impact of papers and refer to bibliometric indicators to incorporate the number of downloads and citations of shared data into their scientific research appraisal evaluation systems. (3) Improve the management system of the organisation. Research organisations' managers should also improve research organisations' internal management systems for SDS, including the data-sharing processes, risk management systems, data property rights systems and sharing incentive systems, to ensure the steady progress of SDS projects.

5.2.5. Improving SDS platforms with information technology

The 'Information Technologies (B)' dimension also has three key influencing factors. With sufficient resources, the Chinese government should fully apply information technology to build a good SDS platform that promotes data sharing and academic innovation effectively. Specifically, the government should (1) introduce emerging platform-building technologies. The Chinese government should actively introduce emerging information technologies such as 5G, big data, blockchain, artificial intelligence and other emerging technologies in developing SDS platforms and focus on building a series of data-driven and intelligent services as the core of the SDS wisdom platform. For example, the Institute of Geographic Sciences and Natural Resources Research of the Chinese Academy of Sciences applied big data technology and distributed storage technology to develop a data-driven, independently distributed earth system SDS platform [44]. (2) Fully apply platform optimisation technology. Establishing an SDS platform should be centred on the actual needs of researchers and user experience. Therefore, during the development of the SDS platform, developers need to fully apply platform optimisation techniques such as search engine optimisation, URL address optimisation, adaptive design, interactive design and interface visual design to optimise the user experience of the platform's functions and services, the platform interface and the appearance design. (3) Strengthen the data security of the platform. Since a considerable amount of scientific data is stored on an SDS platform, its data security also deserves attention. In building the platform, the development team should use firewalls, authentication and authorisation, data encryption, data shielding, data erasure protocols and other data security technologies to enhance the platform's security and achieve data desensitisation while ensuring data security.

6. Conclusions

6.1. Research conclusions

Based on the existing SDS-related studies, we constructed a model for SDS influencing factors by combining grounded theory and information ecology theory and used the fuzzy-DEMATEL approach to explicitly analyse the relationships between the influencing factors and identify the key factors. The findings are as follows. (1) SDS is influenced by 28 factors in five dimensions, including the policy environment, information technologies, research organisations, scientific researchers and scientific data. Furthermore, the mutual influence and joint action among the influencing factors constitute a complex network system. (2) Of the 28 SDS influencing factors, 16 were key influencing factors, among which three belonged to the 'Policy Environment (A)' dimension, three to the 'Information Technologies (B)' dimension, three to the 'Research Organisations (C)' dimension, four to the 'Scientific Researchers (D)' dimension and three to the 'Scientific Data (E)' dimension. (3) The Chinese government, research organisation managers and SDS project developers can follow the path of 'Scientific Researchers' → 'Scientific Data' → 'Policy Environment' → 'Research Organisations' → 'Information Technologies' to promote SDS project development.

6.2. Research contributions

The main contributions of this study are listed as follows. (1) We combine grounded theory and information ecology theory to construct a relatively comprehensive and effective model for SDS influencing factors, demonstrating all the factors involved in the SDS process. (2) We adopt the fuzzy-DEMATEL approach to quantify the actual influence degree of each influencing factor in the SDS process and identify the key influencing factors, reducing the errors brought by the traditional DEMATEL approach. (3) Based on the results on the key SDS influencing factors, we summarise the optimisation path of the SDS programme and propose some suggestions to promote the better development of the SDS programme in China.

6.3. Limitations and future work

The main limitations of this research are as follows. (1) The number of expert samples is relatively small. In this study, we only invited eight experts on SDS for the expert scoring session. The insufficient number of experts may lead to significant errors in the results of expert scoring and make it difficult to ensure quality. We should increase the number of experts to make the research results more precise. In future studies, we will clarify the criteria for identifying experts and expand the number of experts to about 30. (2) The processing of expert scoring data needs to be optimised. In constructing the direct influence matrix M, we used the arithmetic mean of the eight experts' overall standardised influence degree values as the degree of influence between the variables. However, the direct use of averages to process these data may result in uninformative data. Therefore, in future studies, we will explore more effective ways of handling expert rating data. For example, we will combine the D–S evidence theory with the fuzzy-DEMATEL method to enhance uncertain information processing.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability statement

Has data associated with your study been deposited into a publicly available repository?
——No.
Has data associated with your study been deposited into a publicly available repository?
——The authors do not have permission to share data.

Ethics statement

This study was reviewed and approved by Ethics Committee of Lishui University on March 10, 2023, with the approval number: 2023TX006. Written informed consent was obtained from all participants for this study, and the participants did not include minors.

CRediT authorship contribution statement

Zhongyang Xu: Writing – review & editing, Writing – original draft, Methodology, Investigation. **Lingyu Liu:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Investigation, Conceptualization. **Zhiqian Meng:** Writing – review & editing, Writing – original draft, Methodology, Investigation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] The General Office of the State Council, The general Office of the state Council on the issuance of scientific data management methods notice [online], http://www.gov.cn/zhengce/content/2018-04/02/content_5279272.htm, 2020. (Accessed 1 February 2023).
[2] J. Pan, X. Zhang, The Fourth Paradigm: Data-Intensive Scientific Discovery, Science Press, Beijing, China, 2012, pp. 11–19.
[3] P. Du, Y. Li, X. Zhang, Research on conceptual domain and management boundary of data set in China's measures for scientific data management, Libr. J. 41 (4) (2022) 96–105+114, https://doi.org/10.13663/j.cnki.lj.2022.04.013 (In Chinese).
[4] J.T. Klein, H.J. Falk-Krzesinski, Interdisciplinary and collaborative work: framing promotion and tenure practices and policies, Res. Pol. 46 (6) (2017) 1055–1061, https://doi.org/10.1016/j.respol.2017.03.001.
[5] J.O. Hodonu-Wusu, A. Noorhidawati, A. Abrizah, Malaysian researchers on open data: the first national survey on awareness, practices and attitudes, Malays, J. Libr. Inf. Sci. 25 (2) (2020) 1–20, https://doi.org/10.22452/mjlis.vol25no2.1.
[6] L. Si, Y. Li, X. Zhuang, W. Xing, X. Hua, X. Li, J. Xin, An empirical study on the performance evaluation of scientific data sharing platforms in China, Libr, Hist. Technol. 33 (2) (2015) 211–229, https://doi.org/10.1108/LHT-09-2014-0093.
[7] R. Melero, C. Navarro-Molina, Researchers' attitudes and perceptions towards data sharing and data reuse in the field of food science and technology, Learn. Publ. 33 (2) (2020) 163–179, https://doi.org/10.1002/leap.1287.
[8] L. Zhang, L. Ma, Is open science a double-edged sword?: data sharing and the changing citation pattern of Chinese economics articles, Scientometrics 128 (2023) 2803–2818, https://doi.org/10.1007/s11192-023-04684-8.
[9] F. Maritsch, I. Cil, C. McKinnon, J. Potash, N. Baumgartner, V. Philippon, B.G. Pavlova, Data privacy protection in scientific publications: process implementation at a pharmaceutical company, BMC Med. Ethics 23 (1) (2022) 65, https://doi.org/10.1186/s12910-022-00804-w.
[10] P. Dorta-Gonzalez, S.M. Gonzalez-Betancor, M.I. Dorta-Gonzalez, To what extent is researchers' data-sharing motivated by formal mechanisms of recognition and credit? Scientometrics 126 (3) (2021) 2209–2225, https://doi.org/10.1007/s11192-021-03869-3.
[11] Y. Kim, S. Nah, Internet researchers' data sharing behaviors: an integration of data reuse experience, attitudinal beliefs, social norms, and resource factors, Online Inf. Rev. 42 (1) (2018) 124–142, https://doi.org/10.1108/OIR-10-2016-0313.
[12] W.Z. Moltgen, E. Akdeniz, A. Katsanidou, V. Nasshoven, E. Balaban, Factors influencing the data sharing behavior of researchers in sociology and political science, J. Doc. 74 (5) (2018) 1053–1073, https://doi.org/10.1108/JD-09-2017-0126.
[13] J.E. Raffaghelli, S. Manca, Exploring the social activity of open research data on ResearchGate: implications for the data literacy of researchers, Online Inf. Rev. 47 (1) (2022) 197–217, https://doi.org/10.1108/OIR-05-2021-0255.
[14] C. Borgman, J.C. Wallis, N. Enyedy, Research and Advanced Technology for Digital Libraries, Springer-Verlag, Berlin, Heidelberg, Germany, 2006, pp. 13–20.
[15] A. Bearth, M. Siegrist, Psychological factors that determine people's willingness-to-share genetic data for research, Clin. Genet. 97 (3) (2020) 483–491, https://doi.org/10.1111/cge.13686.
[16] Y. Kim, Fostering scientists' data sharing behaviors via data repositories, journal supplements, and personal communication methods, Inf. Process. Manag. 53 (4) (2017) 871–885, https://doi.org/10.1016/j.ipm.2017.03.003.
[17] Y. Kim, J.M. Stanton, Institutional and individual factors affecting scientists' data-sharing behaviors: amultilevel analysis, J. Assoc. Inf. Sci. Technol. 67 (4) (2016) 776–799, https://doi.org/10.1002/asi.23424.
[18] S. Searle, M. Wolski, N. Simons, J. Richardson, Librarians as partners in research data service development at Griffith University, Program-electron. Lib 49 (4) (2015) 440–460, https://doi.org/10.1108/PROG-02-2015-0013.
[19] W.D. Chawinga, S. Zinn, Global perspectives of research data sharing: a systematic literature review, Libr. Inf. Sci. Res. 41 (2) (2019) 109–122, https://doi.org/10.1016/j.lisr.2019.04.004.
[20] L.M. Gonzalez, T. Saorin, A. Ferrer-Sapena, R. Aleixandre-Benavent, F. Peset, Managing and disseminating research data, Prof. Inferm. 22 (5) (2013) 415–423, https://doi.org/10.3145/epi.2013.sep.06.
[21] K.A. Abdullahi, A. Noorhidawati, Attributes that influence academics' data sharing in Nigeria: the effects of organization culture, Inf. Res. 26 (3) (2021) 908, https://doi.org/10.47989/irpaper908.
[22] T. Devriendt, M. Shabani, K. Lekadir, P. Borry, Data sharing platforms: instruments to inform and shape science policy on data sharing? Scientometrics 127 (6) (2022) 3007–3019, https://doi.org/10.1007/s11192-022-04361-2.
[23] L.F. Sales, L.F. Sayao, The big and small science: analysis of differences in research data management, Inform. Soc-Estud. 29 (3) (2019) 151–170.
[24] C. Li, Y. Zhou, X. Zheng, Z. Zhang, L. Jiang, Z. Li, P. Wang, J. Li, S. Xu, S. Wang, Tracing the footsteps of open research data in China, Learn, Public 35 (1) (2022) 46–55, https://doi.org/10.1002/leap.1439.
[25] Y. Kim, A study of the roles of metadata standard and data repository in science, technology, engineering and mathematics researchers' data reuse, Online Inf. Rev. 45 (7) (2021) 1306–1321, https://doi.org/10.1108/OIR-09-2020-0431.
[26] Y. Zhu, Open-access policy and data-sharing practice in UK academia, J. Inf. Sci. 46 (1) (2020) 41–52, https://doi.org/10.1177/0165551518823174.
[27] B.G. Glaser, A.L. Strauss, The Discovery of Grounded Theory: Strategies for Qualitative Research, Aldine Transaction, Chicago, American, 1967, pp. 56–70.
[28] M. Wiesche, M.C. Jurisch, P.W. Yetton, H. Krcmar, Grounded theory methodology in information systems research, MIS Q. 41 (3) (2017) 685–701, https://doi.org/10.25300/misq/2017/41.3.02.
[29] I. Walsh, F. Rowe, BIBGT: combining bibliometrics and grounded theory to conduct a literature review, Eur. J. Inf. Syst. 32 (4) (2022) 653–674, https://doi.org/10.1080/0960085X.2022.2039563.
[30] A. Gabus, E. Fontela, World Problems, an Invitation to Further Thought within the Framework of DEMATEL, Battelle Geneva Research Centre, Geneva, Switzerland, 1972, pp. 1–8.
[31] G. Koca, O. Egilmez, O. Akcakaya, Evaluation of the smart city: applying the dematel technique, Telematics Inf. 62 (2021) 101625, https://doi.org/10.1016/j.tele.2021.101625.
[32] X. Xue, Y. Zhang, L. Zhang, Y. Wang, R. Hou, Evaluation on sustainable development of smart urban rail transit, Mobile Inf. Syst. 2022 (2022) 2737750, https://doi.org/10.1155/2022/2737750.
[33] Y. Zhai, B. Liu, Y. Wang, D. Zhao, Research on the influencing factors of medical and health big data assetisation from the perspective of data life cycle, J. Intell. 43 (1) (2024) 183–190, https://doi.org/10.3969/j.issn.1002-1965.2024.01.025 (In Chinese).
[34] S. Si, X. You, H. Liu, P. Zhang, DEMATEL technique: a systematic review of the state-of-the-art literature on methodologies and applications, Math. Probl Eng. 2018 (2018) 3695457, https://doi.org/10.1155/2018/3696457.
[35] B. Barghi, S. Sikari, Qualitative and quantitative project risk assessment using a hybrid PMBOK model developed under uncertainty conditions, Heliyon 6 (1) (2020) e03097, https://doi.org/10.1016/j.heliyon.2019.e03097.
[36] S. Du, X. Li, A new hybrid hermeneutic-fuzzy-DANP model based key influencing factors evaluation and ranking of soft science institute, Int. J. Fuzzy Syst. 25 (2023) 3016–3035, https://doi.org/10.1007/s40815-023-01553-7.
[37] C. Lou, Research on Information Ecosystem Theory and its Application, China Social Sciences Press, Beijing, China, 2014, pp. 10–19.
[38] C. Chen, Extensions of the TOPSIS for group decision-making under fuzzy environment, Fuzzy Set, Off. Syst. 114 (1) (2000) 1–9, https://doi.org/10.1016/s0165-0114(97)00377-1.
[39] J. Jiao, M. Wei, Y. Yuan, T. Zhao, Risk quantification and analysis of coupled factors based on the DEMATEL model and a bayesian network, Appl. Sci. 10 (1) (2020) 317, https://doi.org/10.3390/app10010317.
[40] Y. He, J. Kang, Y. Pei, B. Ran, Y. Song, Research on influencing factors of fuel consumption on superhighway based on DEMATEL-ISM model, Energy Pol. 158 (2021) 112545, https://doi.org/10.1016/j.enpol.2021.112545.

[41] H. Li, M. Tao, Z. Zuo, Analysis of the key influencing factors of the construction and development of new smart city in China based on DEMATEL method, Fresenius Environ. Bull. 29 (12) (2020) 10915–10923.
[42] H. Li, H. Dietl, J. Li, Identifying key factors influencing sustainable element in healthcare waste management using the interval-valued fuzzy DEMATEL method, J. Mater. Cycles Waste Manag. 23 (5) (2021) 1777–1790, https://doi.org/10.1007/s10163-021-01233-4.
[43] L. Chen, F. Luo, X. Zhu, X. Huang, Y. Liu, Inclusive leadership promotes challenge-oriented organizational citizenship behavior through the mediation of work engagement and moderation of organizational innovative atmosphere, Front. Psychol. 11 (2020) 560594, https://doi.org/10.3389/fpsyg.2020.560594.
[44] Xinhua News Agency, CAS releases earth big data sharing service platform [online], Available at: http://www.xinhuanet.com/politics/2019-01/15/c_1210038890.htm, 2019. (Accessed 23 March 2023).