Received: 11 July 2020 | Accepted: 1 April 2021

DOI: 10.1002/asi.24483

RESEARCH ARTICLE

JASIST WILEY

Understanding the process of data reuse: An extensive review

Xiaoguang Wang[1,2] | **Qingyu Duan**[1] | **Mengli Liang**[1]

[1]School of Information Management, Wuhan University, Wuhan, China

[2]Big Data Institute, Wuhan University, Wuhan, China

Correspondence
Xiaoguang Wang, School of Information Management, Wuhan University, No. 299 Bayi Avenue, Wuhan, China.
Email: wxguang@whu.edu.cn

Funding information
Foundation for Innovative Research Groups of Hubei Province of China, Grant/Award Number: 2019CFA025; National Natural Science Foundation of China, Grant/Award Numbers: 1874129, 71790612, 71921002; Postdoctoral Science Foundation of China, Grant/Award Number: 2019M662727

Abstract

Data reuse has recently become significant in academia and is providing new impetus for academic research. This prompts two questions: What precisely is the data reuse process? What is the connection between each participating element? To address these issues, 42 studies were reviewed to identify the stages and primary data reuse elements. A meta-synthesis was used to locate and analyze the studies, and inductive coding was used to organize the analytical process. We identified three stages of data reuse—initiation, exploration and collection, and repurposing—and explored how they interact and form iterative characteristics. The results illuminated the data reuse at each stage, including issues of data trust, data sources, scaffolds, and barriers. The results indicated that multisource data and human scaffolds promote reuse behavior effectively. Further, two data and information search patterns were extracted: reticular centripetal patterns and decentralized centripetal patterns. Three paths with elements cooperating through flexible functions and motivated by different action items were identified: data centers, human scaffolds, and publications. This study supports improvements for data infrastructure construction, data reuse, and data reuse research by providing a new perspective on the effect of information behavior and clarifying the stages and contextual relationships between various elements.

1 | INTRODUCTION

Modern research is being reshaped by a flood of data, making data reuse a new scholarly debate in the context of the data-driven research paradigm that has attracted attention in academia worldwide. By 2019, the United States, Australia, Canada, China, and 16 European countries have made scientific data on national policies available, based on the expectation of enhanced data reuse (G. Song & Hu, 2016; The Digital Curation Centre, 2019). In addition, reusability was incorporated into the four basic principles of open data recognized and promoted by the Future of Research Communication and e-Science 11 (FORCE11) (FORCE11, 2016; Wilkinson et al., 2016). Reusability is becoming one of the guiding principles for building data resources in various fields (Calamai & Frontini, 2018; Schaaf et al., 2018; Vita et al., 2018). With broader possibilities in data policies and resources, there has also been growth in the positive perception and practice of data reuse in academia (Kriesberg et al., 2013; Tenopir et al., 2015). According to a global survey of 8,000 academics conducted by Figshare and Digital Science in 2019, close to or more than 50% of researchers in Europe, Asia, and North America have reused open data in their work (Fane, 2019).

Data reuse is rarely defined formally; however, two characterizations have emerged. One meaning considers reuse to be the repurposing of data for a new problem other than the initial intention of the data collected (Curty & Qin, 2014; Hinds et al., 1997; Law, 2005;

G. Sun & Khoo, 2016; Zimmerman, 2008). The other meaning refers to the practice of using data by people other than the original collector (I. M. Faniel et al., 2016; Fear, 2013). A broad definition is also provided by some researchers who define reuse as the behavior or the process to use research data, regardless of its purpose or the characteristics of its reuser (Castle, 2003; Sandt et al., 2019).

There are three predominant topics in data reuse research: analysis of the conceptual frameworks of data reuse (Custers & Uršič, 2016; I. V. Pasquetto et al., 2017), exploration of tools or standards supporting data reuse (Ball & Duke, 2015; Callaghan, 2015; Wilkinson et al., 2016; Wilkinson et al., 2017), and summaries of data reuse experiences and problems (Frank et al., 2019; Hsu et al., 2015; Huggett, 2018), including meta-syntheses of data reuse such as research type, barriers, research questions, data resources, and influencing factors (M. Song et al., 2013; Y. Sun et al., 2019). The extraction and conceptualization of specific elements, such as data evaluation, barriers, supporting tools, and rights issues, also focus (Benitez-Paez et al., 2018; Boté & Térmens, 2019; Weiskopf & Weng, 2013; Williams et al., 2017; Zhang et al., 2018). However, few literature reviews have directly focused on how various elements cooperate to accomplish the data reuse process. Although some exploratory articles have referred to the stages or roles of participating elements (Daniels, 2014; Huggett, 2018; X. Song et al., 2020; Yoon, 2014b; Zimmerman, 2003), the descriptions are vague, and relationships are not explored.

Positive awareness of data reuse, infrastructure support, and the breadth of practice in various fields have been improving, and significant achievements have been made in the conceptual category of data reuse and solution for specific elements (Fane, 2019; I. M. Faniel et al., 2016; Tenopir et al., 2015; Wilkinson et al., 2016; Zimmerman, 2008). However, despite some academic interest in the issue of the data reuse process, research on this topic is still fragmentary and obscure, requiring further enhancement of the conceptual framework of data reuse and improvement in data reuse. This study focuses on the general nature, occurrence patterns, and participating elements of data reuse and poses the following questions:

RQ1. *What stages do researchers go through when reusing existing data?*

RQ2. *What are the specifics of each stage?*

RQ3. *How do the primary elements fit together?*

To answer these questions, information behavior theory was used as a reference, and the meta-synthesis and inductive coding methods were applied to analyze existing qualitative and quantitative studies exploring data reuse rather than utilizing data reuse. This review takes a user perspective, that is, the perspective of researchers who reuse data, to explore the entire data reuse process.

2 | INFORMATION SEEKING BEHAVIOR RESEARCH

Information-seeking behavior is searching, obtaining, and using the information in various situations (Pettigrew et al., 2001). Various information-seeking models have been proposed to clarify its features and the relationship between information needs and behavior (Ellis, 1989; Spink, 2006; Wilson, 2016). A classic information behavior theory is Kuhlthau's model, which proposes an information-seeking process consisting of six stages from the user perspective—initiation, selection, exploration, formulation, collection, and presentation—with the realms of feelings, thoughts, and actions common to each stage (Kuhlthau, 1991; Zeng & Xue, 2013). However, the action realm consists only of the stages of initiation, exploration, and collection, which seek background, relevant information, and relevant or focused information, respectively (Kuhlthau, 1991).

Some researchers have focused on searching for available data in the context of information behavior, in which information seeking serves as an important connection that shortens the distance between the reuser and data and facilitates a smooth process (Rolland & Lee, 2013; Whitmore, 2016; Yoon, 2014b; Zimmerman, 2007). Based on the cognition of correlation between information and data reuse behaviors, we used information-seeking behavior theory as the foundation for categorizing and theorizing the coding results, and used Kuhlthau's model to instruct the construction of a theoretical data reuse process model and present the relationship between the theoretical categories.

3 | METHODS

This study used the meta-synthesis method to collect and analyze literature and inductive coding based on grounded theory to extract data and organize data. Meta-synthesis is a method that can include qualitative and quantitative studies for analysis, and consists of four parts: a search, the determination of inclusion and exclusion criteria, literature evaluation, and data extraction

and analysis (Catalano, 2013; Edwards & Kaimal, 2016). This study adopted inductive coding methods based on grounded theory to conduct the fourth part of a meta-synthesis. Grounded theory is an effective scientific qualitative research method used to establish theory from empirical data with a rigorous coding paradigm (Strauss, 1987). Using open, axial, and selective coding, this study extracted data and constructed a theoretical framework for the data reuse process.

3.1 | Search strategies

Two databases were searched in December 2019 to obtain relevant articles, the Web of Science Core Collection and two ProQuest indexes (ProQuest Dissertations & Theses Global A&I: The Humanities and Social Sciences Collection and ProQuest Dissertations & Theses Global A&I: The Sciences and Engineering Collection). These indexes cover the core collection of peer-reviewed, scholarly journals or dissertations published worldwide in science, social science, and humanities. With reference to previous studies (Y. Sun et al., 2019; Zimmerman, 2008) and a preretrieval, keywords were selected, including *data reuse, data reusing, secondary analysis, and data reusability*. After the retrieval, citation and reference tracking was conducted to identify the most relevant articles using Google Scholar.

3.2 | Inclusion/exclusion criteria

The following criteria were used: (a) the primary focus was on researchers' data reuse behaviors; (b) based on empirical data rather than purely criticism or exploration of the conceptual category system; (c) if the differences were insignificant between some dissertations, journal articles, and conference papers written or coauthored by the same author, only the most comprehensive articles were included; that had been (d) peer-reviewed before publication. Studies on public or enterprise reuse behaviors were excluded.

3.3 | Study identification

Figure 1 illustrates the selection process. One author reviewed the title and abstract to identify a smaller subset, which was then read in full by two authors to filter and determine the final sample for analysis.

Two authors coded four articles twice with discussions to confirm the coding agreement (Cohen's kappa = 0.89), and we also reserved 14 articles in advance for theoretical saturation tests. The coding range excluded literature reviews and preserved the content of the authors' findings and conclusions.

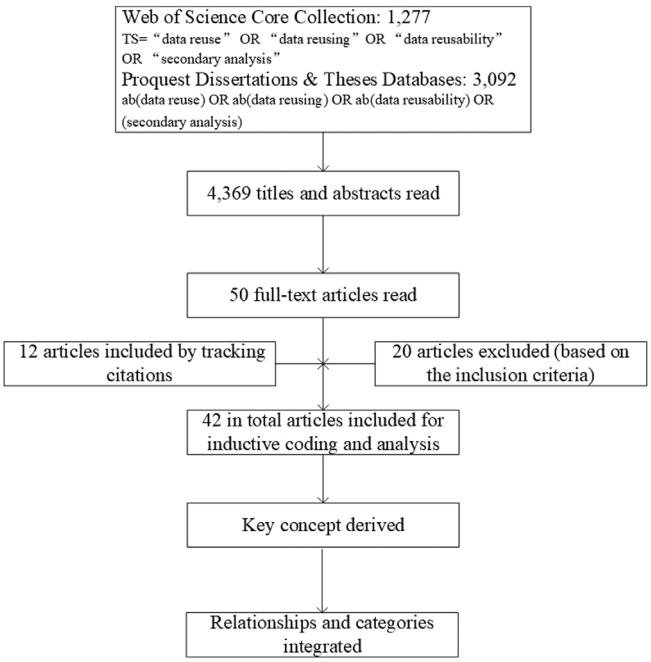

FIGURE 1 Selection and analysis process

3.4 | Data extraction and analysis

The full text of the selected articles was imported into NVivo 11 plus for inductive coding. The following data were extracted: basic data, including year, author(s), topics, the publisher, research method(s) and data collection instrument, sampling methods, sample size, response rate, and discipline of the participant, as well as the relevant characteristics of data reuse, including the stages, motivation, methods, participants, and their correlation. Grounded theory was used to guide the coding, conceptualize the original statements, identify the categories, and build relationships between categories.

4 | DATA ANALYSIS

The 42 articles included in this review were published between 2003 and 2019, and their characteristics are presented in Table 1.

Three types of peer-reviewed papers were selected: journal articles ($N = 22$), conference presentations ($N = 9$), and dissertations ($N = 11$). Figure 2 describes the data reuse focus for different fields, where natural science papers focused on medicine, ecology, physics, and

TABLE 1 Characteristics of the included articles (sorted by year)

Articles	Type	Discipline(s)	Topic of reuse	Sampling method	Sample size *N*	Response rate	Data collection instrument
Zimmerman, 2003	Dissertation	Ecology	The reuse experiences of ecologists	Purposive, random and stratified	20		Semi-structured interview
Birnholtz & Bietz, 2003	Conference paper	Earthquake engineering, HIV/AIDS research, and space physics	The role of data and information context	Purposive	Not reported (≥95.5 hr observation and ≥50 interviews)		Ethnographic observations and semi-structured interview
Zimmerman, 2007	Journal paper	Ecology	Knowledge for reusing data	Purposive	13		Semi-structured interviews
Carlson & Anderson, 2007	Journal paper	Astronomy, anthropology and social science	The data kinds	Not reported	16, respondents not reported	Not reported	Case study and semi-structured interview, and electronic communication
Zimmerman, 2008	Journal paper	Ecology	The affection of standards	Purposive	13		Semi-structured interviews
Niu, 2009	Conference paper	Social science	Inadequate documentation and the reasons and ways seeking outside information	Random	13	384/1260	Unstructured and exploratory interview
I. M. Faniel & Jacobsen, 2010	Journal paper	Earthquake engineering	Assessing the reusability of colleagues' data	Purposive	14		Semi-structured interview
Tenopir et al., 2011	Journal paper	Social science and science	Practice and situation	Snowballed	1,329	9%	A online survey questionnaire
Whyte & Pryor, 2011	Journal paper	Astronomy, bioinformatics, chemistry, epidemiology, language technology and neuroimaging.	Affection of documentation and quality assurance	Not reported	18		Case study and semi-structured interview
I. M. Faniel et al., 2012	Conference paper	Social science	Reuse among novice researcher	Purposive	22		Interview
Sands et al., 2012	Conference paper	Astronomy	The ways to reuse data	Purposive	13		Semi-structured interview
I. Faniel, Barrera-Gomez, et al., 2013	Conference paper	Social science and archaeology	Comparison between disciplines	Not reported	66		Semi-structured interview
I. Faniel, Kansa, et al., 2013	Conference paper	Archaeology	Data context	Convenient and snowballed	22		Semi-structured interview

TABLE 1 (Continued)

Articles	Type	Discipline(s)	Topic of reuse	Sampling method	Sample size *N*	Response rate	Data collection instrument
Kriesberg et al., 2013	Conference paper	Social science and archaeology		Convenient and snowballed	92		Semi-structured interview
Rolland & Lee, 2013	Conference paper	Cancer epidemiology	Reuse practices of postdoctoral researchers in cancer epidemiology	Purposive	11		Semi-structured interview
Atici et al., 2013	Journal paper	Zooarchaeology	The reasons of publishing primary data	Not reported	3		User experiment
Yakel et al., 2013	Journal paper	Social science and archaeology	Factors influencing the trust for depository	Convenient and snowballed	66		Semi-structured interview
Daniels, 2014	Dissertation	Botany and archaeology	Museum contexts experiences	Snowballed	45		Case study and semi-structured interview, and observation
Yoon, 2014a	Journal paper	Social science	Factors influencing the development of users' trust in repositories	Purposive	19	25/213	Semi-structured interview
Yoon, 2014b	Conference paper	Social science	Experience of reusing qualitative data	Purposive and random	8	32.5%	Interview
Yoon, 2015	Dissertation	Public health, social work	Facets of data reusers' trust in data	Purposive	38	25.3%	Semi-structured interview
Curty, 2015	Dissertation	Social science	Factors influencing reuse	Purposive	13, 9, 12		Semi-structured interview, expert judgment and a survey questionnaire in Qualtrics
Federer et al., 2015	Journal paper	Clinical and basic science	Attitudes and practices of biomedical data sharing and reuse	Random	135	Not reported	A questionnaire in SurveyMonkey
Tenopir et al., 2015	Journal paper	Social science, nature science and humanities	Changes in practice and perceptions over the past 4 years	Snowballed	1,329 and 1,015	Not reported	A online survey questionnaire
Weiskopf, 2015	Dissertation	Clinical science	Data quality assessment framework and guideline for reuse	Purposive	9		Semi-structured interview
Stvilia et al., 2015	Journal paper	Physics	Constructs of research project tasks and data quality	Stratified and random	Not reported (12 interviews), 160	24%	Semi-structured interviews and a survey questionnaire in Qualtrics

(Continues)

TABLE 1 (Continued)

Articles	Type	Discipline(s)	Topic of reuse	Sampling method	Sample size *N*	Response rate	Data collection instrument
Shen, 2015	Journal paper	Agriculture and life sciences, architecture and urban studies, business, engineering, arts and human sciences, natural resources and environment science, veterinary medicine	The practice of regional scholars	Random	423	652/2532	A online survey questionnaire
I. M. Faniel et al., 2016	Journal paper	Social science	Factors influencing reusers' satisfaction	Random	249	16.8%	A survey questionnaire in Qualtrics
Murillo, 2016	Dissertation	Nature science	Affection of descriptive information	Random	16	—	Think-aloud guidance and questionnaire
Whitmore, 2016	Dissertation	Archaeology	Data context	Purposive	16	—	Semi-structured interview
S. Joo et al., 2017	Journal paper	Health science	Affection of attitudinal, social, and resource factors	Random	161	6.74%	E-mail survey in Qualtrics
Kim & Yoon, 2017	Journal paper	STEM	Factors influencing reuse	Random and stratified	1,237	11.14%	A online survey questionnaire
Y. K. Joo & Kim, 2017	Journal paper	Engineering	Factors influencing reuse	Random	193	9.07%	A online survey questionnaire
Spreng, 2017	Dissertation	Genetics	Methods for the reuse of public data	Not reported	53, 33, 450	—	Experiment
Wei, 2017	Dissertation	Social science	Qualitative data sharing and reuse	Convenient, purposive and expert	13 and 8 interviewee 66 and 65 respondents	17.4%, 11.8%	Interview and self-assessed questionnaire, focus group, Likert scale questionnaire
Yoon, 2017	Journal paper	Public health, social work	Stages of trust development for data	Purposive	38	25.3%	Semi-structured interview
Yoon & Kim, 2017	Journal paper	Social science	The roles of attitudinal beliefs, attitudes, norms, and data repositories	Random	292	14.91%	A online 5 point Likert scale questionnaire
Poole, 2017	Journal paper	Digital humanities	The challenges and situation	Convenient and snowballed	45	—	Semi-structured interview
I. Pasquetto, 2018	Dissertation	Biomedical science	The socio-technical, epistemic, and ethical challenges.	Purposive	Not reported (50 interviews)		Ethnographic observations and semi-structured interview

TABLE 1 (Continued)

Articles	Type	Discipline(s)	Topic of reuse	Sampling method	Sample size *N*	Response rate	Data collection instrument
Tenopir et al., 2018	Journal paper	Geophysics	Motives, attitudes, and data practices	Random	1,372	2.2%	A survey questionnaire in Qualtrics
Federer, 2019	Dissertation	Biomedicine	Reuser and purpose	Purposive	5,513 requestors and 11,832 requests	—	Content analysis of reusers' requests and text mining
Yoon & Lee, 2019	Journal paper	Nature science, social sciences	Factors of trust	Random	145	8%	A 5 point Likert scale questionnaire

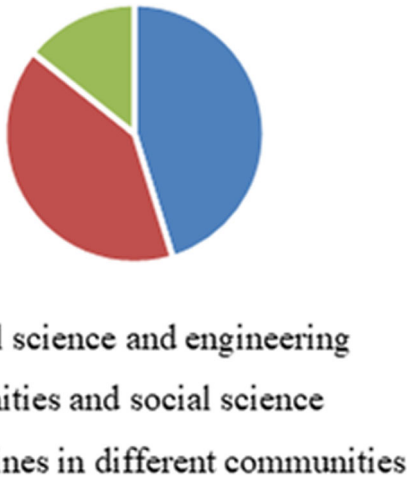

FIGURE 2 Disciplinary distribution of articles and research topics [Color figure can be viewed at wileyonlinelibrary.com]

earthquake engineering; humanities, and social science papers concentrated in the whole community rather than the specific discipline, and the remaining six articles involved general issues across disciplines. Figure 3 depicts the method frequency used by these studies, with interviews being the most common, which were often used together with observations, focus groups, think-aloud, or case studies.

4.1 | Inductive coding results

Data reuse is a delicate and often time-consuming process, involving multiple participating elements, not just reusers. We extracted the text on research methods, results, discussion, and other contributory sections of the 42 articles, excluding the literature review. Based on these texts, three-level coding using grounded theory was implemented to examine the process and data reuse elements.

4.1.1 | Open coding

Open coding is the basic stage in grounded theory and involves a process of conceptualization based on original sentences and categorization of concrete concepts. After repeated comparison, verification, and exploratory questioning by the authors, 61 initial concepts were derived, as displayed in Table 2.

After conceptualizing the original sentences, we combined the concepts for comparison and merged similar concepts into 17 categories, as illustrated in Table 3.

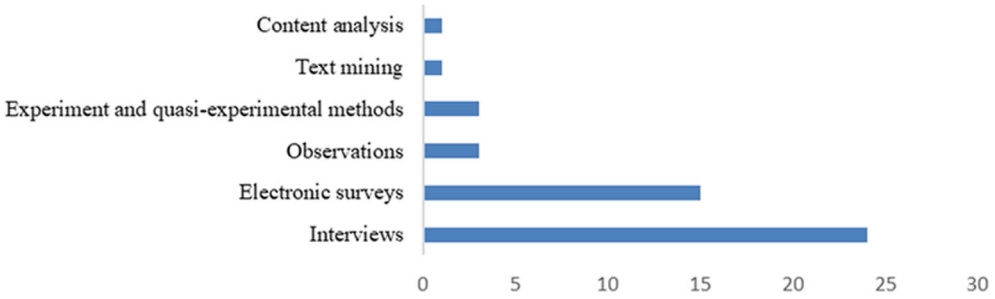

FIGURE 3 Methods used [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 2 Results of conceptualization

Conceptualization	Original statements and sources (examples)
a1 perceived usefulness/benefits	"In terms of individual motivations, perceived usefulness in data reuse ...was found to have a significant positive relationship with attitudes toward data reuse." (Y. K. Joo & Kim, 2017)
a2 perceived concerns	"Perceived concerns about data reuse were found to have a significant negative relationship with engineering researchers' attitudes toward data reuse." (Y. K. Joo & Kim, 2017)
a3 perceived risks	"Perceived risks [fear of being undervalued, fear of infringing ethical codes, Slippage, vulnerability to hidden errors] are considered as foreseeable harmful consequences associated with the reuse of research data." (Curty, 2015)
a4 perceived effort	"Perceived effort was not found to have any significant influence on social scientists' attitudes toward data reuse..." (Yoon & Kim, 2017)
a5 data quality	"Four of the five data quality attributes had significant positive associations with data reusers' satisfaction, including data completeness..., data accessibility ..., data ease of operation..., and data credibility..." (I. M. Faniel et al., 2016)
a6 sufficient context	"The data were always 'cooked'... It appeared to depend largely on what was considered to be 'enough context' to fulfill the needs of the data re-user." (Carlson & Anderson, 2007)
a7 data documentation availability	"The documentation of this genetics data, so far I cannot find any. It's totally mystifying... (Luke, Professor)." (Curty, 2015)
a8 trust in repository	"Trust in the repository is a separate and distinct factor from trust in the data." (Yakel et al., 2013)
a9 publishing room	"Publishing was a major challenge to the qualitative researchers and they said there was not many room for publish secondary data analysis." (Yoon, 2014b)
a10 original value	"There was a value issue going on, yeah, there was actually a value issue and it was not as respected as collecting my own data, doing that [reusing existing data]. (Denise, Professor)." (Curty, 2015)
a11 peer/community encouragement	"When social scientists perceive an expectation of data reuse and acknowledgement by their communities and disciplines that data reuse is common and acceptable, they are more likely to have strong norms of data reuse and to have positive attitudes toward data reuse." (Yoon & Kim, 2017)
a12 institutional support	"As those concerns can be an important impediment to reusing data, it is important to provide proper education and institutional support to address such concerns." (Kim & Yoon, 2017)
a13 availability of repository support	"Social scientists recognize the importance of having the technological infrastructure of data repositories as important for the reuse of data..." (Curty, 2015)
a14 primary data collector reach	"I went to Philadelphia and I met some people who run the bail system there, and they gave me a lot of the institutional details that I needed." (Curty, 2015)
a15 purposes of data	Table showed the purpose: original research study, meta-analysis study, statistical methods study, software or tool development study, validation, reproducibility or reanalysis study, infrastructure (Federer, 2019)
a16 types of data reuse	"This typology..., does not classify data reuse practices by the type of data that is reused (background data vs. foreground data), but rather by the type of research purpose (background reuse vs. foreground reuse)." (I. Pasquetto, 2018)

TABLE 2 (Continued)

Conceptualization	Original statements and sources (examples)
a17 initial promoting conditions of data reuse	"The conditions that initially prompted researchers to seek out spatial information about their sites were divided into five categories: new site, publication, further analysis, expanding study area, and synthesis." (Whitmore, 2016)
a18 databases/repositories as sources a19 publications as sources a20 museums/libraries as sources a21 CD-ROM as sources a22 websites as sources	"Among my interviewees, public sources of data and information about data included natural history museums, published literature, and bibliographic databases, as well as databases available on CD-ROM, over the Internet, or through a public data center." (Zimmerman, 2007) "Actually, [I would be] potentially likely [to reuse data from a data repository]." (I. M. Faniel & Jacobsen, 2010) "[A spatial information source list includes] colleagues, company, data repository, government institutions, libraries, personal libraries, private museums and archives, project websites, research institutions." (Whitmore, 2016)
a23 through publications	"EE researchers learn about potential data for reuse through colleagues' journal articles as well as personal networks." (I. M. Faniel & Jacobsen, 2010)
a24 through personal knowledge	"Ecologists' knowledge was also useful in providing initial direction to their search for sources of data." (Zimmerman, 2007)
a25 through personal network a26 through databases	"They learned about collection strengths through sources including literature search, recommendations from mentors, and databases..., I discuss these methods for finding data and the ways..." (Daniels, 2014)
a27 web/internet search	"Astronomers may use generic Internet search engines to discover necessary datasets in addition to formal literature searches." (Sands et al., 2012)
a28 specific requests via personal network	"Ecologists who requested data directly from individuals or from institutions, such as natural history museums, made very specific requests." (Zimmerman, 2007)
a29 publications search a30 communication with personal network	"These resources can be discovered via searching data release papers and the open literature, by communication with colleagues, and other means." (Sands et al., 2012)
a31 data document search	"She was forced to return to the study documentation and the files on the shared drive to try to identify the correct one." (Rolland & Lee, 2013)
a32 downloading data directly	"While some datasets may be small enough to download to a local laptop over a wireless connection..." (Sands et al., 2012)
a33 sharing via server	"Some university departments have chosen to download large sets of the SDSS data and to keep them locally on the university server cluster." (Sands et al., 2012)
a34 visiting in person via personal relationship	"Collection managers at the Kelsey pull artifacts for researchers, which they can examine in the collection storage room or a study area adjacent to the collection managers' office." (Daniels, 2014)
a35 difficult accessibility	"As archeologist CCU17 described, one primary challenge for data reuse in his field was discovery and access. Not only was there a significant lag between data collection and publication, there was also a lack of centralized access." (Kriesberg et al., 2013)
a36 failure of human scaffold reach in discovery and acquiring	"As seen in the techniques to identify potential sources...Five of the researchers described situations in which they were unable to contact the original data collector because they had either passed away or were not working in archaeology any longer." (Whitmore, 2016)
a37 technical issue	"The size of available bandwidth and CPU space limit the astronomer's decision of how much data to retrieve." (Sands et al., 2012)
a38 encoding method for variables	"He noted that variables were derived for previous papers in a certain way, and it was important to keep that consistent, yet those variables were each derived by individual investigators instead of by the central authority, leaving open the possibility for future mistakes by others." (Rolland & Lee, 2013)
a39 the data processing procedures	"At issue were the data collection and recording procedures." (I. Faniel, Kansa, et al., 2013)
a40 through formal education	"Formal disciplinary training along with insights gained in the field lead to familiarity with particular types of data." (Zimmerman, 2007)
a41 through experience	"Knowledge about what can go wrong comes primarily from ecologists' own data collecting experiences, especially their assessments of the degree of skill required to gather particular data, and from their perceptions of or personal knowledge about the competence and commitment of specific data collectors." (Zimmerman, 2008)

(Continues)

TABLE 2 (Continued)

Conceptualization	Original statements and sources (examples)
a42 through data document	"You know I had to kind of comb through some of the actual files of the survey (Michael, Professor)." (Curty, 2015)
a43 through data manager	"Upon receipt of the dataset, postdocs began to get a feel for the data by talking with the study's data manager, running summary statistics (e.g., mean, median, missing, distribution) and checking those against previously published manuscripts." (Rolland & Lee, 2013)
a44 through personal network	"[There were]68% of survey respondents sought outside information from other people. Some of those people work closely with secondary data users, such as mentors, advisors and colleagues." (Niu, 2009)
a45 through relevant websites	"Besides those main sources [websites of data producers, websites of data archives and so on], other information sources include..." (Niu, 2009)
a46 through metadata to understand	"Metadata allow scientists to verify the quality and accuracy of the data." (I. Pasquetto, 2018) "In particular, 18 interviewees (27.3%) mentioned ...metadata... in conjunction with a trust decision." (Yakel et al., 2013)
a47 by e-mail or telephone a48 face-to-face a49 by working together	"[There were] 68% of survey respondents sought outside information from other people... 80% obtained that information through e-mail or telephone, 55% obtained that information through face-to-face conversations, 31.4% obtained that information by working together with other people." (Niu, 2009)
a50 insufficient data documentation	"While all three faunal analysts determined that the quality of the data was sufficient to move forward with analysis, they lamented that certain data were not present, specifically contextual and methodological information." (Atici et al., 2013)
a51 lack of specialized knowledge	"When researchers use data across disciplines, disciplinary differences present a challenge... It was a disciplinary difference in how the data was collected and what was included in the set." (Yoon, 2014b)
a52 failure of human scaffold reach in helping to understand	"Whereas the person processing the data was aware of what kinds of documentation and metadata would have best supported reuse, they were not in a position to interact or clarify with the original data collector." (Whitmore, 2016)
a53 cleaning and picking the useful portion of data	"I need to clean the data and I need to pick the things that are useful for me (Beth, PhD candidate)." (Curty, 2015) "Choices about data aggregation and disaggregation will depend on the research question(s) being asked" (Atici et al., 2013)
a54 aggregating multiple sources and types data	"Matching and Merging Data across Multiple Datasets...NSSRs were combining data from multiple sources (e.g., repositories)." (I. M. Faniel et al., 2012) "The most common approach was to synthesize multiple types and sources of spatial data in order to determine accuracy..." (Whitmore, 2016)
a55 encoding and extracting data	"Transforming Qualitative Data to Quantitative Data" (I. M. Faniel et al., 2012)
a56 transforming data format	"Nearly every interview contained descriptions of the amount of time required to process these data into a usable format." (Whitmore, 2016)
a57 compensating or losing data granularity	"...faced...the discrepancies between the granularity of the original data collection and the granularity needed by the participant." (Whitmore, 2016)
a58 comparing data	"Comparing variance in gene expression across a collection of public studies to reduce data dimensionality prior to clustering." (Spreng, 2017)
a59 insufficient data	"more often participants identified issues that hindered or prevented them from being able to reuse the spatial data (Table 7 [Accuracy, age of data, changing personnel and systems, disassociated datasets, found errors, granularity or resolution, missing or sparse documentation, ownership of data, time])." (Whitmore, 2016)
a60 a lack of skills	"Nine of the sixteen researchers expressed difficulties [to process and reuse data] due to the original data collection occurring on a higher level than they required for their projects." (Whitmore, 2016)
a61 an iterative process	"Data selection (as an ongoing process throughout the data reuse experience, as participants may stop using the data at any point)..." (Yoon, 2017)

TABLE 3 Result of categorization

Conceptualization	Categorization	Connotation
a1 perceived usefulness/benefits a2 perceived concerns a3 perceived risks a4 perceived effort	A1 attitudes	Perceptions about the pros and cons of data reuse, which embody positive or negative attitudes.
a5 data quality a6 sufficient context a7 data documentation availability a8 trust in a repository	A2 data reusability assessment	Matched with researcher's research needs to obtain a data assessment impression.
a9 publishing room a10 original value a11 peer/community encouragement	A3 disciplinary climate	Cognition and behavioral evaluation of data reuse in specific discipline communities and publishing institutions form an emotional tendency such as encouragement or rejection.
a12 institutional support a13 availability of repository support a14 primary data collector reach	A4 external assistance availability	Based on personal knowledge, researchers perceive the possibility of relevant institutions or individuals contributing to the reuse of existing data for research.
a15 purposes of data a16 types of data reuse a17 initial promoting conditions of data reuse	A5 intentions of data reuse	Data reuse is a goal-oriented behavior. Researchers differ in data requirements and data processing measures for different purposes.
a18 databases/repositories as sources a19 publications as sources a20 museums/libraries as sources a21 CD-ROM as sources a22 websites as sources	A6 data source	Where the researchers obtain the data.
a23 through publications a24 through personal knowledge a25 through personal network a26 through databases	A7 scaffold helping discover data	The mediator indicates or discloses to the researcher whether the required data exist and where it is located.
a27 web/internet search a28 specific requests via a personal network a29 publication search a30 communication with a personal network a31 data document search	A8 data discovery method	The operation the researcher uses to locate data.
a32 downloading data directly a33 sharing via server a34 visiting in person via personal relationship	A9 data access method	The operation after locating data
a35 difficult accessibility a36 failure of human scaffold reach in discovery and acquiring a37 technical issue	A10 barrier preventing data discovery or access	A data discovery or accessing operation is interrupted due to a lack of relevant information, skills, or approaches.
a38 encoding method for variables a39 the data processing procedures	A11 understanding data properties	Researchers understand the procedures of data collection and processing, as well as the meaning of variables to select the most appropriate data and ensure that the data are used correctly
a40 through formal education to understand data a41 through experience to understand data	A12 scaffold helping understand data	The mediator discloses to the researcher the procedures of data collection and processing and the meaning of variables.

(Continues)

TABLE 3 (Continued)

Conceptualization	Categorization	Connotation
a42 through data document to understand data a43 through data manager to understand data a44 through a personal network to understand data a45 through relevant data websites to understand data a46 through metadata to understand data		
a47 by e-mail or telephone a48 face-to-face a49 by working together	A13 method to get scaffold helping data understanding	The operation through which the researcher gets in touch with scaffolds.
a50 insufficient data documentation a51 lack of specialized knowledge a52 failure of human scaffold reach in helping to understand	A14 barriers preventing data understanding	Failure in data properties investigation.
a53 cleaning and picking the useful portion of data a54 aggregating multiple sources and types of data a55 encoding and extracting data a56 transforming data format a57 compensating or losing data granularity a58 comparing data	A15 processing data for research needs	The acquired data are processed in terms of granularity, format, magnitude, and other properties to analyze in accordance with research requirements.
a59 insufficient data a60 a lack of skills	A16 barriers preventing data process and reusing	Impossible to re-process and reuse the required data after they acquire and understand it.
a61 an iterative process	A17 an iterative process	The steps from start to finish are not linear across the data reuse process, with some steps repeated multiple times and interactions between the steps.

4.1.2 | Axial coding

By analyzing the logical relationship among concepts, axial coding expands the category with an "axis" concept and explores the relationship between categories. This study identified five main categories: making a decision, discovering and acquiring data, understanding and choosing data, processing and reusing data, and the iterative property. This was combined into four dimensions: cognition, data context, research context, and behavioral traits, as shown in Table 4.

4.1.3 | Selective coding

In the selective coding stage, relationships between the main categories were studied and established through repeated deliberation and reasoning to construct the theoretical model of the data reuse process. The five main categories developed in the axial coding stage represent the process of data reuse. Theoretical saturation tests were carried out using the reserved 14 articles, and no additional important concepts, categories, or relationships were found. Therefore, the results of this grounded theory demonstrated that the concept category reached the saturation state. Combined with the core categories of the "data reuse process," relationships between the main categories and core categories are displayed in Table 5.

5 | RESULTS

5.1 | Theoretical model of the data reuse process

According to the concepts of Kuhlthau's model, the data reuse process is composed of different stages and

TABLE 4 Results of axial coding

Dimension	Main category	Category
Cognition	Make a decision	A1 attitudes A2 data reusability assessment A3 disciplinary climate A4 external assistance availability A5 intentions of data reuse
Data context and entity	Discover and acquire data	A6 data source A7 scaffold helping discover data A8 data discovery method A9 data access method A10 barrier preventing data discovery or access
	Understand and choose data	A11 understanding data properties A12 scaffold helping understand data A13 method to get scaffold helping data understanding A14 barriers preventing data understanding
Research context	Process and reuse data	A15 processing data for research needs A16 barrier preventing data processing and reusing
Behavior	The iterative property	A17 an iterative process

transitions from the previous stage to the next stage. Each stage contains commonalities over a while during the data reuse process, including feelings, thoughts, and action realms. In the study, we only concluded stages from the action realm, extracting the commonalities of action items and action orientation.

The above inductive coding process demonstrates the reusers' efforts to reuse data. Data reuse can be viewed as four categories—making a decision, discovering and acquiring data, understanding and choosing data, and processing and reusing data—and the relationships between them. To further illustrate how categories constitute the data reuse process and their positions in the process, all categories were grouped into different stages according to the order of action and the action orientation. Combining grounded theory and the coding results with the information-seeking model, the theoretical model of the data reuse process containing the stage, orientation, action item, and relationships between items was obtained (Figure 4).

Data reuse is a dynamic process. Each stage illustrated in Figure 4 contains commonalities over a while, namely, action items and orientation. The initiation stage (a), stimulated by data needs, involves decision making about whether to reuse existing data, with the target of obtaining an initial sense of data reuse. A positive answer will induce the second stage, exploration and collection (b). In this stage, researchers need to discover, acquire, understand, and choose the required data in various ways and from various sources, in which object orientation is the data entity and context. If sufficient relevant, efficient, and available data are located, the determination is enhanced or else abandoned. When a favorable decision is made, before the final data are selected and obtained, the researcher shifts between the discovery and acquisition of data and the action of understanding and choosing data, correcting choices about inappropriate data or searching for additional data to incorporate for the best fit (Rolland & Lee, 2013). After data selection, collected data are matched to the research purpose and secondary processing—that is, repurposing (c), begins. This stage adapts data to the research, adopts various data processing operations, and researches the secondary processed data.

5.2 | Specifics of each stage

5.2.1 | Initiation

In the initiation stage, the researcher becomes aware of data requirements and the data reuse possibility. Uncertainty lies in the researcher's cognition and is overcome by perceiving encouragement, threat, barriers, and realizability of data reuse that are shaped in experience or knowledge. In action, this is called making a decision, in which some factors conspire to have a complex effect on whether this stage begins.

We statistically sorted the literature sources of the five factors shown in Table 4 influencing decision making, which is depicted in Figure 5, to reflect the extent to which different factors affect the decision positivity of data reuse. Research intention, perceived data reusability, and personal attitude top the list. Researchers weigh the benefits, risks, and personal ability and consider whether reusing existing data suits the research nature. Intentions or attitudes cannot always act as independent factors, as these are primarily influenced by the disciplinary climate and other factors. For instance, the value or viability of research reusing data is questioned by some disciplinary communities, resulting in a limited scope for publication

TABLE 5 Relationships between main and core categories

Typical relationships	Relationships structure	Connotation of relationships structure
Make a decision→Discover and acquire data	Inducing relationship	Decisions about data reuse are followed by discovering and acquiring data. The reuse effect induces the next action.
Discover and acquire data→Make a decision	Enhancing relationship	The results of data discovery and acquisition influence the positivity of data reuse decisions. The reuse effect should enhance the last action.
Discover and acquire data→Understand and choose data	Supportive relationship	Based on data discovery and acquisition, researchers understand and choose the appropriate data for correct reuse.
Understand and choose data→Discover and acquire data	Iterative relationship	The negative results of understanding and choosing data induce the next data discovery and acquisition action due to more data requirements or discovery errors based on understanding, which smooths the data reuse process.
Understand and choose data→Process and reuse data	Supportive relationship	Based on the understanding of data variables and collective procedures and the choice of appropriate data, data is processed for correct reuse in accordance with research requirements.
Process and reuse data→Understand and choose data	Iterative relationship	When processing and reusing data, a more specific, definite, and micromesh understanding of what is necessary for analysis optimization that influences the data selection.
The iterative property→Data reuse process	Performance relationship	When data reuse encounters barriers, the researcher iterates over the action of the previous step to smooth the process.

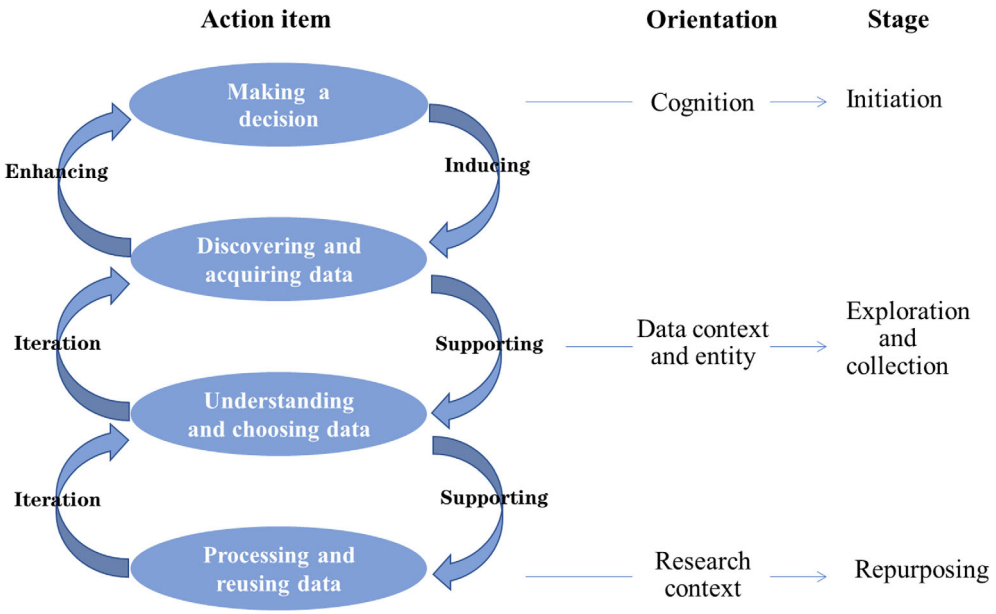

FIGURE 4 Theoretical model of the data reuse process [Color figure can be viewed at wileyonlinelibrary.com]

(Carlson & Anderson, 2007; Curty, 2015; Yoon, 2014b; Yoon & Kim, 2017). In such cases, data reuse plays a minor role in the original or foreground research due to increased perceived risk and reduced perceived availability, leading to a more negative attitude. Therefore, various factors are interwoven, prompting researchers to examine the subtle relationships between data reuse and their research.

5.2.2 | Exploration and collection

In the exploration and collection stage, the researcher performs repeated actions to achieve the optimal solution, including discovering and acquiring data entity, and understanding the data context and choosing appropriate data by seeking relevant or focused information or scaffolds. Thus, the uncertainty about data reuse is

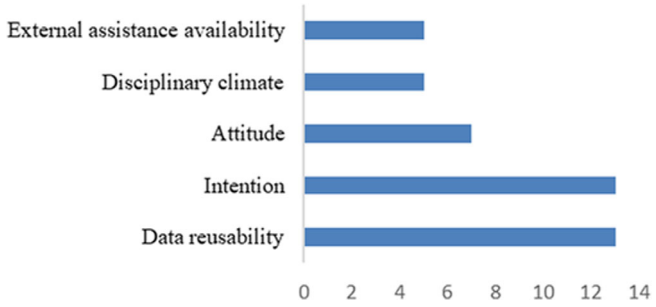

FIGURE 5 Factors involved in data reuse decisions [Color figure can be viewed at wileyonlinelibrary.com]

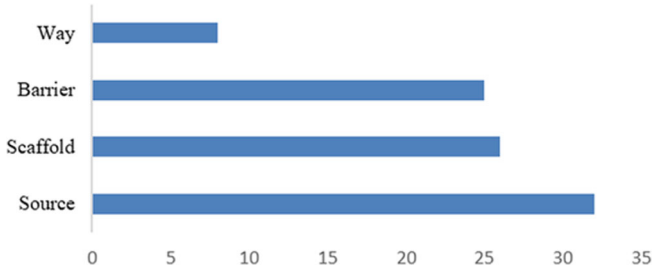

FIGURE 6 The statistics of assistance and barriers [Color figure can be viewed at wileyonlinelibrary.com]

eliminated and leads to enhanced decision positivity for a complete data reuse process, which can be reduced to the following four questions: where to obtain data, what to rely on, how to obtain and understand data, and what obstacles exist.

The above four questions are in accordance with the main categories A6–A14, with the corresponding article number displayed in Figure 6. The main three issues were data source, scaffold, and barriers.

Regarding the concept of source and scaffolds, the literature implies that they share the same contributor in many cases, where personal social networks are prominent. In addition, peer-reviewed repositories and publications, despite a lack of individual researchers and research group repositories, are regarded as particularly common sources because scholars trust them for their quality, volume, and scope (Curty, 2015).

Human scaffolds—such as formal education, experience, data managers, and social networks— and nonhuman scaffolds—such as data documents and publications—act as intermediary parties contributing to identifying and understanding data. Skills and literacy from experience and education are often accumulated by researchers to obtain clues to the existence and understandability of data (Zimmerman, 2007; Zimmerman, 2008). In addition, discussions with mentors or colleagues indicate errors in understanding and help extended datasets, especially when the researcher is inexperienced (Rolland & Lee, 2013). However, an initial contextual understanding of the data, especially the collection program and parameters, relies on data documents, associated peer-reviewed papers, and even primary collectors or data managers of repositories. An assessment of the data quality of a wide range of alternative data in combination with the reputation of the primary collector can help the researcher select the most appropriate data (Curty, 2015; Daniels, 2014; I. Faniel, Kansa, et al., 2013; Niu, 2009).

It is worth mentioning that there are five types of barriers in this stage: insufficient data documentation, lack of accessibility, failure of human scaffolds, lack of specialized knowledge, and technical issues. Three factors cause failure to understand data from data documents. First, data collectors and data reusers use data for different purposes, resulting in different recording requirements for data documents. In addition, some generators are unwilling to expend extra time to provide enough documentation about data collection and processing for secondary use (Niu, 2009). Moreover, important data descriptions are sometimes hidden in the back of metadata records (Murillo, 2016). Lack of accessibility is a common problem. There is less choice in some disciplines due to human data sensitivity, limited accessibility, data citation issues, and other reasons. Considering disciplinary peculiarities, specialized knowledge barriers are particularly significant when accessing and understanding data across disciplines (Yoon, 2014b). These issues force potential reusers to turn to the primary collector or data manager to remove the uncertainty in their understanding, which is a time-consuming process.

Another particularity that merits our focus is that this stage consists of information and data searches for data entity and context, that is, on one hand, the process is meant to search for suitable data entity for reuse, so information behavior is inevitable. On the other hand, information searches not only commit to constructing data context through other resources other than data entity, such as talks with mentors about the data location or data papers, but also involve removing the redundancy in datasets to obtain valuable and relevant data (Rowley, 2007). There are two distinct patterns, referred to in this study as a reticular centripetal pattern and a decentralized centripetal pattern, as illustrated in Figure 7.

In a reticular centripetal pattern, datasets from different sites are linked. Researchers often turn to data consortia, data aggregation websites, or mentors instead of retrieving and obtaining data from multiple sites separately. However, the pattern depends strongly on aggregators' ability to reconstruct context and integrate information from multiple fields and multiple online or

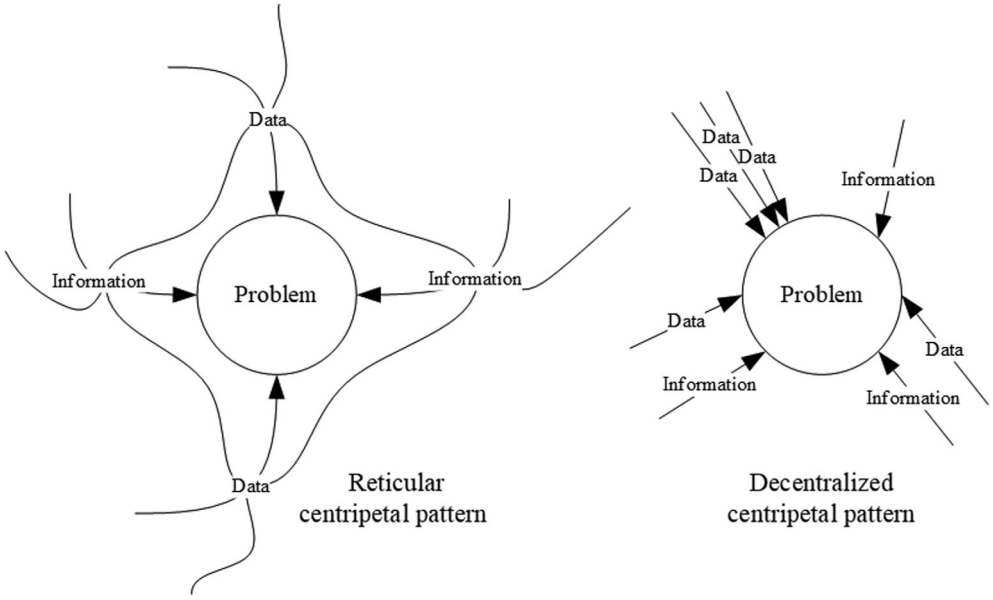

FIGURE 7 Patterns of information and data searches in the data reuse process

physical sites (Daniels, 2014; I. Faniel, Kansa, et al., 2013). In a decentralized centripetal pattern, researchers make a list of possible data or information sources and search for data, browsing and obtaining the content from each source separately (Whitmore, 2016). This requires the reuser's familiarity with relevant data, information skills, and integration skills. Generally, the two patterns can be used interchangeably during data retrieval.

5.2.3 | Repurposing

During the repurposing stage, data are processed to match the new research purpose and then reused, with more focused data understanding, underpinned by professional ability, sometimes required.

On the premise of the particularity of specific disciplines being excluded, the literature hinted at six basic processing procedures: comparison, granularity compensation/loss, aggregation, disaggregation, format transformation, coding, and extraction, as illustrated in Table 3. Generally, one data set does not accomplish the reusability goals, making it necessary for researchers to compare the similarities and differences between predetermined data through documents or parameters (Spreng, 2017), then to balance the granularity differences among multisource data and integrate them. Moreover, the entire data collection does not always fit the requirements, and researchers must extract the required parts and transform the format and granularity (Whitmore, 2016). Particularly in the social sciences, to perform quantitative analyses, the data must be encoded again (I. M. Faniel et al., 2012). Generally, before the data can be used formally in secondary analysis, specific data processing procedures are necessary, which are often related to the accuracy of analyses or the success rate of research results.

Due to the demands for multisource data and professional ability, the two most common barriers are poor data quality, especially poor interoperability (Atici et al., 2013; Curty, 2015; Kriesberg et al., 2012; Poole, 2017; Rolland & Lee, 2013; Stvilia et al., 2015; Tenopir et al., 2011; Weiskopf, 2015; Whitmore, 2016; Yoon, 2015; Zimmerman, 2007, 2008) and lack of skills (Whitmore, 2016). Thus, data standards and cross-disciplinary cooperation are two issues of great concern.

5.3 | Primary elements fitting together

The theoretical model demonstrates the commonalities in the data reuse process, revealing the stages of data reuse. However, the realization of these stages requires further clarification. The above coding results indicate the following elements:

- Reuser or researcher, serving as the agent of cognition and behavior;
- Data, serving as the patient of behavior;
- Data center/data source and data documents and publications, serving as the vectors of the data context and data entity;
- Human scaffolds (individual, team, community forum, and institution entity), education context (research experience and formal education accepted), research context (the intentions and needs of research), and community climate (attitudes to reuse) serving as the mediators to facilitate behavior.

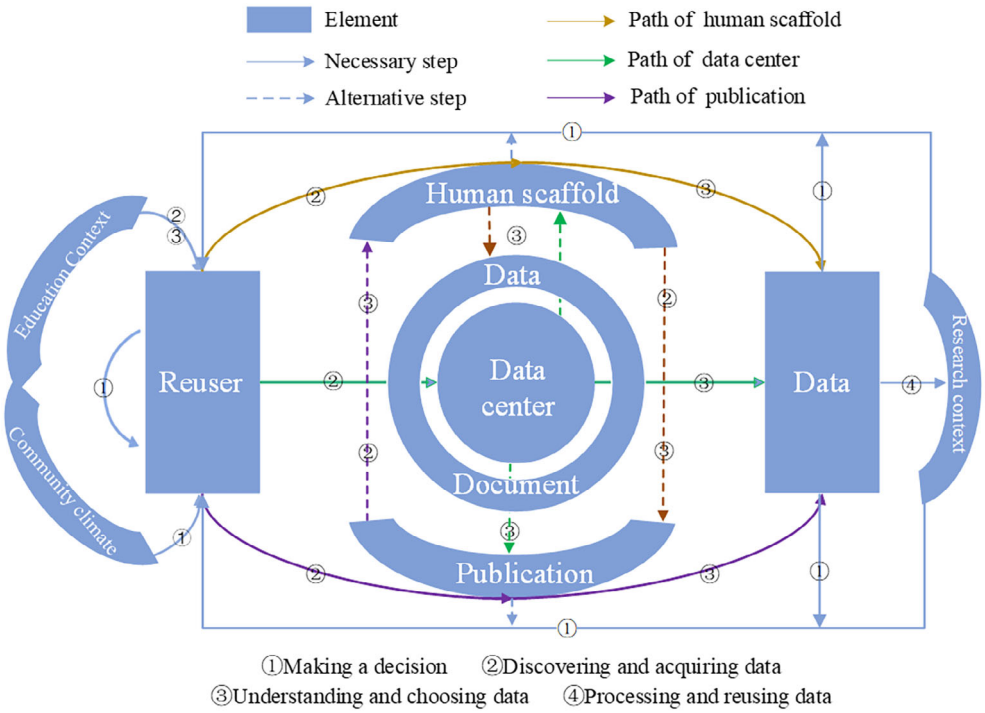

FIGURE 8 Paths based on elements cooperating [Color figure can be viewed at wileyonlinelibrary.com]

As demonstrated in Figure 8, these elements connect in which action items in the theoretical model provide the connection impetus to achieve the action orientation.

Figure 4 summarizes the four action items and three orientations—cognition, data context and entity, and research context. As illustrated in Figure 8, elements—community climate, reuser (reuser's attitude), and data (data conditions from the reuser's impressions)—engage in action item ① to reach a cognition orientation; and both the selected data and research context are interlinked through action item ④, contributing to the research context orientation. However, the same does not apply for data context and entity orientation. Action items ② and ③ connect different elements to reach the data context and entity orientation, divided into necessary steps and alternative steps. Alternative steps extend actions and enhance the efficiency of necessary steps when the latter is insufficient. The following three paths were extracted for data reuse based on differences between elements undertaking necessary steps.

5.3.1 | The path of the data center

In this path, data sources are identified as data consortia, data aggregation websites, institutional repositories, or disciplinary databases, with particular emphasis on the role of freely available data documents organized to ease the understanding and choosing of data, all of which directly contribute to action items ② and ③. In these instances, human scaffolds and publications act as auxiliary discovery or understanding mediators to perform alternative steps (Curty, 2015; Daniels, 2014; Yoon, 2015). In some cases, with data from different sources contained in data consortia, a database can serve as a convenient approach to discover data (Daniels, 2014) so that the reticular centripetal pattern occurs. A high-profile characteristic in data consortia is a dual data source, where some data from a data center have limited open access availability, and the primary collector must be contacted to obtain permission or primary data and data documents (Yoon, 2015). However, these data are easy to use. Part of a well-built data center provides analysis tools, such as visualization and data comparison tools (Curty, 2015; I. Pasquetto, 2018). Among different researcher groups, senior researchers with more experience in reusing data give higher priority to this path than less experienced researchers (Curty, 2015).

5.3.2 | The path of human scaffold

In this path, data sources are primary researchers, individual teams, and personal websites, with reliance on the corresponding data documentation from the primary collector or publications to understand data, all of which directly contribute to action items ② and ③ and reach the data context and entity orientation illustrated in Figure 4. Such data are frequently shared among

colleagues, mentors, students, and project team members on a small scale (Rolland & Lee, 2013; Yoon, 2015). In particular, a collector's help is often required due to the lack of assurance about the data documentation quality (Niu, 2009), causing the original collector to likely play a triple role in the discovery, understanding, and data sources. Publications may also assume responsibility for the first two roles except for the data source, as an element to perform the alternative steps of ② and ③ under the help of a mentor or original collector. Novice researchers may prefer this to compensate for their lack of experience (Yakel et al., 2013), relying primarily on senior researchers to discover and understand data, as senior researchers are more supportive of the accumulated social network.

5.3.3 | The path of publication

Data are extracted from existing publications or appendices, or are converted from publications. In this path, publication plays a central role in locating and acquire targeting data; in other words, it is the immediate executors of action items ② and ③ and a determinant of the realization of the data context and data entity orientation. Mentors and discipline communities are regarded as significant human scaffolds, helping reusers to understand and choose data, which constitutes the alternative steps.

As presented above, when researchers follow different paths to carry out data reuse, the elements involved and their functions are flexible, except for the fixed position of elements contributing to the cognition and research context orientations. It implies that elements fit together through promoting action items and matching functions, and a well-developed data reuse infrastructure and atmosphere permit the activation of necessary functions.

6 | DISCUSSION

This study reviewed 42 researches that reported data reuse behaviors. From the perspective of the reuser and information behavior, and in consideration of the unique characteristics of the data reuse process, this study argues that data reuse can be divided into three stages: initiation, exploration and collection, and repurposing. Researchers can follow the paths of data centers, human scaffolds, and publications, in which the reticular centripetal pattern and decentralized centripetal pattern are used interchangeably to search for data entity or data context. In the following, we clarify the contributions of this review toward answering three questions proposed in the Introduction section.

6.1 | RQ1: What stages do researchers go through when reusing existing data?

Due to the lack of clear discussion and systematic description of the data reuse process, the significance of which has been suggested by previous studies, we posed this question. Based on inductive coding, five main categories and their relationships were derived. On these grounds, our review constructed the theoretical model, clarified the actions and objects commons in the data reuse process, and put forward the entire flow from initiation to repurposing for reuse and the relationship between stages, which improves the conceptual framework of data reuse.

Additionally, this review adds to the data reuse research by providing a new perspective on the effect of information behavior. Previous research has focused on data reuse and information behavior; however, the relationship between them has not been clearly pointed out (Rolland & Lee, 2013; Yoon, 2014b; Zimmerman, 2007). The study holds that they are similar in tasks and activities, such as recognizing the need, retrieval, selection, acquisition, and use, but different in goals and specific actions. The goal of data reuse is to capture data and data-related information. At the level of action, information behavior ends actions if the information collected is deemed useful (Kuhlthau, 1991), whereas data reuse is generally considered to end only after the data collected are processed and reanalyzed.

6.2 | RQ2: What are the specifics of each stage?

Although many previous articles have described the experiences and problems of different researcher groups in the data reuse process, our review summarized the current research and practice situation that enriches the data process from the previous theory model to reality. The following conclusions were drawn. (a) at the initiation stage, trust in open data is a particularly significant factor that affects a decision on whether to carry out reuse. (b) In the exploration and collection stage, data sources, scaffolds, and barriers dominate research and practice, which implies demand for smoothing the approach to acquiring data and searching for help. Human scaffolds attached to social networks are seen as aids in discovering and understanding data. We also identified two universal information and data-seeking methods: the reticular centripetal pattern, particularly emphasizing the perfections of infrastructure, and the decentralized centripetal pattern, underlining the capacity of reusers, such as the familiarity with disciplinary

data and information skills. (c) In the repurposing stage, multisource data are favored by reusers, which suggests the significance of linked-data publishing and data standards.

6.3 | RQ3: How do the primary elements fit together?

The previous studies have focused on specific elements or activities (Huggett, 2018; Weiskopf & Weng, 2013; Williams et al., 2017), with a less explicit emphasis on the role of context relationships among various elements and their interaction in various contexts. Thus, we explored this issue and examined how the elements fit together in specific steps and stages to complete the reuse process. Our review found that researchers practice the reuse process in three paths, in which elements play different roles in different context relationships. Therefore, they connect through the induction of action items requirements and the functional matching degree of different elements. This enriches the context perspective of data reuse research.

6.4 | Implications

This study has practical applications for data infrastructure construction, data reuse, and data reuse research. The action items and orientation clarification of the data reuse process provide a clear direction for improving data sharing and publishing, including data context-oriented, research context-oriented, smooth channel-oriented data reuse, and enhancing the open data environment for reuse. From a practical viewpoint, three paths were proposed in this review, emphasizing the importance of the connection between data entities, data documentation, and human scaffolds. The path centering on data centers received the most attention, which reflects that reusers demand access centralization, including data entity and data context. Therefore, data publishing institutions should provide smooth access to dual licensing channels, conduct data quality assurance, and connect data to data documents and published literature using that data. However, the centralization of access tends to mean selecting and merging data from large-scale datasets. Thus, tools are provided to assist reuse demand trends in data comparison, multisource data merging, and preliminary analysis, except for the selection of data by information retrieval. Data interoperability is required to be enhanced through the standardization of the organization framework. In the academic community, the second path centers on human scaffolds, demonstrating the demands and dependence of academic cooperation for novices and trust in data acquired through trusted relationships, revealing insufficient data literacy and data quality concerns. Accordingly, data literacy and data reusability perception of researchers must be improved. In addition, the theoretical model helps identify needs and tasks for reusers, and the clarification of element functions and matching paths exhibit the alternatives to data reuse, particularly for novices. Moreover, the theoretical model implies interesting data reuse topics, such as feelings and thoughts research in the data reuse process, the relationship between the research context and data context, and the possibility of research data reuse from information behavior.

6.5 | Limitations

This review has three main limitations. We regarded each concept in the articles as equally important; therefore, the importance of concepts and their categories in different disciplines, among novice and experienced groups, and in other different situations cannot be identified. In the coding process, the core concepts were extracted following the terminology that appeared in original articles. There are inevitable differences in the connotation and extension of terms used in different articles, thus reducing the accuracy of the extracted concepts. In future, more scrutiny is required to remedy these deficiencies. As the authors of the selected articles were mostly from the information science domain, our review embodies the information science perspective on data reuse to a large extent, and future studies should enhance the universality of the proposed data process in the fields of natural sciences and humanities.

REFERENCES

Atici, L., Kansa, S. W., Lev-Tov, J., & Kansa, E. C. (2013). Other people's data: A demonstration of the imperative of publishing primary data. *Journal of Archaeological Method and Theory*, *20* (4), 663–681.

Ball, A., & Duke, M. (2015). *How to track the impact of research data with metrics (DCC How-to Guides)*. http://www.dcc.ac.uk/resources/how-guides/track-data-impact-metrics

Benitez-Paez, F., Degbelo, A., Trilles, S., & Huerta, J. (2018). Roadblocks hindering the reuse of open geodata in Colombia and Spain: A data user's perspective. *ISPRS International Journal of Geo-Information*, *7*, 6.

Birnholtz, J. P., & Bietz, M. J. (2003). Data at work: Supporting sharing in science and engineering. *Proceedings of the 2003 International ACM SIGGROUP Conference on Supporting Group Work* (pp. 339–348), Sanibel Island, Florida, USA.

Boté, J. J., & Térmens, M. (2019). Reusing data: Technical and ethical challenges. *DESIDOC Journal of Library & Information Technology*, *39*(6), 329–337.

Calamai, S., & Frontini, F. (2018). FAIR data principles and their application to speech and oral archives. *Journal of New Music Research*, *47*(4), 339–354.

Callaghan, S. (2015). Data without peer: Examples of data peer review in the earth sciences. *D-Lib Magazine*, *21*(1), 9.

Carlson, S., & Anderson, B. (2007). What are data? The many kinds of data and their implications for data re-use. *Journal of Computer-Mediated Communication*, *12*(2), 635–651.

Castle, J. E. (2003). Maximizing research opportunities: Secondary data analysis. *Journal of Neuroscience Nursing*, *35*(5), 287–290.

Catalano, A. (2013). Patterns of graduate students' information seeking behavior: A meta-synthesis of the literature. *Journal of Documentation*, *69*(2), 243–274.

Curty, R. G. (2015). *Beyond "data thrifting": An investigation of factors influencing research data reuse in the social sciences* (Doctoral dissertation, Syracuse University). Available from ProQuest Dissertation and thesis database. (UMI No. 3713677).

Curty, R. G., & Qin, J. (2014). Towards a model for research data reuse behavior. *Proceedings of the American Society for Information Science and Technology*, *51*(1), 1–4.

Custers, B., & Uršič, H. (2016). Big data and data reuse: A taxonomy of data reuse for balancing big data benefits and personal data protection. *International Data Privacy Law*, *6*(1), 4–15.

Daniels, M. G. (2014). *Data reuse in museum contexts: Experiences of archaeologists and botanists* (Doctoral dissertation, University of Michigan). Available from ProQuest Dissertation and thesis database. (UMI No. 3636549).

Edwards, J., & Kaimal, G. (2016). Using meta-synthesis to support application of qualitative methods findings in practice: A discussion of meta-ethnography, narrative synthesis, and critical interpretive synthesis. *The Arts in Psychotherapy*, *51*, 30–35.

Ellis, D. (1989). A behavioral approach to information retrieval system design. *Journal of Documentation*, *45*(3), 171–212.

Fane B. (2019). What is the state of open data in 2019?. In Figshare (Ed.), *The State of Open Data 2019*. https://digitalscience.figshare.com/articles/The_State_of_Open_Data_Report_2019/9980783

Faniel, I., Barrera-Gomez, J., Kriesberg, A., & Yakel, E. (2013). A comparative study of data reuse among quantitative social scientists and archaeologists. *Proceedings of iConference 2013* (pp. 797–800), Texas, USA.

Faniel, I., Kansa, E., Whitcher K. S., Barrera-Gomez, J., & Yakel, E. (2013). The challenges of digging data: A study of context in archaeological data reuse. *Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries* (pp. 295–304), Indianapolis, Indiana, USA.

Faniel, I. M., & Jacobsen, T. E. (2010). Reusing scientific data: How earthquake engineering researchers assess the reusability of colleagues' data. *Computer Supported Cooperative Work*, *19*(3–4), 355–375.

Faniel, I. M., Kriesberg, A., & Yakel, E. (2012). *Data reuse and sensemaking among novice social scientists*. Paper presented at the Proceedings of the American Society for Information Science and Technology, Baltimore, MD.

Faniel, I. M., Kriesberg, A., & Yakel, E. (2016). Social scientists' satisfaction with data reuse. *Journal of the Association for Information Science and Technology*, *67*(6), 1404–1416.

Fear, K. M. (2013). *Measuring and anticipating the impact of data reuse* (Doctoral dissertation, University of Michigan). https://deepblue.lib.umich.edu/bitstream/handle/2027.42/102481/kfear_1.pdf?sequence=1&isAllowed=y

Federer, L. M. (2019). *Who, what, when, where, and why? Quantifying and understanding biomedical data reuse* (Doctoral dissertation, University of Maryland). Available from ProQuest Dissertation and thesis database. (UMI No. 13860396).

Federer, L. M., Lu, Y. L., Joubert, D. J., Welsh, J., & Brandys, B. (2015). Biomedical data sharing and reuse: Attitudes and practices of clinical and scientific research staff. *PLoS One*, *10*(6), e0129506.

Frank, R. D., Tyler, A. R., Gault, A., Suzuka, K., & Yakel, E. (2019). Privacy concerns in qualitative video data reuse. *International Journal of Digital Curation*, *13*(1), 47–72.

Hinds, P. S., Vogel, R. J., & Clarke-Steffen, L. (1997). The possibilities and pitfalls of doing a secondary analysis of a qualitative data set. *Qualitative Health Research*, *7*(3), 408–424.

Hsu, L., Martin, R. L., McElroy, B., Litwin-Miller, K., & Kim, W. (2015). Data management, sharing, and reuse in experimental geomorphology: Challenges, strategies, and scientific opportunities. *Geomorphology*, *244*, 180–189.

Huggett, J. (2018). Reuse remix recycle: Repurposing archaeological digital data. *Advances in Archaeological Practice*, *6*(2), 93–104.

Joo, S., Kim, S., & Kim, Y. (2017). An exploratory study of health scientists' data reuse behaviors. *Aslib Journal of Information Management*, *69*(4), 389–407.

Joo, Y. K., & Kim, Y. (2017). Engineering researchers' data reuse behaviours: A structural equation modelling approach. *The Electronic Library*, *35*(6), 1141–1161.

Kim, Y., & Yoon, A. (2017). Scientists' data reuse behaviors: A multilevel analysis. *Journal of the Association for Information Science and Technology*, *68*(12), 2709–2719.

Kriesberg, A., Frank, R. D., Faniel, I. M., & Yakel, E. (2013). *The role of data reuse in the apprenticeship process*. Paper presented at the Proceedings of the American Society for Information Science and Technology, Montreal, Canada.

Kuhlthau, C. C. (1991). Inside the search process: Information seeking from the user's perspective. *Journal of the American Society for Information Science*, *42*(5), 361–371.

Law, M. (2005). Reduce, reuse, recycle: Issues in the secondary use of research data. *IASSIST Quarterly*, *29*(1), 5–10.

Murillo, A. P. (2016). *Data sharing and data reuse: An investigation of descriptive information facilitators and inhibitors* (Doctoral dissertation, The University of North Carolina at Chapel Hill). Available from ProQuest Dissertation and thesis database. (UMI No. 10245454).

Niu, J. (2009). *Overcoming inadequate documentation*. Paper presented at the Proceedings of the American Society for Information Science and Technology, Hawaii, USA.

Pasquetto, I. (2018). *From open data to knowledge production: Biomedical data sharing and unpredictable data reuses* (Doctoral dissertation, University of California, Los Angeles). Available from ProQuest Dissertation and thesis database. (UMI No. 10977195).

Pasquetto, I. V., Randles, B. M., & Borgman, C. L. (2017). On the reuse of scientific data. *Data Science Journal*, *6*(8), 1–9.

Pettigrew, K. E., Fidel, R., & Bruce, H. (2001). Conceptual frameworks in information behavior. *Annual Review of Information Science and Technology*, *35*, 43–78.

Poole, A. H. (2017). "A greatly unexplored area": Digital curation and innovation in digital humanities. *Journal of the Association for Information Science and Technology*, *68*(7), 1772–1781.

Rolland, B., & Lee, C. P. (2013, February). Beyond trust and reliability: Reusing data in collaborative cancer epidemiology research. *Proceedings of the 2013 Conference on Computer Supported Cooperative Work* (pp. 435–444), San Antonio, TX, USA.

Rowley, J. (2007). The wisdom hierarchy: Representations of the dikw hierarchy. *Journal of Information Science*, *33*(2), 163–180.

Sands, A., Borgman, C. L., Wynholds, L., & Traweek, S. (2012). *Follow the data: How astronomers use and reuse data*. Paper presented at the Proceedings of the American Society for Information Science and Technology, Baltimore, MD.

Sandt, S. V. D., Dallmeier-Tiessen, S., Lavasa, A., & Petras, V. (2019). The definition of reuse. *Data Science Journal*, *18*(1), 22.

Schaaf, J., Kadioglu, D., Goebel, J., Behrendt, C. A., & Storf, H. (2018). OSSE goes Fair - implementation of the fair data principles for an open-source registry for rare diseases. *Studies in Health Technology and Informatics*, *253*, 209–213.

Shen, Y. (2015). Research data sharing and reuse practices of academic faculty researchers: A study of the virginia tech data landscape. *International Journal of Digital Curation*, *10*(2), 157–175.

Song, G., & Hu, W. (2016). The study of mandatory open scientific data policies and implementation suggestions. *Library and Information Service*, *60*(9), 61–69.

Song, M., Liu, K., Abromitis, R., & Schleyer, T. L. (2013). Reusing electronic patient data for dental clinical research: A review of current status. *Journal of Dentistry*, *41*(12), 1148–1163.

Song, X., Li, L., & Zhang, W. (2020). Research on data reuse ecosystem model. *Research on Library Science*, (7), 39–47.

Spink, A. (2006). Human information behavior: Integrating diverse approaches and information use. *Journal of the American Society for Information Science and Technology*, *57*(1), 25–35.

Spreng R. L. (2017). *Methods for the reuse of public data in gene expression studies* (Doctoral dissertation, North Carolina State University). Available from ProQuest Dissertation and thesis database. (UMI No. 10977195).

Strauss, A. (Ed.). (1987). *Qualitative analysis for social scientists*. Cambridge University Press.

Stvilia, B., Hinnant, C. C., Wu, S., Worrall, A., Lee, D. J., Burnett, K., Burnett, G., Kazmer, M. M., & Marty, P. F. (2015). Research project tasks, data, and perceptions of data quality in a condensed matter physics community. *Journal of the Association for Information Science and Technology*, *66*(2), 246–263.

Sun, G., & Khoo, C. S. G. (2016). Social science research data curation: Issues of reuse. *Libellarium: Journal for the Research of Writing, Books, and Cultural Heritage Institutions*, *9*(2), 59–80.

Sun, Y., Cheng, Y., & Xie, J. (2019). A review on the data reuse behavior of scholars: System review and meta synthesis. *Journal of Library Science in China*, *45*(241), 110–130.

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.

Tenopir, C., Christian, L., Allard, S., & Borycz, J. (2018). Research data sharing: Practices and attitudes of geophysicists. *Earth and Space Science*, *5*(12), 891–902.

Tenopir, C., Dalton, E. D., Allard, S., Frame, M., Pjesivac, I., Birch, B., Pollock, D., & Dorsett, K. (2015). Changes in data sharing and data reuse practices and perceptions among scientists worldwide. *PLoS One*, *10*(8), e0134826.

The Digital Curation Centre (DCC). (2019). *An Analysis of Open Data and Open Science Policies in Europe v4*. https://zenodo.org/record/3379705#.XuNn2Pkzbb0

The Future of Research Communication and e-Science 11. (2016). *FAIR principles*. https://www.force11.org/group/fairgroup/fairprinciples

Vita, R., Overton, J. A., Mungall, C. J., Sette, A., & Peters, B. (2018). FAIR principles and the IEDB: Short-term improvements and a long-term vision of OBO-foundry mediated machine-actionable interoperability. *Database: The Journal of Biological Databases and Curation*, *2018*, bax105.

Wei, J. (2017). *Qualitative data sharing practice in social science* (Doctoral dissertation, University of Pittsburgh). Available from ProQuest Dissertation and thesis database. (UMI No. 10645840).

Weiskopf, N. G. (2015). *Enabling the reuse of electronic health record data through data quality assessment and transparency* (Doctoral dissertation, Columbia University). Available from ProQuest Dissertation and thesis database. (UMI No. 3667382).

Weiskopf, N. G., & Weng, C. (2013). Methods and dimensions of electronic health record data quality assessment: Enabling reuse for clinical research. *Journal of the American Medical Informatics Association*, *20*(1), 144–151.

Whitmore, D. A. (2016). *Seeking context: Archaeological practices surrounding the reuse of spatial information* (Doctoral dissertation, University of California, Los Angeles). Available from ProQuest Dissertation and thesis database. (UMI No. 10125041).

Whyte, A., & Pryor, G. (2011). Open science in practice: Researcher perspectives and participation. *The International Journal of Digital Curation*, *6*(1), 199–212.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Bouwman, J. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*, 160018.

Wilkinson, M. D., Verborgh, R., da Silva Santos, L. O. B., Clark, T., Swertz, M. A., Kelpin, F. D., Gray, A. J. G., Schultes, E. A., van Mulligen, E. M., Ciccarese, P., Kuzniar, A., Gavai, A., Thompson, M., Kaliyaperumal, R., Bolleman, J. T., & Dumontier, M. (2017). Interoperability and FAIRness through a novel combination of Web technologies. *PeerJ Computer Science*, *3*, e110.

Williams, R., Kontopantelis, E., Buchan, I., & Peek, N. (2017). Clinical code set engineering for reusing EHR data for research: A review. *Journal of Biomedical Informatics*, *70*, 1–13.

Wilson, T. D. (2016). A general theory of human information behaviour. *Information Research*, *21*(4), 20–23.

Yakel, E., Faniel, I. M., Kriesberg, A., & Yoon, A. (2013). Trust in digital repositories. *The International Journal of Digital Curation*, *8*(1), 143–156.

Yoon, A. (2014a). End users' trust in data repositories: Definition and influences on trust development. *Archival Science*, *14*(1), 17–34.

Yoon, A. (2014b). *"Making a square fit into a circle": Researchers' experiences reusing qualitative data*. Paper presented at the Proceedings of the American Society for Information Science and Technology, Seattle, USA.

Yoon, A. (2015). *Data reuse and users' trust judgments: Toward trusted data curation* (Doctoral dissertation, University of North Carolina at Chapel Hill). Available from ProQuest Dissertation and thesis database. (UMI No. 3719920).

Yoon, A. (2017). Data reusers' trust development. *Journal of the Association for Information Science and Technology*, *68*(4), 946–956.

Yoon, A., & Kim, Y. (2017). Social scientists' data reuse behaviors: Exploring the roles of attitudinal beliefs, attitudes, norms, and data repositories. *Library & Information Science Research*, *39*(3), 224–233.

Yoon, A., & Lee, Y. Y. (2019). Factors of trust in data reuse. *Online Information Review*, *43*(7), 1245–1262.

Zeng, D., & Xue, S. (2013). *Model analysis in information behavior*. Paper presented at International Conference on Computer Sciences and Applications, Wuhan, China.

Zhang, S., Liu, J., Gu, L., Cui, W., & Zhang, Z. (2018). A research on the rights and interests issues of research data reuse. *Documentation, Information & Knowledge*, *94*, 105–113.

Zimmerman, A. S. (2003). *Data sharing and secondary use of scientific data: Experiences of ecologists* (Doctoral dissertation, University of Michigan). Available from ProQuest Dissertation and thesis database. (UMI No. 3079559).

Zimmerman, A. S. (2007). Not by metadata alone: The use of diverse forms of knowledge to locate data for reuse. *International Journal on Digital Libraries*, *7*(1–2), 5–16.

Zimmerman, A. S. (2008). New knowledge from old data: The role of standards in the sharing and reuse of ecological data. *Science, Technology & Human Values*, *33*(5), 631–652.

How to cite this article: Wang X, Duan Q, Liang M. Understanding the process of data reuse: An extensive review. *J Assoc Inf Sci Technol*. 2021; 1–22. https://doi.org/10.1002/asi.24483