

## **Research Data Management Frameworks: A Systematic Literature Review**

### **Mohammad Azami**

Associate prof., Medical Library and Information Science Department, Faculty of Management and Medical Information Sciences, Kerman University of Medical Sciences, Kerman, Iran.

ORCIDiD: <https://orcid.org/0000-0002-8971-5971>  
[mohammadaazami14@gmail.com](mailto:mohammadaazami14@gmail.com)

### **Ali Sadatmoosavi**

Assistant Prof., Medical Library and Information Science Department, Faculty of Management and Medical Information Sciences, Kerman University of Medical Sciences, Kerman, Iran.

ORCID iD: <https://orcid.org/0000-0001-6800-3345>  
[moosavi56@gmail.com](mailto:moosavi56@gmail.com)

### **Mohamad Reza Chashmyazdan**

Student in Medical Library and Information Science, Faculty of Management and Medical Information Sciences, Kerman University of Medical Sciences, Kerman, Iran.

ORCID iD: <https://orcid.org/0000-0003-2558-5319>  
Corresponding Author: [m.chashmyazdan@yahoo.com](mailto:m.chashmyazdan@yahoo.com)

Received: 01 April 2022

Accepted: 02 July 2022

### **Abstract**

Several models and frameworks have been proposed by researchers for the management of research data over the past few years. These models and frameworks may help organizations and researchers evaluate various aspects of implementing research data management systems. Thus, this study aims to identify the principal dimensions and necessary practices of research data management based on a systematic review. A systematic search of PubMed, Web of Science, Scopus, and LISTA databases was conducted by October 2021 to identify relevant studies. Two authors screened all retrieved articles in three stages (title, abstract, and full text). Based on the refined studies, data were extracted to fulfil the objectives and answer the research questions. According to the analysis of articles, research data management can be categorized into four main dimensions: data, researchers, organizations, and technology. Data processing, data protection, data legal issues, and data sharing are the main practices of the data dimension. The researcher's dimension is educating, changing the mindset, and motivating researchers. On the organizational dimension, the essential practices supply human resources, policy-making, interaction, and support. On the technology dimension, the most significant practices include: Establishing software and hardware infrastructures. This study demonstrates that research data management is a systematic process involving various stakeholders and facilities, each cooperating and coordinating various activities. To manage research data, researchers, policymakers, research organizations, and trained staff are among the most important participants, and technical facilities are vital to its success.

**Keywords:** Research Data Management, Data Sharing, Data Management, Dimension Identifying, Systematic Literature Review.

### Introduction

Research data is considered a valuable resource and requires significant time and investment. A range of data is collected and analyzed to produce and justify conclusions in the process of research (Singh, Monu & Dhingra, 2018). With the development of information technology, research processes have changed, and electronic research has become possible. As a result, science and technology activities generate more data than ever. Research data in itself can represent unique and considerable value. Some argue that scientific data has become one of the most important outputs of technology and science activities (Wilms, Stieglitz, Ross & Meske, 2020; Zhao, Peng & Huo, 2009). Due to the volume of data generated, these resources must be managed effectively. As with any non-renewable resource, the more effectively data is managed, the greater the value it generates (Laughton & Du Plessis, 2013).

Much research data can be used beyond its original purposes, such as defining novel research areas and preventing duplication and overlap. Therefore, with the potential for data reuse and the growth of data-driven research, increasing attention is being paid to the governance and management of data (Krahe, Toohey, Wolski, Scuffham & Reilly, 2020). In this context, research data management (RDM) is recognized internationally as a significant factor in scientific advancement (Redkina, 2019) and an increasingly important priority for universities (Gunjal & Gaitanou, 2017). There is a growing movement among universities to address this strategic goal. Many organizational stakeholders, like academic libraries, have developed policies and infrastructure to support this endeavour (Cox, Kennan, Lyon & Pinfield, 2017).

RDM can be defined as the process of managing, protecting and facilitating the use of research data, from the point of creation through to its use and eventual reuse. RDM allows research results to be independently validated, expands research scope and visibility, and maximizes research investment by preventing duplication (Borghi & Van Gulick, 2021). It also opens opportunities for researcher collaborations from different countries and institutions. RDM enhances data validation and transparency, increases accessibility, encourages data reuse and sharing and ensures data privacy (Goben & Griffin, 2019; Tripathi & Pandey, 2018). Researchers can interrogate existing data to ask new questions and use new analytical techniques to address significant scientific and social concerns (Mayernik et al., 2012).

The management of research data represents a major challenge for research organizations. A significant amount of research data in various forms is generated by universities, creating so-called "volume," "diversity," and "speed" challenges (Pinfield, Cox & Smith, 2014). It is essential to build technical, communication, personnel and policy capabilities to manage such data at the project and organizational levels. But research data management requires more detailed instructions for research projects and institutions (Qin, Crowston & Kirkland, 2017).

A range of approaches to RDM has been developed in universities and research organizations, such as the open central database, distributed network of approved users, and networking with peer-to-peer sharing at local, institutional and national levels (Arend et al., 2014; Goban & Griffin, 2019). It created a confusing environment for various stakeholders, including support staff and universities. Academic libraries are one of the main participants in the RDM space, particularly in providing research data services (Pinfield et al., 2014). Libraries' provision of organizational support for RDM reflects previous experience developing digital services and open-access repositories. Nevertheless, RDM presents new challenges for librarians (Cox et al., 2017).

Researchers have recently proposed various models and frameworks for managing research data (Jeng, He & Oh, 2016; Laughton & Du Plessis, 2013; Tripathi & Pandey, 2018). Organizations and researchers may be better equipped to understand the various aspects of implementing a research data management system by adopting these frameworks. While numerous frameworks and models are available, the UK data archive introduced only one approved international standard organization model for data for research data management. This model has six steps: data creation, data processing, data analysis, data protection, data access, and data reuse, referred to as the data lifecycle model (Krahe et al., 2020; Laughton & Du Plessis, 2013). Nevertheless, this model poses several challenges and problems, such as providing details of low-level processes and being closed and one-way, which were its two most significant disadvantages (Cox & Tam, 2018).

Since there is no general agreement at the international level, academic institutions have adopted multiple models and frameworks for managing research data according to their circumstances and context (Arend et al., 2014; Patel, 2016; Shakeri & Gracy, 2014). By analyzing the existing literature on RDM, this paper aims to guide librarians and university administrators. This study aims to gather the available evidence through a systematic literature review. A systematic literature review collates all relevant studies on a research question to synthesize the available evidence (Tricco et al., 2015). One of the significant objectives of review studies is to summarise and interpret the research results (Arksey & O'Malley, 2005). Therefore, this study combines the available evidence through a systematic literature review to identify the major dimensions of research data management and establish the necessary practices for effective research data management.

### **Research Questions**

In this article, we attempt to answer the following research questions:

- 1) What are the main components and dimensions of research data management?
- 2) What are the main practices and activities associated with each dimension of research data management?

### **Materials and Methods**

This systematic literature review examines the dimensions and practices of research data management. The literature search was conducted using electronic databases, and retrieved articles were screened for inclusion. Contents of retrieved articles were analyzed using a thematic analysis approach in conjunction with the above research questions. This study followed the guidelines for "preferred practices in reporting systematic reviews and meta-analyses" (Moher, Liberati, Tetzlaff, Altman & Group, 2009).

### **Inclusion and exclusion criteria**

Articles included in this study needed to meet the following criteria:

1. Must be scientific articles
2. It must have been published in English
3. Full text must be available
4. Articles must be published after 2000
5. Articles must propose a new or revised model or framework for research data management

6. The dimensions, processes, and practices of the framework or models must be clearly outlined and extractable.

Articles were excluded if they were letters, editorials, books, retracted publications and workshops.

### Data source and search strategy

A systematic search was conducted in October 2021 across four databases: LISTA, Web of Science, PubMed, and Scopus. Manual searches were also conducted in Google Scholar. After screening the retrieved results, references and citations of included articles were also reviewed, and sources relevant to these articles were added to the study. Keywords were identified after reviewing relevant sources and performing an initial search to improve the relevance of the search results. There were no restrictions on the publication type or timeframe. The search strategy is summarised in Table 1.

Table 1  
Search strategies

Database	Search string	Results
Web of Science	TS= ("Research Data" OR RDM OR "Data Management" OR "Research Management" OR "data curation" OR "Open Data" OR "Data Sharing" OR "data reus*" OR "data reutilization" OR "data Re*usage") AND TI=(Model* OR Pattern OR prototype OR template OR framework)	2893
Scopus	TITLE-ABS-KEY ("Research Data" OR RDM OR "Data Management" OR "Research Management" OR "data curation" OR "Open Data" OR "Data Sharing" OR "data reus*" OR "data reutilization" OR "data Re*usage") AND TITLE-ABS-KEY (Model* OR Pattern OR prototype OR template OR framework)	3718
PubMed	("Research Data"[Title/Abstract] OR RDM[Title/Abstract] OR "Data Management"[Title/Abstract] OR "Research Management"[Title/Abstract] OR "data curation"[Title/Abstract] OR "Open Data"[Title/Abstract] OR "Data Sharing"[Title/Abstract] OR "data reus*" [Title/Abstract] OR "data reutilization"[Title/Abstract] OR "data Re*usage"[Title/Abstract]) AND "model*" [Title/Abstract] OR "Pattern"[Title/Abstract] OR "prototype" [Title/Abstract] OR "template"[Title/Abstract] OR "framework" [Title/Abstract]	4031
LISTA	S1 TI ("Research Data" OR RDM OR "Data Management" OR "Research Management" OR "data curation" OR "Open Data" OR "Data Sharing" OR "data reus*" OR "data reutilization" OR "data Re*usage") OR KW ("Research Data" OR RDM OR "Data Management" OR "Research Management" OR "data curation" OR "Open Data" OR "Data Sharing" OR "data reus*" OR "data reutilization" OR "data Re*usage") S2 TI (Model* OR Pattern OR prototype OR template OR framework) OR KW (Model* OR Pattern OR prototype OR template OR framework) S3 S1 AND S2	738

### Study selection

The search results were entered into EndNote v.X8 software for review and screening. In the first step, duplicate articles were removed. This was followed by examining the titles and

abstracts of the remaining articles to determine if the topic was relevant to the research objectives. Selected articles were screened according to the inclusion and exclusion criteria. Finally, the references of all included articles were searched to identify additional research. Two researchers (MC and AS) independently reviewed all the articles in three stages (title, abstract, and full-text screening). Disagreements between the two researchers were resolved by discussion between the two. A third researcher (MA) was consulted if the two researchers could not reach a consensus.

### **Data collection process**

Data was extracted from refined studies to achieve the research objectives. First, a data extraction form was designed, and the data from two of the included articles were extracted and entered into the form. The form was then revised and used for the remaining studies. Two researchers independently extracted the data based on four sections: bibliographic information, methodology, dimensions, components and practices presented in each model, and key findings. The included studies were examined separately through a thematic analysis approach.

### **Synthesis method**

A thematic analysis was conducted to identify and categorize the dimensions of research data management. Thematic analysis is an inductive method of analyzing qualitative data using open coding to identify and describe a specific phenomenon. MAXQDA-12 software was used to analyze the extracted data. The thematic analysis process involved six stages: (1) familiarisation with data, (2) generating primary codes, (3) using codes to find themes, (4) reviewing the themes, (5) defining and naming themes, and (6) creating the report.

In this study, research data management frameworks were regarded as broader themes, and the included studies were coded to identify the associated dimensions, components, and practices. The research results were synthesized and qualitatively described, providing an overall picture of the different aspects of research data management.

### **Results**

The first search yielded 11,748 articles through scientific databases and manual searches. Of these, 7,511 articles were reviewed after removing duplicates. In the second stage, 5,429 articles were excluded due to their lack of relevance based on their titles. After reviewing the abstracts, the full text of 228 articles was reviewed, with three additional articles added through backward and forward citation searches. Finally, after applying the inclusion and exclusion criteria, 23 studies were eligible for this study. Figure 1 illustrates the steps involved in searching and selecting articles.

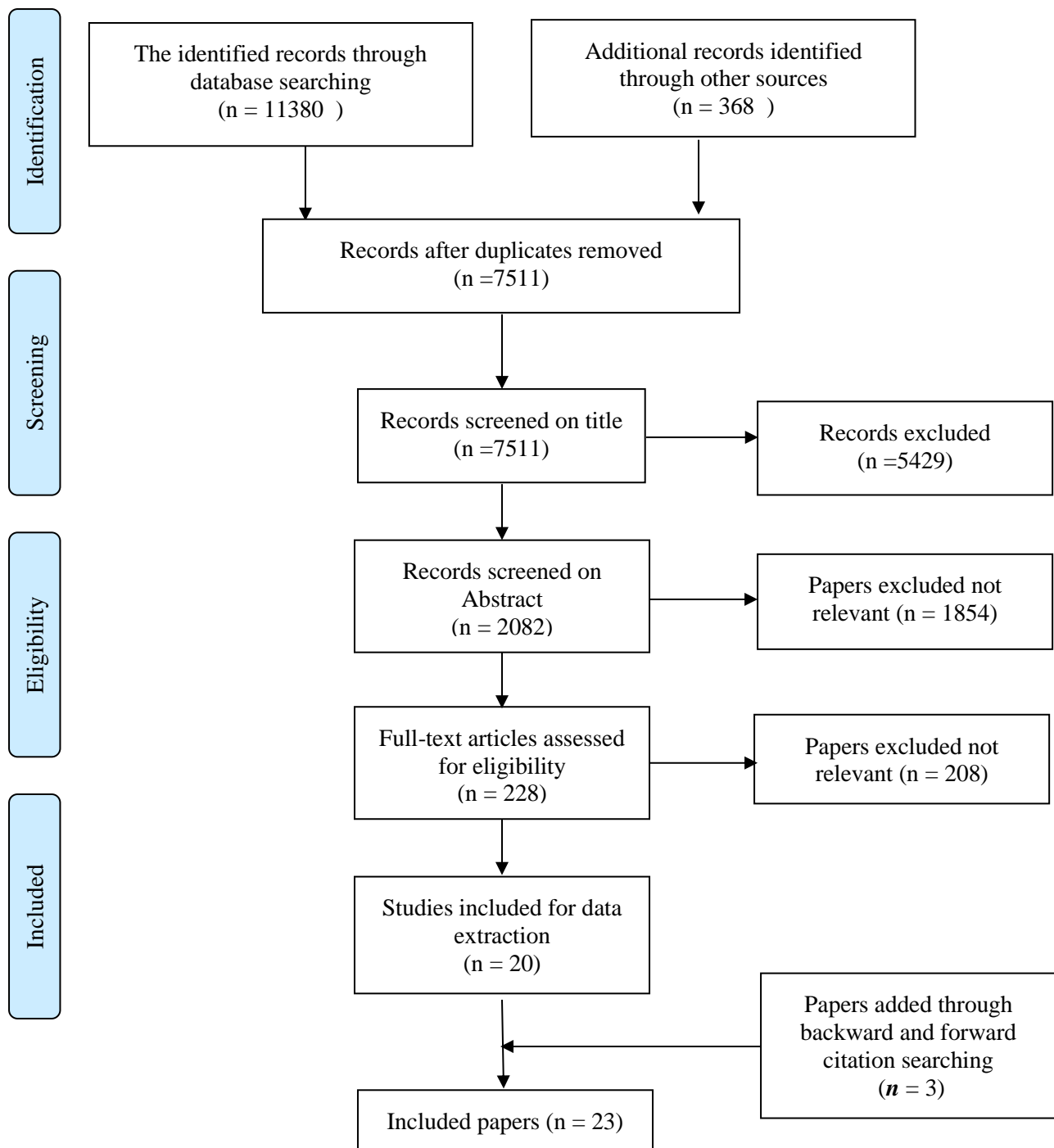


Figure 1: PRISMA (Preferred reporting items for systematic reviews and meta-analyses) flow diagram

### Characteristics of included studies

Methodologically, several research methods have been used to develop a model of RDM. Of the 23 articles, seven were written using a review method, six used mixed methods, four were conceptual papers, two were case studies, two were surveys, one was qualitative, and one was a cross-sectional study. Articles were published between 2008 and 2020, with the highest number appearing in 2014. Most articles were published in the United States and India (five per country), and one study was conducted across multiple countries (Table 2).

Table 2  
 Characteristics of included studies

ID	Source (Year)	Country	Design/Methods	Model context	Model dimensions
1	Higgins (2008)	United Kingdom	Conceptual paper	International level	Conceptualize; create and receive; appraise and select; ingest; preservation action; store; access, use and reuse; transform
2	Yunus, et al. (2009)	Malaysia	Survey	Social science	Data creation; data analysis/use; data output/ maintenance; data preservation; data distribution; data reuse; policies/acts; people; technology
3	Crowston, et al. (2011)	USA	Literature review	Institutional level	Data acquisition, processing and quality assurance; data description and representation; data dissemination; repository services/ preservation
4	Mayernik, et al. (2012)	USA	Literature review, case study	Disciplinary level and institutional level	Data conservancy; technical framework; organizational framework; sustainability strategies
5	Laughton, et al (2013)	South Africa	Case study and interview	International level	Pre ingest; ingest; archival storage; data management; administration; preservation planning; access
6	Arend, et al. (2014)	Germany	Case study, system analysis	Institutional level	Version management; metadata; information retrieval; data publication and citation; data security; generality; technology
7	Dillo, et al. (2014)	Netherlands	Practice paper	National level	Basic technical infrastructure; back-office functions; front office functions
8	Fechera, et al. (2014)	Germany	Systematic review	International level	Data donor; research organization; research community; norms; data infrastructure; data recipients
9	Ndume, et al. (2014)	South Africa	Cross-sectional study	Field level	Technical components; social components
10	Pinfield, et al. (2014)	United Kingdom	Qualitative study	Institutional level	Strategies; policies; guidelines; processes; technologies; services
11	Shakeri & Gracy (2014)	USA	Case study	Small sciences	Data characteristics; data management; data management practices
12	Jeng, et al. (2016)	USA	Case study, survey	Social sciences	Data characteristics; technical infrastructure; organizational and research context; individual characteristics and motivations; research product sharing practices
13	Patel (2016)	India	Conceptual paper	Institutional level	Data management; data storage and hosting; data usage
14	Yves Burgi, et al. (2016)	Switzerland	Literature review and qualitative study	National level	Initial data and workflow; analysis and data exploration; publication, archiving and long-term data management
15	Cox, et al. (2017)	Multi-country	Survey	Academic libraries	Basic level; developing level; extensive level
16	Gunjal, et al. (2017)	India	Literature review	Institutional level	Data creation and deposit; managing active data; data repositories and archives; data catalogues and registries
17	Huh, et al. (2017)	South Korea	Literature review	National level and field (ecology, environment,	Entering and validating of datasets; metadata management; search functions for datasets and data; management of controlled vocabulary; EML conversion of

ID	Source (Year)	Country	Design/Methods	Model context	Model dimensions
				and biodiversity) level	metadata; data conversion and synthesis; data visualization; management of scalable data; user convenience function.
18	Qin, et al. (2017)	USA	Literature review	Project level and community level	Data management in general; data acquisition, processing, and quality assurance; data description and representation; data dissemination; repository services and preservation
19	Tripathi, et al. (2017)	India	Survey, observation study	National level	Develop RDM plans; research data management policies; collaboration of the library and computer department; metadata and appropriate format; confidentiality and sensitivity of data; publish and transfer data to data repositories
20	Kumar Singh, et al. (2018)	India	Literature review	Institutional level	Framing data sharing policy; motivate the researcher; accessibility of infrastructure; orientation programme; costing; well trained staff
21	Shrestha, et al. (2018)	Canada	Literature review	International level	Data owner; smart contract; data repository
22	Tripathi, et al. (2018)	India	Conceptual paper	Institutional level	Data management; data storage and hosting; data usage
23	Van Wyk, et al. (2020)	South Africa	Empirical and nonempirical	University level	Human components layer; hardware components layer; software components layer; management services component layer; standards, specification and protocols layer; policy components layer; research lifecycle and research data lifecycle

### Context of models

Models developed in the studies had been applied to various contexts and levels. Generally, the models can be classified into five levels. Most studies (n = 9) were conducted at the institutional level. Five studies developed a model for data management within a specific field, three for data management within a country, one for data management at the project level, and five for all levels.

### Analyzing models and identifying dimensions of research data management

The four dimensions of research data management identified in the literature can be summarised as data, researchers, organizations, and technology, with 12 subcategories described in Table 3.

Table 3  
Dimensions and practices identified from model analysis

Dimensions	Practices	Activities / References
Data	Data processing	<p><b>Acquisition and transfer data</b> (Crowston &amp; Qin, 2011; Dillo &amp; Doorn, 2014; Higgins, 2008; Laughton &amp; Du Plessis, 2013; Qin et al., 2017; Van Wyk, Bothma, &amp; Holmner, 2020)</p> <p><b>Validation (quality assessment)</b> (A. M Cox et al., 2017; Crowston &amp; Qin, 2011; Dillo &amp; Doorn, 2014; Fecher, Friesike, &amp; Hebing, 2015; Higgins, 2008; Huh, Park, Ahn, Hwang, &amp; Jung, 2017; Laughton &amp; Du Plessis, 2013; Pinfield et al., 2014; Qin et al., 2017; Shakeri &amp; Gracy, 2014; Van Wyk et al., 2020)</p> <p><b>Vocabulary control</b> (Burgi, Blumer, &amp; Makhlof-Shabou, 2017; A. M Cox et al., 2017; Huh et al., 2017; Laughton &amp; Du Plessis, 2013; Pinfield et al., 2014; Shakeri &amp; Gracy, 2014; Van Wyk et al., 2020)</p> <p><b>Metadata allocation</b> (Arend et al., 2014; Burgi et al., 2017; A. M Cox et al., 2017; Crowston &amp; Qin, 2011; Fecher et al., 2015; Higgins, 2008; Huh et al., 2017; Laughton &amp; Du Plessis, 2013; Pinfield et al., 2014; Shakeri &amp; Gracy, 2014; Van Wyk et al., 2020; Yunus &amp; Johare, 2009)</p>
	Data protection	<p><b>Data security</b> (Arend et al., 2014; Burgi et al., 2017; Crowston &amp; Qin, 2011; Fecher et al., 2015; Higgins, 2008; Huh et al., 2017; Ndume, NKANSAH-GYEKYE, Ko, &amp; Selemani, 2014; Patel, 2016; Shakeri &amp; Gracy, 2014; Shrestha &amp; Vassileva, 2018; D. Tripathi &amp; Pandey, 2018; Van Wyk et al., 2020; Yunus &amp; Johare, 2009)</p> <p><b>Suitable and safe format</b> (Burgi et al., 2017; A. M Cox et al., 2017; Crowston &amp; Qin, 2011; Dillo &amp; Doorn, 2014; Fecher et al., 2015; Gunjal &amp; Gaitanou, 2017; Patel, 2016; Qin et al., 2017; Shakeri &amp; Gracy, 2014; M. Tripathi, Shukla, &amp; Sonkar, 2017; Van Wyk et al., 2020; Yunus &amp; Johare, 2009)</p> <p><b>Data backup</b> (Crowston &amp; Qin, 2011; Dillo &amp; Doorn, 2014; Huh et al., 2017; Laughton &amp; Du Plessis, 2013; Patel, 2016; D. Tripathi &amp; Pandey, 2018; M. Tripathi et al., 2017)</p>
	Legal issues	<p><b>Data ownership</b> (A. M Cox et al., 2017; Fecher et al., 2015; Jeng et al., 2016; Laughton &amp; Du Plessis, 2013; Patel, 2016; Pinfield et al., 2014; Shakeri &amp; Gracy, 2014; D. Tripathi &amp; Pandey, 2018)</p> <p><b>Determining the sensitivity and confidentiality of data</b> (Fecher et al., 2015; Huh et al., 2017; Jeng et al., 2016; Laughton &amp; Du Plessis, 2013; Patel, 2016; D. Tripathi &amp; Pandey, 2018; M. Tripathi et al., 2017)</p> <p><b>Copyright and license to use the data</b> (Fecher et al., 2015; Patel, 2016; Shakeri &amp; Gracy, 2014; D. Tripathi &amp; Pandey, 2018; Van Wyk et al., 2020)</p> <p><b>Data Citation</b> (Arend et al., 2014; Burgi et al., 2017; A. M Cox et al., 2017; Gunjal &amp; Gaitanou, 2017; M. Tripathi et al., 2017)</p> <p><b>Data anonymization</b> (Fecher et al., 2015; Patel, 2016; M. Tripathi et al., 2017)</p>
	Data sharing	<p><b>Determine the source of data</b> (Crowston &amp; Qin, 2011; Gunjal &amp; Gaitanou, 2017; Jeng et al., 2016; Patel, 2016; D. Tripathi &amp; Pandey, 2018)</p> <p><b>Create data access</b> (Burgi et al., 2017; Crowston &amp; Qin, 2011; Huh et al., 2017; Jeng et al., 2016; Patel, 2016; Qin et al., 2017; Shakeri &amp; Gracy, 2014; Van Wyk et al., 2020)</p>
Researchers	Researchers training	<b>Training management and sharing processes</b> (Burgi et al., 2017; A. M Cox et al., 2017; Fecher et al., 2015; Laughton & Du Plessis, 2013; Pinfield et al., 2014; M. Tripathi et al., 2017)

Dimensions	Practices	Activities / References
	Motivate researchers	<b>Encouragement to data sharing</b> (Crowston & Qin, 2011; Jeng et al., 2016; Patel, 2016; Singh et al., 2018; Van Wyk et al., 2020)
	Changing the mindset of researchers	<b>Changing the mindset of researchers</b> (Crowston & Qin, 2011; D. Tripathi & Pandey, 2018)
Organization	Supply human resources	<b>Skilled human resources</b> (Jeng et al., 2016; Mayernik et al., 2012; Ndume et al., 2014; Singh et al., 2018) <b>Empowerment of human resources</b> (Jeng et al., 2016; M. Tripathi et al., 2017)
	Policy making	<b>Data sharing policy</b> (A. M Cox et al., 2017; Fecher et al., 2015; Higgins, 2008; Patel, 2016; Pinfield et al., 2014; Qin et al., 2017; Singh et al., 2018; D. Tripathi & Pandey, 2018; M. Tripathi et al., 2017; Yunus & Johare, 2009) <b>Data management plan</b> (Burgi et al., 2017; Gunjal & Gaitanou, 2017; Pinfield et al., 2014; Shakeri & Gracy, 2014) <b>Data protection policies</b> (Higgins, 2008; Laughton & Du Plessis, 2013) <b>Technology policies and procedures</b> (Ndume et al., 2014)
	Interaction and support	<b>Technical support</b> (A. M Cox et al., 2017; Fecher et al., 2015; Jeng et al., 2016) <b>Interaction and cooperation with other centers</b> (Fecher et al., 2015; Gunjal & Gaitanou, 2017; Laughton & Du Plessis, 2013; D. Tripathi & Pandey, 2018) <b>Guarantee compliance with standards</b> (Crowston & Qin, 2011) <b>Handle users' questions and complaints</b> (Laughton & Du Plessis, 2013)
Technology	Creating hardware infrastructure	<b>High performance computing facilities</b> (Ndume et al., 2014; Van Wyk et al., 2020) <b>Hardware infrastructure</b> (Arend et al., 2014; A. M Cox et al., 2017; Crowston & Qin, 2011; Fecher et al., 2015; Gunjal & Gaitanou, 2017; Higgins, 2008; Laughton & Du Plessis, 2013; Mayernik et al., 2012; Ndume et al., 2014; Patel, 2016; Qin et al., 2017; Shrestha & Vassileva, 2018; Singh et al., 2018; D. Tripathi & Pandey, 2018; Van Wyk et al., 2020; Yunus & Johare, 2009) <b>Data repositories</b> (Fecher et al., 2015) <b>Network infrastructure</b> (Pinfield et al., 2014; Van Wyk et al., 2020)
	Creating software infrastructure	<b>Data hosting service</b> (A. M Cox et al., 2017; Patel, 2016; D. Tripathi & Pandey, 2018; Van Wyk et al., 2020) <b>Software infrastructure</b> (Burgi et al., 2017; A. M Cox et al., 2017; Crowston & Qin, 2011; Fecher et al., 2015; Gunjal & Gaitanou, 2017; Higgins, 2008; Laughton & Du Plessis, 2013; Mayernik et al., 2012; Ndume et al., 2014; Patel, 2016; Pinfield et al., 2014; Qin et al., 2017; Shrestha & Vassileva, 2018; Singh et al., 2018; D. Tripathi & Pandey, 2018; Van Wyk et al., 2020)

## Data

One of the most critical dimensions of research data management is the data. This dimension includes practices and activities related to data preparation for management and sharing. This main category contains four subcategories: data processing, data protection, legal issues, and data sharing. In total, 14 studies discussed the data processing phase and its subsets. Nineteen studies focused on data protection and its subsets. The legal implications of data were addressed in 15 studies, and 10 addressed data sharing and its subsets.

### **Researchers**

The researchers dimension included educating, motivating, and changing their mindset. In total, six studies investigated the issue of educating researchers and its subsets, five focused on motivating the researchers, and two discussed the possibility of changing the mindset of researchers.

### **Organizations**

Organizations were identified as the third dimension of research data management. Three primary practices were identified within this dimension: human resource supply, policy-making, and support and interaction. Five studies addressed human resource issues or subsets. Fifteen studies examined policy-making and its subsidiaries, and seven focused on the role of support and interaction within the organization.

### **Technology**

Technology was the fourth dimension in this study, containing two primary practices, including creating software and hardware infrastructure. Seventeen studies addressed hardware and its subsets, and 18 focused on software and its subsets.

## **Discussion**

This study was conducted to identify the dimensions and practices needed for managing research data. An analysis of 23 articles revealed different frameworks and approaches to managing research data. Different organizational policies, plans, priorities, and technical infrastructure may cause differences between organizations or countries. Despite this, a summary of existing frameworks may be helpful for policymakers, librarians, and research institutions. We have aimed to present the dimensions of RDM as a comprehensive set. However, it should be acknowledged that RDM is a multidimensional issue, which should be considered in all its dimensions. Therefore, the topics used were categorized into four main topics, the importance, role, and position of each discussed below.

Data has been identified as the central dimension of research data management. All research data management systems aim to share the data (Pinfield et al., 2014). The data dimension includes four practices: processing data, protecting them, resolving legal matters, and sharing data through the RDM system/facilitating its use. Data processing practices convert them into suitable output formats. Arend Lange, Chen, Colmsee, Flemming, Hecht and Scholz (2014) suggest that well-described data are more valuable than data obtained from articles. A well-defined metadata schema is the most effective way to create and manage data sets (Zhao et al., 2009). Institutions must store data that is valuable to them due to the cost of storing it. Hence, data selection and evaluation is one of the most basic data processing practices (Osareh et al., 2021). Burgi, Blumer and Makhlof-Shabou (2017) argue that some data cannot be reproduced or are very expensive, so the value of data must be considered when storing it. Therefore, the processing methods must be carefully documented to ensure the data's validity and integrity.

Data protection practices must ensure that data remains valid, reliable, and usable while maintaining integrity (Higgins, 2008). This practice should also ensure the security of data. Data misuse has always been an issue when researchers share data, as various studies have demonstrated (Burgi et al., 2017; Campbell et al., 2002; Ismail, Xu, Wood & Welch, 2013;

Tenopir et al., 2011). Access to data over the long term also depends on the file format. According to Tripathi and Pandey, open file formats are useful since they do not have a dedicated format or data lock. The open file format helps to recover data if the file format becomes obsolete (Dillo & Doorn, 2014; Tripathi & Pandey, 2018).

The practice of legal issues of data in research data management is the most challenging (Hashim, 2019). The copyright and data ownership issues are extremely complex because they involve researchers, data collectors, data analysts, universities or research institutes, and funding agencies. Data ownership and copyrights should be clearly defined when sharing data (Jeng et al., 2016; Tripathi & Pandey, 2018). The presence of copyright statements and unambiguous licenses is essential for a research data management system. According to Patel, one approach is that any research project should clearly define the terms and conditions of data ownership. Authors must provide proof of ownership of the copyright to the data before the data are made available in the repositories (Patel, 2016).

Data-sharing practices are associated with identifying the data source and creating data access measures. Identifiable data sources can address confidentiality issues or ambiguity concerning data ownership. Alternatively, this could be viewed as a protocol for citing raw data and facilitating the citation process (Gunjal & Gaitanou, 2017; Jeng et al., 2016). Patel (2016) discusses that it is important for every data set to be linked to the methodology through which the data has been collected, the resources consulted and the limitations, if any. This will enable the subsequent researcher to understand the limitations of the data set before it is used (ibid).

We suggest that research institutes and policymakers follow a standardized protocol for collecting and validating data and provide long-term data access, readability, usability, retrieval, search, and filtering metadata.

The second dimension of research data management identified in this study is researchers. In this dimension, the training of researchers is the first practice. Burgi et al. (2017) note that research data management is a complex process, which researchers are unaware of when working with their data. Researchers should be aware of these issues through training and assistance from information professionals and/or data operators to better manage data (ibid). Various methods have been proposed for educating users regarding the reviewed documents, including links to case studies, training materials, guides, DMP tools, consultation groups, and checkpoints for researchers. Additionally, workshops, face-to-face interactions, and special training programs can help educate researchers (Burgi et al., 2017; Cox et al., 2017; Tripathi et al., 2017).

Motivating researchers is the second identified practice in this dimension. Ideally, researchers should publish data alongside articles so that the results can be reproduced and the data can be used in a new context. Despite some journals and research organizations requiring researchers to share their data, most researchers do not. This indicates that their perceived risks outweigh the potential individual benefits they expect from data sharing (Eynden, Corti, Woollard, Bishop & Horton, 2011; Fecher, Friesike & Hebing, 2015; Strier et al., 2010). Among the primary reasons for not sharing data are concerns about patient privacy (in the medical field), the possibility of losing future publishing opportunities, the tendency to retain patents for data obtained over a long period, and legal issues related to abuse (Campbell & Bendavid, 2003). Jeng et al. state that the absence of reward models hinders data sharing. Researchers heavily rely on a reward system that allows identification, funding, and credibility to facilitate data sharing (Jeng et al., 2016).

Change of mindset is the third practice in this dimension. The default mentality of researchers is to publish research articles rather than research data (Crowston & Qin, 2011). According to Patel, the most difficult challenge in managing research data is persuading researchers of the importance of sharing the data and making it available to other researchers or users so they may use it for other research purposes. Researchers should know the benefits of sharing their data to expand their knowledge in a particular field and help the community (Patel, 2016; Tripathi & Pandey, 2018).

Research data sharing occurs when researchers, motivated by full knowledge, interest, and understanding, make their data available to other researchers for use or related scientific purposes. Hence, to implement a research data management system, it is recommended that research institutes and policymakers create mechanisms so that researchers have access to the necessary training and, secondly, create sufficient motivation to share research data. Encouraging researchers to share their research data by adopting secure data protection, designing data citation mechanisms, and preventing data abuse is possible. These measures can enhance the success of research data management systems.

The organization was identified as the third dimension of the research data management process. For the RDM program to be successful, the organization must take on several roles as the designer and planner of the program. The presented studies and frameworks identified three main roles for the organization.

The first practice in this dimension is to supply human resources. Access to skilled and qualified personnel facilitates the management of data. Academic librarians, system administrators, ICT technicians, senior university managers, research investors, and executives play an essential role in research data management and contribute to its success (Mayernik et al., 2012; Singh et al., 2018; Van Wyk et al., 2020). The most effective data management systems enable the mutual creation of expertise, practices, and skills among employees. Mayernik et al. discuss that effective data management systems enable the mutual creation of employees' skills, knowledge, and expertise (Mayernik et al., 2012). Research data management quality can be enhanced through workforce planning, training, recruitment, retention, and performance of skilled human resources and identifying regulatory options (Ndume et al., 2014).

The second practice of this dimension is policy-making. Over 60% of studies conclude that organizational policies are essential for successful research data management programs. A comprehensive policy framework is key to the success of research data management. Before implementing research data management, an organization should develop policies and guidelines (Ahmadi, Jano & Khamis, 2016). Several stakeholders are developing the RDM policy (including librarians, IT services, research bureaus, law firms, and academic institutions). The comprehensive policy document should clearly state the purpose, scope, application, and guidelines for data participants regarding data transmission, licensing, metadata entry, data classification, copyright agreements, and data use terms and conditions. This practice must address how research data are protected against security breaches, intellectual property concerns, and all activities involved in the research data management process (Ahmadi et al., 2016; Patel, 2016; Tripathi et al., 2017; Yunus & Johare, 2009).

The third practice of the organization is support and interaction. These services should be accessible to both researchers and end-users of the system. As a result of the complexity of research data management, technical support is required at various stages (A. M Cox et al.,

2017). They must be clearly defined in organizational policy and are integrated across all areas of the organization (Huh, Park, Ahn, Hwang & Jung, 2017; Laughton & Du Plessis, 2013; Pinfield et al., 2014; Tripathi & Pandey, 2018)

Consequently, research institutions and policymakers should develop a comprehensive and specific policy for research data management that would include all aspects and issues related to data, researchers, and technology. Additionally, hiring and employing skilled personnel with different specialities involved in research data management and providing appropriate training increases system security and is critical in ensuring its success. Finally, continuous interaction with users and researchers can improve system performance.

We have identified technology and infrastructure as the fourth dimension. Effectively Managing research data requires a technical infrastructure. Software and hardware infrastructure development are crucial to facilitate data storage and access (Henderson & Knott, 2015).

The first practice in this dimension is the creation of software infrastructure. Van Wyk et al. (2020) argue that the software sector should include data management planning tools, data collection tools, data analysis tools, data flow tools, data visualization tools, data citation generators, data warehouse software, and data storage tools. Nevertheless, the most important factor in evaluating potential software platforms is the software's ability to integrate with organizational needs and synchronize between different systems and hardware (Shrestha & Vassileva, 2018; Van Wyk et al., 2020). One of the most critical decisions is which software will host the data. Several studies have demonstrated that open-source software can be an excellent choice for hosting data to avoid dependency on others and prevent the vendor from locking data. For instance, Dataverse software is open source and available for free. According to some studies, the best option for an organization is to store data on its server (Patel, 2016; Tripathi & Pandey, 2018). Van Wyk et al. (2020) have recommended using cloud services to store and manage large data.

The second practice in this dimension is the development of hardware infrastructure. This sector's research data management necessitates high-performance computing facilities, hardware infrastructure, data warehouses, and network infrastructure (Mayernik et al., 2012). The hardware requirements for data management may differ depending on the range of data management services used. Before implementation, it is necessary to evaluate the hardware and software requirements of a research data management system. Regular system performance evaluation is needed to identify and correct weaknesses (Gunjal & Gaitanou, 2017; Mayernik et al., 2012; Tripathi & Pandey, 2018).

Research institutes and policymakers must be aware that long-term storage, maintenance, security, and the long-term use of data require appropriate tools and technologies to build networks, enterprise repositories, databases, software, and hardware. Also, hardware and software requirements are influenced by the nature of the organization. Furthermore, the specifications of the data management system determine the appropriate hardware and software.

### **Conclusion**

Re-accessing and reusing research data has apparent benefits and preserves the value of the data. This study demonstrates that RDM is a multidimensional concept that can be categorized into four main dimensions: data, researchers, organizations, and technology. Data processing,

data protection, data legal issues, and data sharing are the main practices of the data dimension. The researcher's dimension is educating, changing the mindset, and motivating researchers. On the organizational dimension, the essential practices supply human resources, policy-making, interaction, and support. On the technology dimension, the most significant practices include: Establishing software and hardware infrastructures. This study demonstrates that research data management is a systematic process involving various stakeholders and facilities, each cooperating and coordinating various activities. Researchers, policymakers, research organizations, and trained staff are among the most important participants in managing research data, and technical facilities are vital to its success.

Our results show that the models presented in these articles focused mainly on data, technology, and related issues. Fewer studies addressed the dimensions of researchers and organizations. Of the studies at the organizational level, these tended to concentrate on policy-making over other roles. RDM can be effectively achieved if operational dimensions and practices are well-defined and matched with comprehensive, transparent, expert policies and processes. As a result, the findings of this study can provide managers, policymakers, and researchers with an overview of the dimensions, practices, and processes involved in RDM. This will aid them in gaining a comprehensive understanding of the topic. This allows stakeholders to evaluate the requirements and aspects of an RDM system before implementing it and to plan with a full understanding of its requirements.

#### **Disclosure of potential conflicts of interest**

The authors declare that there is no conflict of interest regarding the publication of this article.

#### **Acknowledgements**

This study is part of a dissertation for a PhD in Medical Library, and Information Sciences approved (Code: IR.KMU.REC.1400.115) in Kerman University of Medical Sciences, Kerman, Iran. There has been no significant financial support for this work that could have influenced its outcome.

#### **References**

- Ahmadi, N. A., Jano, Z. & Khamis, N. (2016). Analyzing crucial elements of research data management policy. *International Business Management*, 10(17), 3847-3852. <https://dx.doi.org/10.36478/ibm.2016.3847.3852>
- Arend, D., Lange, M., Chen, J., Colmsee, C., Flemming, S., Hecht, D., & Scholz, U. (2014). e! DAL-a framework to store, share and publish research data. *BMC Bioinformatics*, 15(1). <https://doi.org/10.1186/1471-2105-15-214>
- Arksey, H. & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology*, 8(1), 19-32. <https://doi.org/10.1080/1364557032000119616>
- Borghi, J. A. & Van Gulick, A. E. (2021). Promoting open science through research data management. arXiv preprint arXiv:2110.00888.
- Burgi, P.-Y., Blumer, E. & Makhlof-Shabou, B. (2017). Research data management in Switzerland: National efforts to guarantee the sustainability of research outputs. *IFLA journal*, 43(1), 5-21. <https://doi.org/10.1177/0340035216678238>

- Campbell, E. G. & Bendavid, E. (2003). Data-sharing and data-withholding in genetics and the life sciences: results of a national survey of technology transfer officers. *Journal of Health Care Law & Policy*, 6(2), 241–255.
- Campbell, E. G., Clarridge, B. R., Gokhale, M., Birenbaum, L., Hilgartner, S., Holtzman, N. A. & Blumenthal, D. (2002). Data withholding in academic genetics: Evidence from a national survey. *JAMA*, 287(4), 473-480. <https://doi.org/10.1001/jama.287.4.473>
- Cox, A. M., Kennan, M. A., Lyon, L. & Pinfield, S. (2017). Developments in research data management in academic libraries: Towards an understanding of research data service maturity. *Journal of the Association for Information Science and Technology*, 68(9), 2182-2200. <https://doi.org/10.1002/asi.23781>
- Cox, A. M. & Tam, W. W. (2018). A critical analysis of lifecycle models of the research process and research data management. *Aslib Journal of Information Management*, 70(2), 142-157. <https://doi.org/10.1108/AJIM-11-2017-0251>
- Crowston, K. & Qin, J. (2011). A capability maturity model for scientific data management: Evidence from the literature. *Proceedings of the American Society for Information Science*, 48(1), 1-9. <https://doi.org/10.1002/meet.2011.14504801036>
- Dillo, I. & Doorn, P. (2014). The front office–back office model: Supporting research data management in the Netherlands. *International Journal of Digital Curation*, 9(2), 39-46. <http://dx.doi.org/10.2218/ijdc.v9i2.333>
- Eynden, V. V. d., Corti, L., Woollard, M., Bishop, L. & Horton, L. (2011). Managing and sharing data: Best practice for researchers. Essex: UK Data Archive. Retrieved from <https://dam.ukdataservice.ac.uk/media/622417/managingsharing.pdf>
- Fecher, B., Friesike, S. & Hebing, M. (2015). What drives academic data sharing? *PLoS One*, 10(2). e0118053. <https://doi.org/10.1371/journal.pone.0118053>
- Goben, A. & Griffin, T. (2019). In aggregate: Trends, needs, and opportunities from research data management surveys. *College & Research Libraries*, 80(7), 903. Retrieved from <https://crl.acrl.org/index.php/crl/article/view/23610/30923>
- Gunjal, B. & Gaitanou, P. (2017). Research data management: A proposed framework to boost research in higher educational institutes. *IASSIST Quarterly* 41(1-4), 12. <https://doi.org/10.29173/iq12>
- Hashim, H. N. M. (2019). Developing a model guidelines addressing legal impediments to open access to publicly funded research data in Malaysia. *Data Science Journal*, 18(27), 1-17. <https://doi.org/10.5334/dsj-2019-027>
- Henderson, M. E. & Knott, T. L. (2015). Starting a research data management program based in a university library. *Medical Reference Services Quarterly*, 34(1), 47-59. <https://doi.org/10.1080/02763869.2015.986783>
- Higgins, S. (2008). The DCC curation lifecycle model. *International Journal of Digital Curation*, 3(1), 134-140. <https://doi.org/10.2218/ijdc.v3i1.48>
- Huh, T., Park, G., Ahn, S., Hwang, S. & Jung, H. (2017). Design criteria of Korean LTER data platform model for full life-cycle data management. *International Journal of Applied Engineering Research*, 12(3), 336-342. Retrieved from [https://www.ripublication.com/ijaer17/ijaerv12v3\\_10.pdf](https://www.ripublication.com/ijaer17/ijaerv12v3_10.pdf)

- Ismail, N. A. M., Xu, M. X., Wood, M. & Welch, C. (2013). To share or not to share? Research-knowledge sharing in higher education institution: preliminary results. *International Journal of Information Technology Management*, 12(3-4), 169-188. <https://doi.org/10.1504/IJITM.2013.054809>
- Jeng, W., He, D. & Oh, J. S. (2016). Toward a conceptual framework for data sharing practices in social sciences: A profile approach. *Proceedings of the Association for Information Science Technology*, 53(1), 1-10. <https://doi.org/10.1002/pras.2016.14505301037>
- Krahe, M. A., Toohey, J., Wolski, M., Scuffham, P. A. & Reilly, S. (2020). Research data management in practice: Results from a cross-sectional survey of health and medical researchers from an academic institution in Australia. *Health Information Management Journal*, 49(2-3), 108-116. <https://doi.org/10.1177/1833358319831318>
- Laughton, P. & Du Plessis, T. (2013). Data curation in the World Data System: Proposed framework. *Data Science Journal*, 12, 56-70. Retrieved from [https://www.jstage.jst.go.jp/article/dsj/12/0/12\\_13-029/\\_pdf](https://www.jstage.jst.go.jp/article/dsj/12/0/12_13-029/_pdf)
- Mayernik, M. S., Choudhury, G. S., DiLauro, T., Metsger, E., Pralle, B., Rippin, M. & Duerr, R. (2012). The data conservancy instance: Infrastructure and organizational services for research data curation. *D-Lib Magazine*, 18(9/10). Retrieved from <http://www.dlib.org/dlib/september12/mayernik/09mayernik.html>
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G. & Group, P. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of Internal Medicine*, 151(4), 264-269. <https://doi.org/10.1136/bmj.b2535>
- Ndume, V., Nkansah-gyekye, Y., Ko, J. & Selemani, M. (2014). A framework for enhancing e-health data integration and sharing in distributed environments. *International Journal of Computing ICT Research*, 8(1), 10-21. Retrieved from <file:///C:/Users/drghane/Downloads/ndume-IJCIR-article2.pdf>
- Osareh, F., Saam, S. & Ghayebi, A. (2021). Identifying the dimensions of the development of research data management services in scientific centers. *Library Information Sciences*, 24(2), 85-111. <https://doi.org/10.30481/lis.2020.225614.1699> [in Persian]
- Patel, D. (2016). Research data management: A conceptual framework. *Library Review*, 65(4/5), 226-241. <https://doi.org/10.1108/LR-01-2016-0001>
- Pinfield, S., Cox, A. M. & Smith, J. (2014). Research data management and libraries: Relationships, activities, drivers and influences. *PLoS One*, 9(12), e114734. <https://doi.org/10.1371/journal.pone.0114734>
- Qin, J., Crowston, K. & Kirkland, A. (2017). Pursuing best performance in research data management by using the capability maturity model and rubrics. *Journal of eScience Librarianship*, 6(2). e1113 (2017). Retrieved from <https://crowston.syr.edu/node/679>
- Redkina, N. (2019). Current trends in research data management. *Scientific and Technical Information Processing*, 46(2), 53-58. <https://doi.org/10.3103/S0147688219020035>
- Shakeri, S. & Gracy, K. F. (2014). A model for data curation research in small sciences. *Proceedings of the American Society for Information Science Technology*, 51(1), 1-5. <https://doi.org/10.1002/meet.2014.14505101094>
- Shrestha, A. K. & Vassileva, J. (2018). Blockchain-based research data sharing framework for incentivizing the data owners. In *Blockchain-ICBC 2018: First International Conference, Held as Part of the Services Conference Federation, SCF 2018, Seattle, WA, USA, June 25-30, 2018, Proceedings 1* (pp. 259-266). Springer International Publishing.

- Singh, N. K., Monu, H. & Dhingra, N. (2018, February). Research data management policy and institutional framework. In *2018 5th International Symposium on Emerging Trends and Technologies in Libraries and Information Services (ETTLIS)* (pp. 111-115). IEEE.
- Strier, K. B., Altmann, J., Brockman, D. K., Bronikowski, A. M., Cords, M., Fedigan, L. M., Pusey, A. E. (2010). The Primate Life History Database: a unique shared ecological data resource. *Methods in Ecology Evolution*, 1(2), 199-211. <https://doi.org/10.1111/j.2041-210X.2010.00023.x>
- Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., Manoff, M. & Frame, M. (2011). Data sharing by scientists: practices and perceptions. *PLoS One*, 6(6), e21101. <https://doi.org/10.1371/journal.pone.0021101>
- Tricco, A. C., Cardoso, R., Thomas, S. M., Motiwala, S., Sullivan, S., Kealey, M. R., . . . Straus, S. E. (2015). Barriers and facilitators to uptake of systematic reviews by policy makers and health care managers: A scoping review. *Implementation Science*, 11(4). <https://doi.org/10.1186/s13012-016-0370-1>
- Tripathi, D. P. & Pandey, S. R. (2018, February). Developing a conceptual framework of research data management for higher educational institutions. In *2018 5th International Symposium on Emerging Trends and Technologies in Libraries and Information Services (ETTLIS)* (pp. 105-110). IEEE.
- Tripathi, M., Shukla, A., & Sonkar, S. K. (2017). Research Data Management practices in university libraries: A study. *Journal of Library Information Technology*, 37(6), 417.
- Van Wyk, J., Bothma, T., & Holmner, M. (2020). A conceptual virtual research environment model for the management of research data, a South African perspective. *Library Management*.
- Wilms, K. L., Stieglitz, S., Ross, B. & Meske, C. (2020). A value-based perspective on supporting and hindering factors for research data management. *International Journal of Information Management*, 54, 102174. <https://doi.org/10.1016/j.ijinfomgt.2020.102174>
- Yunus, A. M. & Johare, R. (2009). Preservation of primary research data and records for research: Development of a conceptual framework model. Paper presented at the International Symposium of InterPARES 3 Summit. Sungkyunkwan University, Seoul, South Korea. Retrieved from <file:///C:/Users/drghane/Downloads/Preservationofprimaryresearchdataandrecordsforresearch-Developmentofaconceptualframework.pdf>
- Zhao, H., Peng, J. & Huo, G. (2009). Virtual Models of Scientific Data Sharing. Paper presented at the 2009 Fourth International Conference on Cooperation and Promotion of Information Resources in Science and Technology. Beijing, China, (pp. 387-390). <https://doi.org/10.1109/COINFO.2009.30>