



Kharazmi University



Research Data Evaluation And Selection Criteria For Preservation In Data Repositories

Kimiya Taghizadeh Milani¹ | Masoumeh Karbala Aghaei Kamran^{2✉} | Amir Ghaebi³

1. PhD Candidate, Department of Information science, Faculty of Education and Psychology, Alzahra University, Tehran. Iran E-mail: kimiyamilani@gmail.com
2. Corresponding author, Associate Professor, Department of Information science, Faculty of Education and Psychology, Alzahra University, Tehran. Iran E-mail: mkamran@alzahra.ac.ir
3. Associate Professor, Department of Information science, Faculty of Education and Psychology, Alzahra University, Tehran. Iran E-mail: ghaebi@alzahra.ac.ir

Article Info

Article type:

Research Article

Article history:

Received 20 September 2024

Received in revised form

10 February 2025

Accepted 13 February 2025

Published online 15 March

2025

Keywords:

Research data,
research data management,
research data evaluation,
research data selection,
data management.

ABSTRACT

Purpose: The primary objective of this study is to identify criteria for evaluating and selecting research data for preservation in data repositories based on global authoritative literature.

Method: This research was conducted using the meta-synthesis method and the Sandelowski and Barroso (2007) framework. Relevant keywords were utilized to search for studies in databases of various publishers and through citation tracking to ensure comprehensive coverage of identified sources. After multi-stage screening and applying inclusion and exclusion criteria, 48 English-language sources were selected from an initial pool of 3,080 sources. Additionally, the Kappa coefficient was employed to ensure the quality of extracted codes and maintain the reliability of the designed model.

Findings: The criteria for evaluating and selecting research data for preservation in data repositories were categorized into eight components, including Data preparation, data quality, physical conditions and technical features, metadata management and characteristics, ethical principles of data, document-related criteria, alignment with FAIR principles, and repository policies and issues, encompassing a total of 52 indicators.

Conclusion: This study identified criteria for evaluating and selecting research data for preservation in data repositories based on global authoritative literature. Applying these criteria can lead to sustainable preservation of data in repositories, enhance data quality, and strengthen research infrastructure.

Keywords: Research data, research data management, research data evaluation, research data selection, data management, data repositories

Cite this article: Taghizadeh Milani, K., Karbala Aghaei Kamran, M., & Ghaebi, A. (2024). Research Data Evaluation And Selection Criteria For Preservation In Data Repositories. *Human-Information Interaction*, 11(4), 90-111.

© The Author(s). Publisher: University of Kharazmi.





Kharazmi University



Human-Information
INTERACTION

Extended Abstract

Introduction

The rapid proliferation of digital data in the research landscape has underscored the critical need for sustainable data curation strategies, especially regarding the long-term preservation of valuable datasets. Research data repositories, as key infrastructures for data stewardship, face mounting challenges in determining which datasets should be preserved for future reuse, validation, and scientific advancement. Given the constraints of storage, funding, and technical resources, not all data generated by research activities can or should be preserved indefinitely. Consequently, defining rigorous, transparent, and contextually appropriate evaluation and selection criteria has emerged as a vital concern within the broader scope of research data management (RDM) and digital curation.

This study aims to identify and categorize the key criteria used to evaluate and select research data for long-term preservation in repositories. By conducting a systematic review of existing literature and practices, it seeks to offer a conceptual framework that supports repository managers, librarians, archivists, and data stewards in making informed and consistent decisions about what data to retain. The research further addresses the implications of these criteria on policy development, data sharing, and FAIR data principles (Findable, Accessible, Interoperable, and Reusable). Ultimately, the study contributes to improving data lifecycle management strategies and ensuring that preserved data retains its scientific, legal, ethical, and cultural value.

Methods and Material

This research adopted a qualitative content analysis approach based on a systematic literature review. The primary goal was to identify, classify, and synthesize the evaluation and selection criteria applied by data repositories in preserving research data. The review focused on peer-reviewed journal articles, white papers, policy documents, and institutional guidelines published between 2000 and 2024. Major databases such as Scopus, Web of Science, ScienceDirect, and Google Scholar were searched using combinations of keywords including "research data preservation," "data selection criteria," "data curation," and "digital repositories."

Inclusion criteria for the literature involved the presence of explicit or implicit discussion on the assessment or selection of research data for long-term storage, including frameworks, models, or institutional case studies. A total of 67 relevant documents were identified and analyzed. Through iterative coding and constant comparison, the evaluation criteria were distilled into several thematic clusters, such as scientific value, legal and ethical considerations, technical characteristics, economic feasibility, data usability, and policy alignment.

Resultss and Discussion

The findings of this study, based on a systematic review of the literature and a meta-synthesis of previous studies, identify a comprehensive set of criteria and components for



Kharazmi University

Journal of Human-Information Interaction

Online ISSN: 2423-7418

<https://hi.khu.ac.ir/>



evaluating and selecting research data for retention in data repositories. These criteria are categorized into eight main components: data preparation, data quality, physical conditions and technical features, metadata management and features, ethical principles of data, document-related criteria, compliance with FAIR principles, and repository policies and issues.

In the “Data Preparation” component, indicators such as data cleaning, data scale, presence of missing data, and evaluation of survey biases are highlighted. This component emphasizes the necessity of eliminating errors and inconsistencies, assessing the scale of data, and addressing missing values. It also stresses the importance of identifying and evaluating biases in survey data, such as sampling errors, non-response, and other confounding factors.

The “Data Quality” component includes indicators such as accuracy, reliability, completeness, validity, documentation of limitations, and timeliness of data. Accuracy and correctness of information must be carefully assessed, and data reliability should be evaluated based on how the data was produced and analyzed. Completeness refers to the presence of all necessary elements in the dataset, and validity relates to the soundness of data collection tools and the extent to which findings reflect reality. Acknowledging study limitations helps clarify weaknesses, and up-to-date data are valued for their relevance in terms of collection time.

The remaining components and their indicators are as follows:

- **Physical Conditions and Technical Features of Data:** Includes data formats, future readability, required software for access, and compatibility with technical standards.
- **Metadata Management and Features:** Covers the presence of sufficient metadata, use of standardized structures for data description, supplementary documentation, and necessary information for data reuse.
- **Ethical Principles of Data:** Encompasses protection of participants' privacy, anonymization or encryption of sensitive information, obtaining informed consent, and respect for intellectual property rights.
- **Document-Related Criteria:** Includes the association of data with specific research projects, traceability of data to published scholarly articles, and documentation of data collection methods.
- **Compliance with FAIR Principles:** Covers Findability, Accessibility, Interoperability, and Reusability of the data.
- **Repository Policies and Issues:** Involves adherence to legal requirements and repository policies, access licenses, data sharing conditions, and security considerations for data storage.

These eight components and their corresponding indicators provide a comprehensive and evidence-based framework for evaluating and selecting suitable research data for long-term retention in data repositories.



Kharazmi University



Discussion and Conclusion

The discussion and conclusion of this paper emphasize the importance of various components in the evaluation and selection of research data for storage in data repositories. In the data preparation phase, accuracy in data cleaning and screening, particularly in quantitative research, is crucial. Challenges such as missing data and potential biases, such as sampling errors, can complicate analyses and reduce the quality of the data. Therefore, adherence to precise standards in cleaning and verifying data is essential.

In evaluating data quality, accuracy and precision of information, reliability, and completeness of the data are key criteria. Data that is properly collected and analyzed can facilitate more effective research and reuse of data. Especially in both qualitative and quantitative data, the use of standardized formats and compatibility with various systems are significant technical issues that impact storage quality.

Metadata documentation also plays a critical role in data evaluation. Metadata provides essential information about the data, enhancing transparency, collaboration, and trust. Furthermore, adhering to ethical principles, such as obtaining informed consent from participants and ensuring their privacy during the use of data, is crucial. These actions help maintain public trust and prevent misuse of data.

The paper also emphasizes the importance of alignment with the FAIR principles (Findability, Accessibility, Interoperability, and Reusability) in data evaluation. Adherence to these principles ensures that data will be more effective and accessible for future use. Additionally, policies related to data repositories must consider user needs and technical limitations, preserving high-value research data for future use.

The conclusion reveals that the evaluation and selection of research data for storage should be conducted with care and adherence to standardized criteria to improve the quality and effectiveness of data utilization in future research. Furthermore, practical recommendations such as developing data evaluation guidelines, training data specialists, and implementing technological tools to enhance the data evaluation and storage processes are proposed.



Kharazmi University



معیارهای ارزیابی و گزینش داده‌های پژوهشی برای حفظ و نگهداری در مخازن داده

کیمیا تقی زاده میلانی^۱، معصومه کربلاآقایی کامران^۲، امیر غائبی^۳

۱. دانشجوی دکتری گروه علم اطلاعات و دانش‌شناسی، دانشکده علوم تربیتی و روانشناسی، دانشگاه الزهراء(س)، تهران، ایران. رایانامه:

KimiyaMilani@gmail.com

۲. نویسنده مسئول، دانشیار گروه علم اطلاعات و دانش‌شناسی، دانشکده علوم تربیتی و روانشناسی، دانشگاه الزهراء(س)، تهران، ایران.

رایانامه: mkamran@alzahra.ac.ir

۳. دانشیار، گروه علم اطلاعات و دانش‌شناسی، دانشکده علوم تربیتی و روانشناسی، دانشگاه الزهراء(س)، تهران، ایران. رایانامه:

ghaebi@alzahra.ac.ir

اطلاعات مقاله	چکیده
<p>نوع مقاله: مقاله پژوهشی</p> <p>تاریخ دریافت: ۱۴۰۳/۰۶/۳۰ تاریخ بازنگری: ۱۴۰۳/۱۱/۲۲ تاریخ پذیرش: ۱۴۰۳/۱۱/۲۵ تاریخ انتشار: ۱۴۰۳/۱۲/۲۵</p> <p>کلیدواژه‌ها: داده‌های پژوهشی، مدیریت داده‌های پژوهشی، ارزیابی داده‌های پژوهشی، گزینش داده‌های پژوهشی، مدیریت داده‌ها.</p>	<p>هدف: هدف اصلی پژوهش حاضر شناسایی معیارهای ارزیابی و گزینش داده‌های پژوهشی برای حفظ و نگهداری در مخازن داده براساس متون معتبر جهانی است.</p> <p>روش: پژوهش حاضر با استفاده از روش فراترکیب و چارچوب سندلوسکی و باروسو (۲۰۰۷) انجام شده است. در جستجوی منابع از کلیدواژه‌های مرتبط برای یافتن پژوهش‌ها در پایگاه‌های ناشران مختلف و نیز جستجوی استنادی استفاده شده است تا جامعیت منابع شناسایی شده تأمین شود. همچنین با کنترل چند مرحله‌ای و اعمال معیارهای ورود و خروج از ۳۰۸۰ منبع، در نهایت ۴۹ منبع انگلیسی انتخاب شدند. همچنین برای کنترل کیفیت کدهای استخراج‌شده و حفظ پایایی مدل طراحی‌شده از شاخص کاپا استفاده شده است.</p> <p>یافته‌ها: معیارهای ارزیابی و گزینش داده‌های پژوهشی برای حفظ و نگهداری در مخازن داده در هشت مؤلفه شامل آماده‌سازی داده‌ها، کیفیت داده‌ها، شرایط فیزیکی و ویژگی‌های فنی، مدیریت و ویژگی‌های فراداده‌ای، اصول اخلاقی داده‌ها، معیارهای مربوط به مدرک، هماهنگی با اصول FAIR، و خطومشی و مسائل پایگاه شناسایی شد که شامل ۳۸ گویه بود.</p> <p>نتیجه‌گیری: در این پژوهش معیارهای ارزیابی و گزینش داده‌های پژوهشی برای حفظ و نگهداری در مخازن داده براساس متون معتبر جهانی شناسایی شدند. به‌کارگیری این معیارها می‌تواند به حفظ و نگهداری پایدار داده‌ها در مخازن داده، ارتقای کیفیت داده‌ها و تقویت زیرساخت‌های پژوهشی منجر شود.</p> <p>کلیدواژه‌ها: داده‌های پژوهشی، مدیریت داده‌های پژوهشی، ارزیابی داده‌های پژوهشی، گزینش داده‌های پژوهشی، مدیریت داده‌ها، مخازن داده‌ها</p>

استناد: تقی زاده میلانی، کیمیا؛ کربلاآقایی کامران، معصومه؛ و غائبی، امیر (۱۴۰۳). معیارهای ارزیابی و گزینش داده‌های پژوهشی برای حفظ و نگهداری در مخازن داده. *تعامل انسان و اطلاعات*، ۱۱(۴)، ۹۰-۱۱۱.



مقدمه

داده‌های پژوهشی^۱ به عنوان ستون اصلی پژوهش‌ها عمل می‌کنند و نقش اساسی در پیشرفت علم دارند. داده‌های پژوهشی نه تنها اطلاعات اولیه را فراهم می‌کنند، بلکه به محققان امکان تحلیل، تفسیر و نتیجه‌گیری را می‌دهند. همچنین داده‌های پژوهشی، توسعه نظریه‌ها و مدل‌ها، تأیید یا رد فرضیات و تولید دانش جدید را امکان‌پذیر می‌کنند. از این نظر، مدیریت و اشتراک‌گذاری داده‌های پژوهشی ضرورت پیدا کرده است. با این حال، تخمین زده می‌شود که ۸۰ درصد از داده‌هایی که پژوهشگران برای انجام پژوهش خود جمع‌آوری و به اشتراک می‌گذارند، «بدون استفاده» هستند؛ به دلیل اینکه اکثر مجموعه داده‌ها قابلیت استفاده مجدد ندارند (مارسلک و همکاران^۲، ۲۰۲۳؛ مانز و همکاران^۳، ۲۰۱۷). به نظر می‌رسد یکی از دلایل عدم استفاده از داده‌های پژوهشی در مطالعات ثانویه نادیده گرفته شدن مرحله ارزیابی و گزینش در چرخه مدیریت داده‌های پژوهشی است.

باید در نظر داشت که متولیان مدیریت داده‌های پژوهشی و متصدیان مخازن داده‌های پژوهشی، به دلیل نبود یا ناکافی بودن «دستورالعمل‌های جامع و کمیود دانش ارزیابی و گزینش» در چرخه مدیریت داده‌های پژوهشی، از سیاست‌های ارزیابی و گزینش برای داده‌ها استفاده نمی‌کنند (ماسینده و همکاران^۴، ۲۰۲۱؛ کیم و فالکنر، ۲۰۲۱؛ ال‌جرادات^۵، ۲۰۲۱؛ ماسنیا، ۲۰۲۱، رکانکو و همکاران^۶، ۲۰۲۲).

ابعاد و معیارهای کیفیت داده‌های پژوهشی (به عنوان شاخص‌های ارزیابی و گزینش) در حوزه‌های مختلف موضوعی یکسان نیستند و در مواردی با هم همخوانی ندارند (کیندلینگ و استرکر^۷، ۲۰۲۲)؛ همچنین یکی از چالش‌های اصلی در مدیریت داده‌های پژوهشی علاوه بر فقدان معیارهای ارزیابی و گزینش برای ارزیابان، تفاوت معیارها از سازمانی به سازمان دیگر و از حوزه‌ای به حوزه دیگر متفاوت است (شن و وارول^۸، ۲۰۲۰).

مطالعات نشان می‌دهد معیارهای ارزیابی و گزینش داده‌های پژوهشی در حوزه‌های موضوعی مختلف نیز متفاوت است. بخشی از مطالعاتی که معیارهای و شاخص‌های متفاوت و گوناگونی برای کیفیت داده‌های پژوهشی در حوزه علوم پزشکی گزارش کرده‌اند به شرح زیر است:

کوموتینی و همکاران^۹ (۲۰۱۹) با تدوین دستورالعمل پیشنهادی برای حداقل اطلاعات تحقیقات سکتور مغزی و گزارش داده‌های بالینی؛ آکجو و همکاران^{۱۰} (۲۰۲۲) با مطالعه کیفی استفاده و نیاز به داده‌ها و اطلاعات توسط متخصصان بهداشت مراقبت‌های تسکینی سرطان در جنوب صحرای آفریقا؛ بکر و همکاران^{۱۱} (۲۰۲۲) در مطالعه استفاده ثانویه از داده‌های سلامت شخصی در اتحادیه اروپا؛ وندروالده و همکاران^{۱۲} (۲۰۲۲) با کار بر روی طرحواره ابرداده ژنوم FAIR و توالی نسل‌ها؛ آلوارز-

1. Research data

2. Marsolek et al.

3. Mons et al.

4. Masinde et. Al.

5. Al-Jaradat

6. Recunenko

7. Kindling & Strecker

8. Shen & Varvel

9. Kumuthini et al.

10. Akeju et al.

11. Becker et al.

12. van der Velde et al.

رومرو و همکاران^{۱۳} (۲۰۲۳) با تجزیه و تحلیل ویژگی‌های حاکمیت داده‌های موجود برای تحقیق در مرکز داده‌های سلامت؛ زاس و همکاران^{۱۴} (۲۰۲۳) با توسعه استانداردهای جمع‌آوری داده‌های فنوتیپ بالینی برای تحقیقات در آفریقا؛ کوموتینی و همکاران^{۱۵} (۲۰۲۳) با استانداردهای داده‌ها در زمینه اومیکس؛ سرورا د لا کروز و شعبانی^{۱۶} (۲۰۲۳) مفهوم‌سازی FAIR در استفاده ثانویه از داده‌های سلامت برای تحقیق؛ و ال‌مستاری و همکاران^{۱۷} (۲۰۲۳) با بررسی جنبه‌های فنی و حقوقی مربوط به استفاده مجدد از داده‌های سلامت هنگام استفاده مجدد از آنها.

با اینکه ارزیابی و گزینش داده‌های پژوهشی نقش مهمی در تحقیقات حوزه‌های مختلف علوم دارد (رکانکو و همکاران، ۲۰۲۲) اما پژوهشگران عمدتاً تمرکز خود را بر ذخیره و اشتراک‌گذاری داده‌های پژوهشی قرار داده‌اند و سایر بخش‌های فرایند مدیریت داده‌های پژوهشی به صورت فردی انجام می‌شود (کیم و سین^{۱۸}، ۲۰۲۱). پژوهشگران و متخصصانی که با داده‌های پژوهشی سروکار دارند و آنها را تجزیه و تحلیل می‌کنند به دلیل ابهام در نحوه گردآوری، عدم کیفیت فراداده و ناتوانی فیزیکی و قانونی در دسترسی به داده‌ها، اغلب با مشکل استفاده مجدد از این داده‌ها مواجه هستند.

دستورالعمل‌ها و سیاهه‌های واری که برای ارزیابی و گزینش داده‌های پژوهشی ارائه شده‌اند، اغلب حوزه‌های خاصی از علوم را در برمی‌گیرند و شامل انواع مختلفی از داده‌های پژوهشی برخاسته از حوزه‌های مختلف علمی نیستند و اکثر مخازن داده‌های پژوهشی نیز معیارهای صریحی برای گزینش داده‌ها ندارند (کیم و فالکنر، ۲۰۲۱؛ ماسینده و همکاران، ۲۰۲۱؛ دی.سی.سی. اسکارپ، ۲۰۱۰). شرایط و الزاماتی که برای مدت زمان حفظ و نگهداری داده‌ها ذکر شده است، به بازه زمانی مشخصی اشاره دارد که در آینده منقضی خواهد شد. بنابراین، نیاز به بررسی مجدد تعداد فزاینده‌ای از مجموعه داده‌ها احساس خواهد شد و باید تصمیمات پیچیده‌تری در مورد گزینش مجدد/یا عدم گزینش مجدد داده‌ها اتخاذ شود (دی.سی.ان، ۲۰۲۲).

تاکنون، مطالعاتی که به صورت نظام‌مند به ارائه راهکار در جهت تدوین دستورالعمل و یا راهنماهای جامع و پویا برای نحوه تصمیم‌گیری برای ارزیابی و گزینش داده‌های پژوهشی با در نظر گرفتن روش‌شناسی باشد انجام نشده است (عظیمی و همکاران^{۱۹}، ۲۰۲۳).

وزیری و همکاران (۱۳۹۷) در مقاله خود به این نتیجه رسیدند که نمی‌توان به صورت مشخص مسائل مربوط به اشتراک‌گذاری و مدیریت داده‌های پژوهشی را در اسناد بالادستی کشور رصد کرد. در ایران، سالیانه حجم عظیمی از داده‌های پژوهشی در حوزه‌های مختلف علمی تولید می‌شود، اما سیاست یا برنامه‌ای برای اشتراک‌گذاری و به‌ویژه ارزیابی و گزینش این داده‌ها برای نگهداری وجود ندارد. هرچند نظام نوین اطلاعات پژوهش‌های پزشکی ایران (نوپا)^{۲۰} دارای ۱۲ سامانه از قبیل نتایج و اخبار پژوهش‌های سلامت؛ کتاب‌های دیجیتال؛ مقالات علوم پزشکی؛ داده‌های علم‌سنجی، پایان‌نامه‌های حوزه علوم پزشکی را ذخیره و نگهداری است؛ و یا نمونه آن در وزارت علوم، تحقیقات و فناوری (سامانه میفا^{۲۱}) به وجود دارد؛ اما پایگاهی مختص ذخیره و نگهداری و اشتراک‌گذاری داده‌های پژوهشی وجود ندارد. یکی از دلایل این کمبود، عدم وجود ادبیات کافی و دستورالعمل‌های مناسب در ابعاد مختلف مدیریت داده‌های پژوهشی و همچنین عدم وجود سیاست‌های بالادستی برای آن است. همچنین مجیدی، نقشینه، اسمعیلی‌گیوی و هاشمی و همکاران (۱۳۹۶) اشاره می‌کنند که امروزه گزینش و مدیریت

¹³. Alvarez-Romero et al.

¹⁴. Zass et al

¹⁵. Kumuthini et al.

¹⁶. Cervera De La Cruz & Shabani

¹⁷. El Mestari et al.

¹⁸. Kim & Syn

¹⁹. Azami et al.

²⁰. research.ac.ir

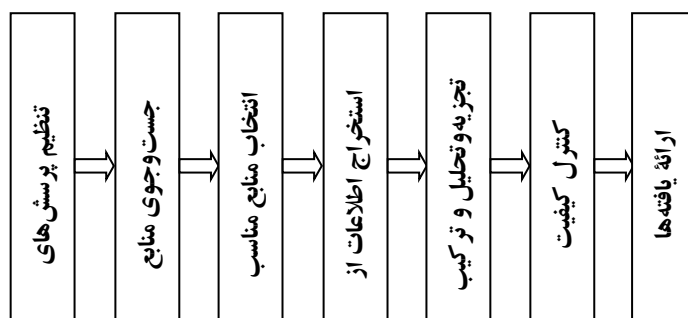
²¹. <https://mapfalogin.msrt.ir>

داده‌های پژوهشی به عنوان یکی از اجزای اصلی سیاست‌گذاری کلان علم و فناوری مطرح شده و نیازمند حمایت دولت و نهادهای قانون‌گذار و سیاست‌گذار است. مدل‌ها و چهارچوب‌های مختلفی در سطوح مختلف ملی، مؤسسه یا اجتماع خاص و رشته برای درک ابعاد گزینش و مدیریت داده‌های پژوهشی و اجرای آن ترسیم شده‌اند. با این حال، گزینش و مدیریت داده‌های پژوهشی در اجرا و عمل با چالش‌هایی روبه‌رو است که یکی از مهم‌ترین آنها عدم آگاهی و شناخت و دانش درباره فرایند ارزیابی و گزینش داده‌های پژوهشی است.

مسئله اصلی در این پژوهش، فقدان معیارها و دستورالعمل‌های جامع و مناسب برای ارزیابی و گزینش داده‌های پژوهشی برای حفظ و نگهداری در مخازن داده است. بحث کیفیت و ارزیابی داده در زنجیره مدیریت داده، قبل از اشتراک‌گذاری آنها قرار دارد. در هنگام جمع‌آوری و مستندسازی داده‌ها، باید اقداماتی در جهت افزایش کیفیت و ارزش داده‌ها انجام شود. به منظور تضمین کیفیت و ارزش نگهداری و استفاده مجدد داده‌های پژوهشی، لازم است یک چارچوب مناسب برای ارزیابی و گزینش داده‌ها تدوین شود. این چارچوب می‌تواند شامل معیارها و روش‌هایی برای ارزیابی داده‌ها، مانند پیاده‌سازی درست روش تحقیق و روش‌های نمونه‌گیری، دقت جمع‌آوری داده‌ها، اخذ رضایت از مشارکت‌کنندگان، فراداده و سایر عوامل مرتبط با داده باشد. در این پژوهش با روش فراترکیب معیارهای ارزیابی و گزینش داده‌ها پژوهشی از متون استخراج می‌شود. سؤال اصلی پژوهش حاضر این است که مؤلفه‌های ارزیابی و گزینش داده‌های پژوهشی برای حفظ و نگهداری در مخازن داده‌های پژوهشی کدام هستند؟

روش پژوهش

پژوهش حاضر به روش فراترکیب^{۲۲} است. به منظور انجام فراترکیب از روش هفت‌مرحله‌ای سندلوسکی و باروسو^{۲۳} (۲۰۰۷) استفاده شد که بیشترین روش مورد استفاده برای انجام پژوهش‌هایی با روش فراترکیب است.



شکل (۱). مراحل هفت‌گانه فراترکیب (سندلوسکی و باروسو، ۲۰۰۷)

22. Meta synthesis

23. Sandelowski & Barroso

مراحل اجرای روش فراترکیب مطابق با شکل زیر است که در ادامه شرح داده می‌شود:

۱. تنظیم پرسش‌های پژوهش: پرسش‌های مبنایی پژوهش به شرح زیر تنظیم شدند.

جدول (۱). پرسش‌های گام نخست فراترکیب

پرسش‌ها	پاسخ
چیستی کار (What)	شناسایی مؤلفه‌های ارزیابی و گزینش داده‌های پژوهشی برای مخازن داده‌های پژوهشی
جامعه مورد مطالعه (Who)	آثار مختلف علمی اعم از مقاله (در مجلات و کنفرانس‌ها) و پایان‌نامه به زبان انگلیسی
محدودیت زمانی (When)	کلیه آثار موجود بین سال‌های ۲۰۰۰ تا ۲۰۲۴ میلادی تا زمان انجام پژوهش
چگونگی روش (How)	بررسی موضوعی آثار، شناسایی و یادداشت کدهای استخراج‌شده، تعیین مفاهیم، دسته‌بندی مفاهیم و ایجاد مقوله‌ها

۲. جست‌وجوی منابع: در این پژوهش، مقالات و پایان‌نامه‌های منتشرشده در نه پایگاه اطلاعاتی انگلیسی بین سال‌های ۲۰۰۰ تا ۲۰۲۴ میلادی جست‌وجو شدند که این بازه زمانی و تعدد پایگاه‌های مطرح و مهم مورد بررسی سبب دستیابی به آثار مناسب، به‌روز، جامع و متعددی در این حوزه شد. کلیدواژه‌های متنوعی مطابق با جدول ۲ برای جست‌وجوی منابع در نظر گرفته شد تا بتوان حداکثر بازبازی ممکن را میسر ساخت. علاوه بر این کلیدواژه‌ها با توجه به گستردگی حوزه، از دو پایگاه استنادی اسکوپوس و گوگل اسکالر برای بررسی زنجیره استنادی منابع یافت شده استفاده شد. همچنین منابع و مآخذ مقالات یافت شده نیز مورد بررسی قرار گرفت و منابع مرتبط استخراج شدند. این جست‌وجوها در فروردین تا شهریورماه ۱۴۰۳ انجام شدند.

جدول (۲). کلیدواژه‌های مورد استفاده در جست‌وجوی منابع

کلیدواژه‌ها
research data
secondary research/ secondary analysis
data reuse/ repurposing
Secondary Data
Data quality
Assessment of data
initial data analyses (IDA)
Qualitative secondary analysis
Quality assurance of data
Data Seliction/ Data Evaluation
ترکیبی از موارد بالا

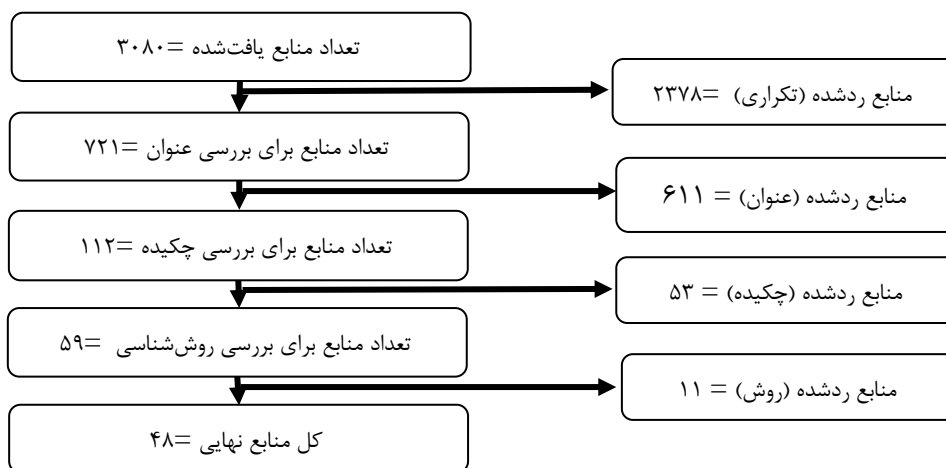
کلیدواژه‌های ذکرشده در بالا همراه با ترکیبات مختلف آن‌ها، متناسب با شرایط موردنظر در پژوهش حاضر و راهبرد جست‌وجوی پیشرفته در هریک از پایگاه‌های اطلاعاتی ذکرشده در جدول ۳-۳ وارد شد. بعد از حذف منابع تکراری بازبازی شده در هر پایگاه، تعداد منابع بازبازی شده در هریک از پایگاه‌های اطلاعاتی مورد جست‌وجو به شرح جدول ۳ به دست آمد. در مجموع، ۳۰۸۰ منبع برای بررسی و غربالگری بازبازی شدند.

جدول (۳). پایگاه‌های اطلاعاتی مورد جست‌وجو

نام پایگاه	تعداد مقاله	تعداد پایان‌نامه
Emerald	۲۳	۰
Sage	۶۱	۰
Wiley	۱۰۱	۰
Taylor and Francis	۱۰۲	۰
Science Direct	۲۰۳	۰
Jstore	۴۵	۰
ProQuest	۱۳۶	۲۱
Scopus	۸۷۱	۰
google Scholar	۱۵۳۶	۰
مجموع	۳۰۷۸	۲۱

کل مدارک بازیابی شده: ۳۰۸۰

۳. انتخاب منابع مناسب: در مرحله نخست منابع تکراری کنار گذاشته شدند. انتخاب منابع بر اساس معیارهای ورود و خروج به فراترکیب و از طریق غربالگری عنوان، چکیده، محتوا و روش‌شناسی انجام شد. معیارهای ورود به فراترکیب پژوهش‌هایی بودند که در آن‌ها به مؤلفه‌ها، ابعاد و عوامل اثرگذار بر ارزیابی و گزینش داده‌های پژوهشی به‌روش پژوهشی کیفی یا آمیخته پرداخته شده بود. معیارهای خروج از فراترکیب شامل پژوهش‌هایی بود که به انتخاب داده‌ها برای پژوهش‌هایی با مقصود خاص پرداخته‌اند. با در نظر گرفتن معیارهای ورود و خروج، پالایش و غربالگری پژوهش‌ها به‌ترتیب بر اساس عنوان، چکیده، محتوا و روش پژوهش انجام شد. در هر مرحله غربالگری، تعدادی از منابع حذف شدند (شکل ۳-۲).



شکل (۲). شیوه‌شناسایی و انتخاب آثار در مرحله فراترکیب

از مجموع ۳۰۸۰ منبع، در نهایت ۴۸ منبع مورد تأیید قرار گرفت. از این ۴۸ منبع، ۴۱ منبع مقاله پژوهشی، پنج منبع مقاله همایشی و سه منبع نیز پایان‌نامه بودند.

۴. استخراج اطلاعات از متون: پس از انتخاب منابع مناسب، با مطالعه چندباره کل متن ۴۸ منبع منتخب، اطلاعات آنها شامل اطلاعات کتاب‌شناختی، اطلاعات روش‌شناختی و همچنین مؤلفه‌ها یا موضوعات و متغیرهای مهم مورد اشاره در منابع استخراج شد. در این بخش، تعداد ۵۲ کد یا معیار منحصر به فرد مربوط به ارزیابی و گزینش داده‌های پژوهشی شناسایی شد.

۵. تجزیه و تحلیل و ترکیب یافته‌ها: در این مرحله، کدهای استخراج شده در گام قبل بر اساس مشابهت در دسته‌ای تحت عنوان مفاهیم قرار گرفتند، و مفاهیم مشابه نیز تحت عنوان یک مقوله دسته‌بندی شدند.

۶. کنترل کیفیت کدهای استخراجی: برای کنترل کیفیت کدهای استخراج شده و حفظ پایایی مدل طراحی شده از شاخص کاپا^{۲۴} استفاده شد، به این ترتیب که از شخصی دیگر درخواست شد دسته‌بندی اطلاعات استخراج شده از متون را بدون اطلاع از دسته‌بندی قبلی انجام دهد. سپس، دسته‌بندی ایشان با دسته‌بندی پژوهشگران مقایسه و شاخص کاپا محاسبه شد. ضریب کاپا اندازه‌ای بین ۱- تا ۱+ است. هرچه میزان به دست آمده به ۱+ نزدیک‌تر باشد، نشان‌دهنده توافقی بیشتر بین دو کدگذار است. چنانچه شاخص مورد نظر بیشتر از ۰/۶۱ باشد، از مطلوبیت و معتبر بودن برخوردار است. در این راستا، متون مورد بررسی برای کدگذاری در اختیار یکی از خبرگان (دکتری رشته علم اطلاعات و دانش‌شناسی و عضو هیئت علمی دانشگاه با تجربه تدوین مقالاتی با استفاده از روش فراترکیب) قرار گرفت و ضریب توافقی دو کدگذار با استفاده از ضریب کاپا در نرم‌افزار اسپاس^{۲۵} محاسبه شد؛ با توجه به سطح معناداری که برابر با ۰/۰۰۰ و کوچک‌تر از ۰/۰۵ است و میزان ضریب کاپا که ۰/۸۷ و بیشتر از ۰/۶ بود، رابطه‌ای معنادار بین نظرات دو گروه کدگذار وجود داشته و بنابراین استخراج کدها از پایایی مناسبی برخوردار است.

۷. ارائه یافته‌ها: در این مرحله، یافته‌های به دست آمده در مراحل قبل ارائه شد.

یافته‌های پژوهش

بر اساس مرور متون و فراترکیب نتایج معیارهای ارزیابی و گزینش داده‌های پژوهشی برای حفظ و نگهداری در مخازن داده در هشت مؤلفه آماده‌سازی داده‌ها، کیفیت داده‌ها، شرایط فیزیکی و ویژگی‌های فنی، مدیریت و ویژگی‌های فراداده‌ای، اصول اخلاقی داده‌ها، معیارهای مربوط به مدرک، هماهنگی با اصول FAIR^{۲۵}، و خطومشی و مسائل پایگاه شناسایی شد.

جدول (۶). مؤلفه آماده‌سازی داده‌ها

مؤلفه	گویه	منابع
آماده‌سازی داده‌ها	پاک‌سازی داده‌ها	Lusa et al., 2023; Stvilia & Lee, 2024;
	مقیاس داده‌ها	Panchenko & Samovilova, 2020;
	داده‌های از دست رفته	Manu et al., 2021; Kumuthini, 2019;
	ارزیابی سوگیری‌های پیمایش‌ها	Stratton, 2015; Lusa et al., 2023; Rabianski 2003; Zuiderwijk et al., 2024;

مؤلفه «آماده‌سازی داده‌ها» از مراحل مهم ارزیابی و گزینش داده‌های پژوهشی برای مخازن است. پاک‌سازی داده‌ها (شناسایی خطاها و ناسازگاری‌ها)، غربالگری داده‌ها و به‌روزرسانی‌های احتمالی تحلیل‌های آماری با دقت انجام شده باشد. برای

^{۲۴}. Kappa Indicator

^{۲۵}. Findability, Accessibility, Interoperability, and Reusability

داده‌های کمی مقیاس‌های جزئی داده‌ها مهم است و امکان تبدیل به مقیاس‌های کلی را می‌دهد. ممکن است بخش‌هایی در مجموعه داده‌های موجود وجود نداشته باشد (مثلاً مجموعه‌های داده از دست رفته در داده‌های سری زمانی) که تخمین مقادیر فرضی را دشوار می‌سازد. ارزیابی سوگیری‌های تحقیقات پیمایشی در گزینش داده‌های پژوهشی تأثیر دارد. این سوگیری‌ها شامل خطای نمونه‌گیری، حاشیه خطای آماری، خطای عدم پاسخ‌دهی، سوگیری پاسخ و نوسان اندازه‌گیری، ویژگی‌های نمونه پژوهش، تعداد نمونه از دست رفته، اطلاعات مربوط به عوامل مخدوش‌کننده، خلاصه زمان پیگیری و اقدامات خلاصه در طول زمان است. همچنین ارزیابان باید خطورود سوگیری به برخی از مجموعه‌های داده از قبل موجود (مانند داده‌های رسانه‌های اجتماعی، داده‌های دولتی) را به دلیل دستکاری احتمالی داده‌ها برای اهداف سیاسی در عصر اخبار جعلی و پسا حقیقت در نظر بگیرند.

جدول (۷). مؤلفه کیفیت داده‌ها

مؤلفه	گویه	منابع
کیفیت داده‌ها	دقت	Niu, 2014; Cai & Zhu, 2015; Kindling & Strecker, 2022; Manu et al., 2021; Rabianski, 2003; Stvilia & Lee, 2024; Sherif, 2018;
	قابلیت اعتماد	Boté-Vericad & Térmens i Graells, 2019; Rolland & Lee, 2013; Yoon, 2017; Manu et al., 2021; Rabianski, 2003; Stvilia & Lee, 2024; Kumuthini, 2019; Sherif, 2018; Niu, 2014; Mannheimer, 2022; Batini & Scannapieco, 2016;
	کامل بودن	Gregory et al., 2020; Kindling & Strecker, 2022; Key et al., 2023; Stvilia & Lee, 2024; Oza, 2023; Peng et al. 2022;
	اعتبار	Sherif, 2018; Harris, 2001; Mannheimer, 2022; Oza, 2023; Manu et al., 2021; Rabianski, 2003; Stvilia & Lee, 2024; Kumuthini, 2019; Busetto et al., 2020;
	ذکر محدودیت‌ها	Sherif, 2018; Manu et al., 2021; Stvilia & Lee, 2024; Manu et al., 2021;
	به روز بودن	Manu et al., 2021; Rabianski, 2003; Omukuti, 2021; Sherif, 2018; Corti & Backhouse, 2005; Whiteside et al., 2012; Woolard et al., 2004;

«کیفیت داده‌ها» مؤلفه دیگری است که ارزیابان باید با وسواس بیشتری آن را بررسی کنند. دقت و صحت اطلاعات ارائه شده را می‌توان با مشاهده جزئیات کشف کرد. باید اطمینان حاصل شود که داده‌های کیفی و کمی ارائه شده برای مخازن دقیق هستند. سطح قابل اعتماد بودن داده‌ها عامل مهمی برای محققینی است که می‌خواهند از داده‌های تحقیق مجدد استفاده کنند. آنها باید بتوانند به قابلیت اطمینان داده‌هایی که قرار است استفاده کنند اعتماد کنند. قابل اعتماد بودن به این موضوع اشاره دارد که آیا فرآیند مورد استفاده برای ایجاد داده‌ها معتبر است، آیا داده‌ها به طور مداوم تولید می‌شوند و آیا روش‌های تجزیه و تحلیل دقیق برای اطمینان از معتبر بودن داده‌ها برای استفاده مجدد دنبال می‌شوند. شواهد برای قابل اعتماد بودن داده‌ها اغلب به ارزیابی داده‌ها و اصول اخلاقی مطالعه اصلی محدود می‌شود. معیار دیگر کیفیت داده‌ها کامل بودن آنها است یعنی تمام اجزای لازم در داده‌ها وجود داشته باشد. کامل بودن مجموعه داده شامل شرایط فنی نیز است. پرداختن به مسائلی مانند تشخیص فرمت‌های دیجیتال می‌تواند مشکل‌ساز باشد. از آنجایی که فرمت‌های دیجیتالی متفاوتی وجود دارد، داده‌ها (به‌ویژه داده‌های کیفی) ممکن است قالب‌های مختلفی داشته باشند (مانند داده‌های گفتاری، تصاویر و ویدئوها) و ممکن است همیشه استاندارد نباشند. اعتبار داده‌ها به روایی ابزار بستگی دارد و میزان و درجه اعتماد به واقعی بودن یافته‌ها برای شرکت‌کنندگان در پژوهش است. تشریح محدودیت‌های پژوهش می‌تواند محقق ثانویه را در اعتماد به واقعی بودن داده‌ها کمک کند. در نهایت برای تکمیل کیفیت پژوهش ارزیابان باید داده‌هایی را گزینش کنند که به‌روز باشد یا احتمال انقضای تاریخ داده‌ها کم‌تر باشد.

جدول (۸). مؤلفه شرایط فیزیکی و ویژگی‌های فنی

مؤلفه	گویه	منابع
شرایط فیزیکی و ویژگی‌های فنی	همسانی و سازگاری داده‌ها	Kindling & Strecker, 2022; El Mestari, 2023; Key et al., 2023; Stvilia & Lee, 2024; Busetto et al., 2020
	ساختاربندی و یکپارچگی داده‌ها	Niu, 2014; Vuokko et al., 2017; Kumuthini et al., 2023; Alvarez-Romero et al., 2023; El Mestari, 2023; Zuiderwijk et al., 2024; Key et al., 2023; Stvilia & Lee, 2024
	ویژگی داده‌ها	Lusa et al., 2023; El Mestari, 2023

مؤلفه «شرایط فیزیکی و ویژگی‌های فنی داده‌ها» از نظر پایگاه‌ها شرط لازم برای ذخیره داده‌های پژوهشی است. این معیار شامل همسانی و سازگاری داده‌ها، ساختاربندی و یکپارچگی داده‌ها، و ویژگی داده‌ها است. برای منابع دیجیتال نه تنها محتوا، شرایط فیزیکی و ویژگی‌های فنی نیز باید ارزیابی شود، زیرا آنها مستقیماً بر هزینه‌ها و امکان نگهداری داده‌ها تأثیر می‌گذارند. داده‌ها برای ذخیره باید همسانی داشته و تناقض یا ناسازگاری نداشته باشند. داده‌ها باید در قالب باز و غیر اختصاصی در دسترس باشند. همچنین ویژگی‌های داده بر تجزیه و تحلیل و تفسیر نتایج تأثیر دارد.

جدول (۹). مؤلفه مدیریت و ویژگی‌های فراداده‌ای

مؤلفه	گویه	منابع
مدیریت و ویژگی‌های فراداده‌ای	جامعیت فراداده‌ها	Gregory et al., 2020; Mckenna-Foster, 2022; Kumuthini, 2019; van der Velde et al., 2022; Stvilia & Lee, 2024;
	مناسب بودن فراداده‌ها	Kindling & Strecker, 2022; Wilkinson et al., 2016; Oza, 2023; Kumuthini, 2019;
	کیفیت فراداده‌ها	Kindling & Strecker, 2022; Lusa et al., 2023; Stvilia & Lee, 2024; Sakai et al., 2023;
	توصیف دقیق فراداده‌ها	Corti et al. 2019; Assante et al., 2016; Koltay, 2020; Lusa et al., 2023; Key et al., 2023; Stvilia & Lee, 2024;
	فراداده مالکیت داده‌ها	Gregory et al., 2020; Omukuti, 2021; Mannheimer, 2022;

ارزیابان و متصدیان پایگاه‌های اشتراک داده می‌دانند که بررسی فراداده قبل از تصمیم به دانلود داده‌ها اتفاق می‌افتد و در تصمیم به استفاده محققان مؤثر است. فراداده به داده‌ها اجازه می‌دهند تا راحت‌تر یافت شوند. فراداده در مجموعه داده‌ها نه تنها قابلیت همکاری را تسهیل می‌کند، بلکه اطلاعاتی درباره بافت آن نیز فراهم می‌کند. حتی شیوه‌های ضعیف مستندسازی فراداده‌ها می‌تواند منجر به سوء استفاده یا تفسیر نادرست از داده‌ها شود.

در مقابل، شیوه‌های مستندسازی خوب و استفاده از استانداردها، قابلیت همکاری بین انواع مختلف مجموعه داده‌ها را تسهیل می‌کند. علاوه بر این، باعث اعتماد، شفافیت و قابل تأیید می‌شود. فراداده‌ها باید جامع بوده و غنای کافی را داشته باشند. لازم است آنها اطلاعاتی نظیر روش‌های نمونه‌گیری، وزن‌دهی، قوانین رمزگذاری مجدد، الگوهای پرش، متغیرهای ساخته شده و روش‌های گردآوری داده‌ها را در برگیرند. تناسب فراداده‌ها برای داده‌های مختلف باید در نظر گرفته شود، فراداده‌های پژوهش‌های کمی با فراداده‌های پژوهش‌های کیفی تفاوت‌هایی دارند، همچنین رشته‌های موضوعی مختلف ممکن است اقتضائات و فراداده‌های خاص خود را داشته باشند. همچنین دقت در توصیف فراداده‌ها و کیفیت آنها نیز اهمیت دارد. کیفیت

فراداده بر ارزش منابع تأثیر می‌گذارد. مالکیت داده‌های پژوهشی باید به روشنی قابل تشخیص باشد تا محققان ثانویه بتوانند به راحتی به داده‌ها استناد کنند.

جدول (۱۰). مؤلفه اصول اخلاقی داده‌ها

مؤلفه	گویه	منابع
اصول اخلاقی داده‌ها	مجوز استفاده مؤلف	Kindling & Strecker, 2022; Stvilia & Lee, 2024; Stvilia & Lee, 2024;
	کسب رضایت آگاهانه	Boté-Vericad & Térmens i Graells, 2019; Yen et al. 2016; Grady, 2015; Stvilia & Lee, 2024;
	حفظ گمنامی شرکت‌کنندگان / رعایت حریم خصوصی	Boté-Vericad & Térmens i Graells, 2019; Gregory et al., 2020; Busetto et al., 2020; Mannheimer, 2022; Mannheimer, 2022; Bradshaw et al., 2023; Zuiderwijk et al., 2024;
	قانونی بودن داده‌ها	Boté-Vericad & Térmens i Graells, 2019; Busetto et al., 2020; Oza, 2023; Bradshaw et al., 2023; El Mestari, 2023; Stvilia & Lee, 2024;
	اخلاقی بودن داده‌ها (رعایت قوانین مرتبط با داده‌ها)	Busetto et al., 2020; Kumuthini, 2019; Bradshaw et al., 2023; El Mestari, 2023;
	مالکیت معنوی داده‌ها	Boté-Vericad & Térmens i Graells, 2019; Rantasaari, 2021; Stvilia & Lee, 2024;

یکی از مؤلفه‌های کلیدی در ارزیابی و گزینش داده‌های پژوهشی، «اصول اخلاقی» مرتبط با داده‌ها است که شامل چندین جنبه حیاتی است. نخست، ارائه مجوز پژوهشگران اولیه است که پژوهشگران ثانویه باید به آنها استناد کنند. کسب رضایت آگاهانه از شرکت‌کنندگان برای انتشار و استفاده مجدد از داده‌ها اهمیت دارد. این مؤلفه شامل انواع مختلف رضایت آگاهانه است که بسته به موقعیت و نوع پژوهش ممکن است متفاوت باشد. هفت نوع مختلف رضایت آگاهانه در ادبیات یافت شده است: رضایت صریح، رضایت پویا، رضایت فردی، رضایت فراتر، رضایت برای تماس، موافقت‌نامه رضایت با پزشک عمومی و انصراف. همچنین، حفظ گمنامی و حریم خصوصی شرکت‌کنندگان از اصول اساسی است که باید رعایت شود تا از هرگونه آسیب احتمالی به شرکت‌کنندگان جلوگیری شود. در این راستا، ناشناس‌سازی داده‌ها و دقت در به اشتراک‌گذاری داده‌های حساس ضروری است. علاوه بر این، رعایت قوانین مرتبط با مالکیت معنوی و استفاده از مجوزهای کپی‌رایت و کرییتیو کامانز^{۲۶} برای اطمینان از قانونی و اخلاقی بودن داده‌ها در فرآیند ارزیابی و گزینش داده‌های پژوهشی اهمیت زیادی دارد.

جدول (۱۱). مؤلفه معیارهای مربوط به مدرک

مؤلفه	گویه	منابع
معیارهای مربوط به مدرک	اعتبار نویسنده	Fecher et al., 2015; Panchenko & Samovilova, 2020;
	توجه به تأثیر داده‌های پژوهشی	Boté-Vericad & Térmens i Graells, 2019;
	داوری داده‌ها	Niu, 2014; Costello, 2013; Costello et al., 2013; Boté-Vericad & Térmens i Graells, 2019; Koltay, 2020; Downs, 2021; Filip et al., 2022; Busetto et al., 2020;

²⁶. Creative Commons

یکی از مؤلفه‌های مهم در ارزیابی و گزینش داده‌های پژوهشی، «معیارهای مرتبط با خود مدرک» است. این معیار شامل اعتبار نویسنده و گردآورنده داده‌ها، ردیابی تأثیر داده‌ها و داوری آن‌ها است. اعتبار نویسنده نقش مهمی در افزایش استفاده و استناد به داده‌ها دارد و همکاری علمی را تقویت می‌کند، به طوری که در برخی موارد نویسنده داده‌ها به‌عنوان همکار مقاله نیز دعوت می‌شود، به‌ویژه در مطالعات کیفی که نویسنده اطلاعات عمیقی از بستر پژوهش دارد. همچنین، ردیابی تأثیر داده‌ها با استفاده از آلت‌متریک‌ها و دستگاه‌های آماری برای شناسایی الگوهای استفاده مجدد و تعداد دانلودها به سنجش کاربردی و مفید بودن داده‌ها کمک می‌کند. برای داده‌های علمی، مشارکت متخصصان موضوع برای ارزیابی ارزش فعلی و بالقوه آینده داده‌ها ضروری است. داوری داده‌ها انگیزه‌ای برای افزایش رویت‌پذیری ایجاد می‌کند. از این رو، پیشنهاد شده است که داده‌های تحقیق نیز در حین فرایند داوری انتشار مقالات و انتشارات علمی، مشمول کنترل کیفیت ویرایشی و یک سیستم مستقل بررسی هم‌تا باشد.

جدول (۱۲). مؤلفه هماهنگی با اصول FAIR

مؤلفه	گویه	منابع
هماهنگی با اصول FAIR	اشاره به اصول چهارگانه FAIR	Boté-Vericad & Térmens i Graells, 2019; Gregory et al., 2020; Mckenna-Foster, 2022; Rantasaari, 2022; van der Velde et al., 2022; Zuiderwijk et al., 2024; Rehnert & Takors, 2023; Stvilia & Lee, 2024; Harper, 2023; Mckenna-Foster et al., 2022; Harper, 2023;
	قابلیت یافتن	Omukuti, 2021; Niu, 2014; Kindling & Strecker, 2022; Kumuthini et al., 2023; Oza, 2023;
	قابلیت دسترسی	Oza, 2023;
	قابلیت تعامل / همکاری	Palmer, 2011; Niu, 2014; Kindling & Strecker, 2022;
	قابلیت استفاده مجدد	

یکی از مؤلفه‌های اساسی در ارزیابی و گزینش داده‌های پژوهشی برای مدیریت و ذخیره‌سازی، هماهنگی با اصول FAIR است. این اصول که به‌منظور ارتقاء کیفیت مدیریت داده‌ها تدوین شده‌اند، چهار ویژگی کلیدی را برای داده‌های پژوهشی تعیین می‌کنند: قابلیت یافتن، قابلیت دسترسی، قابلیت تعامل و همکاری، و قابلیت استفاده مجدد. هر یک از این ویژگی‌ها به‌طور مستقیم بر ارزش و کارایی داده‌های پژوهشی تأثیرگذار است و در فرآیند ارزیابی اهمیت دارد. برای قابلیت یافتن، داده‌ها باید با شناسه‌های منحصر به فرد و فراداده‌های دقیق به‌راحتی قابل شناسایی باشند. قابلیت دسترسی مستلزم وجود مکانیزم‌های شفاف و بدون موانع غیرضروری برای دسترسی به داده‌ها است. همچنین، قابلیت تعامل و همکاری به معنای امکان یکپارچگی داده‌ها با سیستم‌ها و نرم‌افزارهای دیگر از طریق استانداردهای مشترک است. در نهایت، قابلیت استفاده مجدد داده‌ها از طریق مستندسازی کامل و استفاده از استانداردهای مناسب، امکان تفسیر، فهم و بهره‌برداری مجدد توسط سایر محققان را فراهم می‌سازد.

جدول (۱۳). خط‌مشی و مسائل پایگاه

مؤلفه	گویه	منابع
خط‌مشی و مسائل	نیازهای کاربران	Niu, 2014
پایگاه	خط‌مشی پایگاه	Niu, 2014
	هزینه نگهداری داده‌ها	Niu, 2014

Niu, 2014	ارزش و سودمندی داده‌ها
Niu, 2014	منحصربه‌فرد بودن
Busetto et al., 2020	قابلیت استفاده پژوهشی

مؤلفه « خطمشی و مسائل مرتبط با مخازن داده‌ها » نقش مهمی در فرآیند ارزیابی و گزینش داده‌های پژوهشی ایفا می‌کنند. درحالی‌که نیازهای کاربران همچنان یکی از عوامل کلیدی در حفظ داده‌های پژوهشی محسوب می‌شود، مخازن با محدودیت‌هایی در زمینه نگهداری داده‌ها مواجه هستند و همه اسناد قابل ذخیره‌سازی نیستند. خطمشی‌های مشخص هر مخزن می‌تواند به‌طور مستقیم بر راهبردهای ارزیابی و گزینش داده‌ها تأثیر بگذارد. هزینه نگهداری داده‌ها نیز متغیر است و با توجه به پیچیدگی و حجم اطلاعات، ممکن است متفاوت باشد. از سوی دیگر، ارزش و کاربرد داده‌ها نیز تفاوت دارد؛ برخی از داده‌ها به دلیل ماهیت منحصربه‌فردشان اهمیت بیشتری دارند و انتظار می‌رود بیشتر مورد استفاده قرار گیرند. درنهایت، ضروری است که داده‌های گزینش‌شده برای نگهداری، از قابلیت استفاده پژوهشی بالایی برخوردار باشند تا به نحو مؤثری در تحقیقات آتی مورد بهره‌برداری قرار گیرند.

بحث و نتیجه‌گیری

در مؤلفه آماده‌سازی داده‌ها استفاده دقیق و صحیح از روش‌های پاک‌سازی داده‌ها، به‌ویژه در پژوهش‌های کمی، اهمیت زیادی دارد. پاک‌سازی داده‌ها و غربالگری آن‌ها، بخش‌هایی از مؤلفه آماده‌سازی داده‌ها هستند که باید به‌دقت و بدون خطا انجام شوند. در داده‌های کمی، وجود مقیاس‌های جزئی و امکان تبدیل آن‌ها به مقیاس‌های کلی نیز مورد توجه است. با این حال، چالش‌هایی نظیر داده‌های از دست‌رفته، به‌ویژه در داده‌های سری زمانی، می‌تواند تحلیل را پیچیده کند و به‌دقت نتایج لطمه بزند (لوسا و همکاران، ۲۰۲۳). همچنین، سوگیری‌هایی مانند خطای نمونه‌گیری، حاشیه خطای آماری و خطای عدم پاسخ‌دهی می‌توانند کیفیت داده‌ها را تحت تأثیر قرار دهند (استراتن، ۲۰۱۵). به‌علاوه، ویژگی‌های نمونه، اطلاعات درباره عوامل مخدوش‌کننده و مدت‌زمان پیگیری داده‌ها نیز باید به‌درستی مستندسازی شده و در ارزیابی لحاظ شوند. در عصر اخبار جعلی و پسا حقیقت، مجموعه‌های داده‌ای که از منابع مانند رسانه‌های اجتماعی و داده‌های دولتی تهیه می‌شوند، ممکن است در معرض دست‌کاری قرار گیرند (رابیانسکی، ۲۰۰۳) و این امر می‌تواند بر صحت داده‌ها تأثیر بگذارد. ارزیابان باید به این خطرات توجه ویژه‌ای داشته باشند.

بحث در مورد معیار کیفیت داده‌ها نشان می‌دهد که چگونه ارزیابی دقیق و جامع کیفیت داده‌ها می‌تواند تأثیر زیادی بر استفاده مجدد از داده‌ها داشته باشد. کیفیت داده‌ها یکی از معیارهای کلیدی در ارزیابی و گزینش داده‌های پژوهشی است که باید با وسواس خاصی مورد بررسی قرار گیرد. دقت و صحت اطلاعات، دو جنبه اصلی کیفیت داده‌ها هستند که می‌توان آن‌ها را با بررسی دقیق جزئیات و اطمینان از صحت و دقت داده‌ها مشخص کرد. برای داده‌های کیفی و کمی، این دقت به معنای تأیید این موضوع است که داده‌ها به‌طور صحیح جمع‌آوری شده و به‌درستی تحلیل شده‌اند. قابل اعتماد بودن داده‌ها یکی دیگر از عوامل مهم در ارزیابی کیفیت است. محققانی که قصد دارند از داده‌های تحقیق مجدد استفاده کنند، باید اطمینان حاصل کنند که مؤلفه تولید داده‌ها معتبر و قابل اعتماد بوده است. این امر شامل بررسی مداوم بودن تولید داده‌ها و دقت روش‌های تجزیه و تحلیل است که به اعتبار داده‌ها کمک می‌کند (بوت-ورکاد و ترنز آی گرالز، ۲۰۱۹). شواهد مربوط به قابل اعتماد بودن داده‌ها معمولاً به اصول اخلاقی مطالعه اصلی و ارزیابی دقیق آن محدود می‌شود (یون، ۲۰۱۷). کامل بودن داده‌ها به معنای وجود تمامی اجزای لازم در مجموعه داده‌ها است (کیندلینگ و استرکر، ۲۰۲۲). این شامل شرایط فنی و تنوع فرمت‌های دیجیتال نیز می‌شود. داده‌ها، به‌ویژه داده‌های کیفی، ممکن است در قالب‌های مختلفی مانند داده‌های گفتاری، تصاویر و ویدئوها وجود داشته

باشند و ممکن است استانداردهای یکسانی نداشته باشند (گرگوری و همکاران، ۲۰۲۰). بنابراین، اطمینان از استاندارد بودن و سازگاری فرمت‌ها اهمیت زیادی دارد. اعتبار داده‌ها نیز به روایی ابزارهای مورد استفاده برای جمع‌آوری داده‌ها و میزان اعتماد به واقعی بودن یافته‌ها بستگی دارد. تشریح محدودیت‌های پژوهش می‌تواند به محققان ثانویه در ارزیابی اعتبار داده‌ها کمک کند. در نهایت، به‌روز بودن داده‌ها و کاهش احتمال انقضای تاریخ آن‌ها، برای تضمین کیفیت پژوهش بسیار حیاتی است. گزینش داده‌هایی که به‌روز و مناسب برای استفاده مجدد باشند، به ارتقای کیفیت پژوهش کمک کرده و امکان بهره‌برداری مؤثر از داده‌ها را تسهیل می‌کنند.

شرایط فیزیکی و ویژگی‌های فنی داده‌ها نقش کلیدی در فرآیند ذخیره‌سازی و نگهداری ایفا می‌کنند. برای منابع دیجیتال، بررسی شرایط فیزیکی و فنی داده‌ها اهمیت زیادی دارد، چراکه این عوامل به‌طور مستقیم بر هزینه‌های نگهداری و امکان دسترسی به داده‌ها تأثیر می‌گذارند (نیو، ۲۰۱۴). داده‌ها باید از نظر ساختاری یکپارچه و همسان باشند تا از بروز مشکلات در تجزیه و تحلیل و استفاده مجدد جلوگیری شود. همچنین، داده‌ها باید در قالب‌های باز و غیراختصاصی ذخیره شوند تا دسترسی و بهره‌برداری از آن‌ها در آینده برای کاربران مختلف ساده‌تر باشد. این مسئله به‌ویژه در زمینه‌هایی که فناوری‌ها و استانداردها به‌سرعت تغییر می‌کنند، از اهمیت بالایی برخوردار است. ویژگی‌های فنی داده‌ها، شامل فرمت‌های مختلف و قابلیت سازگاری با سیستم‌های گوناگون، نیز تأثیر قابل توجهی بر نحوه تجزیه و تحلیل و تفسیر نتایج دارد (لوسا و همکاران، ۲۰۲۳). داده‌هایی که به‌طور صحیح و باکیفیت مناسب ذخیره شده‌اند، می‌توانند به راحتی مورد استفاده قرار گیرند و به محققان امکان می‌دهند تا به نتایج دقیق و قابل اعتمادی دست یابند. در نتیجه، توجه به شرایط فیزیکی و ویژگی‌های فنی داده‌ها نه تنها به حفظ کیفیت و یکپارچگی داده‌ها کمک می‌کند، بلکه مدیریت و نگهداری مؤثر از آن‌ها را در طول زمان‌های مختلف تسهیل می‌سازد.

بحث در خصوص اهمیت فراداده‌ها در ارزیابی و گزینش داده‌های پژوهشی برای ذخیره‌سازی در مخازن داده، به نقش حیاتی فراداده‌ها در تسهیل استفاده و اعتماد به داده‌ها می‌پردازد. فراداده‌ها به‌عنوان داده‌های توصیفی که اطلاعاتی درباره داده‌های اصلی فراهم می‌آورند، نقشی کلیدی در ارزیابی و گزینش داده‌ها ایفا می‌کنند. بررسی دقیق فراداده‌ها پیش از تصمیم‌گیری برای دانلود و استفاده از داده‌ها بسیار مهم است (گرگوری و همکاران، ۲۰۲۰)، چراکه فراداده‌ها امکان شناسایی و دسترسی به داده‌ها را تسهیل می‌کنند. به‌علاوه، فراداده‌ها اطلاعات ضروری درباره بافت داده‌ها را فراهم کرده و موجب تسهیل قابلیت همکاری بین مجموعه‌های مختلف داده می‌شوند. در عین حال، شیوه‌های ضعیف مستندسازی فراداده‌ها می‌تواند منجر به سوءاستفاده یا تفسیر نادرست از داده‌ها شود (آسانته و همکاران، ۲۰۱۶). بنابراین، استفاده از استانداردهای مستندسازی و شیوه‌های صحیح در این زمینه، به تسهیل قابلیت همکاری، افزایش اعتماد، شفافیت و قابل‌تأیید بودن داده‌ها کمک می‌کند. فراداده‌ها باید جامع و دقیق باشند و شامل اطلاعاتی نظیر روش‌های نمونه‌گیری، وزن دهی، قوانین رمزگذاری مجدد، الگوهای پرس، متغیرهای ساخته‌شده و روش‌های گردآوری داده‌ها باشند (نیو، ۲۰۱۴). به‌علاوه، تناسب فراداده‌ها با نوع داده‌ها و رشته‌های موضوعی مختلف باید در نظر گرفته شود، چراکه فراداده‌های پژوهش‌های کمی با فراداده‌های پژوهش‌های کیفی متفاوت هستند و هر رشته موضوعی ممکن است نیاز به فراداده‌های خاص خود داشته باشد. دقت در توصیف و کیفیت فراداده‌ها بر ارزش و کاربرد منابع تأثیر بسزایی دارد. همچنین، شفافیت در مالکیت داده‌ها برای محققان ثانویه اهمیت زیادی دارد، زیرا به آن‌ها امکان می‌دهد به راحتی به داده‌ها استناد کرده و از آن‌ها استفاده کنند.

بحث در خصوص اصول اخلاقی مرتبط با داده‌ها در ارزیابی و گزینش داده‌های پژوهشی، بر اهمیت رعایت استانداردهای اخلاقی در مراحل مختلف مدیریت داده‌ها تأکید می‌کند. این اصول به چندین جنبه کلیدی تقسیم می‌شود که نقش مهمی در حفظ اعتبار و اعتبار داده‌ها دارند. اولین و شاید مهم‌ترین جنبه، ارائه مجوز و استناد به پژوهشگران اولیه است. این مسئله

موجب می‌شود که محققان ثانویه بتوانند به‌درستی از داده‌ها استفاده کنند و به منابع اصلی استناد کنند، که خود از اصول اساسی تحقیق علمی است. علاوه بر این، کسب رضایت آگاهانه از شرکت‌کنندگان برای انتشار و استفاده مجدد از داده‌ها نیز از اهمیت بالایی برخوردار است (بوت-ورکاد و ترنز آی گرالز، ۲۰۱۹). رضایت آگاهانه به شرکت‌کنندگان این امکان را می‌دهد که از اهداف و نحوه استفاده از داده‌های خودآگاه باشند و تصمیم‌گیری آگاهانه‌ای در این خصوص داشته باشند. این رضایت ممکن است به انواع مختلفی مانند رضایت صریح، رضایت پویا و یا رضایت برای تماس (اسکوگارد و همکاران، ۲۰۱۹) تقسیم شود، که بسته به نوع تحقیق و نیازهای خاص آن متفاوت است. حفظ گمنامی و حریم خصوصی شرکت‌کنندگان نیز از اصول کلیدی اخلاقی است که باید به‌دقت رعایت شود. جلوگیری از آسیب‌های احتمالی به شرکت‌کنندگان از طریق ناشناس‌سازی داده‌ها و دقت در به اشتراک‌گذاری داده‌های حساس از اهمیت ویژه‌ای برخوردار است. این اقدامات به حفظ اعتماد شرکت‌کنندگان و جلوگیری از مشکلات قانونی و اخلاقی کمک می‌کند. علاوه بر این، رعایت قوانین مرتبط با مالکیت معنوی و استفاده از مجوزهای کپی‌رایت برای اطمینان از قانونی بودن و اخلاقی بودن داده‌ها ضروری است (بوت-ورکاد و ترنز آی گرالز، ۲۰۱۹). این قوانین به تعیین حقوق و مسئولیت‌ها در مورد استفاده از داده‌ها و جلوگیری از نقض حقوق مالکیت معنوی کمک می‌کنند.

بحث در مورد معیارهای مرتبط با خود مدرک در ارزیابی و گزینش داده‌های پژوهشی، به تأثیرات و اهمیت عواملی نظیر اعتبار نویسنده، ردیابی تأثیر داده‌ها، و فرآیند داوری می‌پردازد. اعتبار نویسنده به‌عنوان یک معیار کلیدی، تأثیر بسزایی در افزایش استفاده و استناد به داده‌ها دارد (فچر و همکاران، ۲۰۱۵). پژوهشگران با اعتبار بالا، نه تنها می‌توانند به افزایش اعتبار داده‌های خود کمک کنند، بلکه همکاری‌های علمی بیشتری را نیز جذب خواهند کرد. در پژوهش‌های کیفی، اطلاعات عمیق و زمینه‌شناسی که نویسنده داده‌ها از آن برخوردار است، می‌تواند نقش قابل توجهی در تقویت کیفیت و اعتبار داده‌ها داشته باشد (بوت-ورکاد و ترنز آی گرالز، ۲۰۱۹). به‌این ترتیب، در برخی موارد، نویسنده داده‌ها ممکن است به‌عنوان همکار مقاله نیز دعوت شود، که این خود نشان‌دهنده اهمیت و تأثیر اعتبار نویسنده است. ردیابی تأثیر داده‌ها نیز به‌عنوان یکی از معیارهای مهم، به ارزیابی کاربردی و مفید بودن داده‌ها کمک می‌کند. با استفاده از آلت‌متریک‌ها و دستگاه‌های آماری، می‌توان الگوهای استفاده مجدد از داده‌ها و تعداد دانلودهای آن‌ها را شناسایی کرد. این داده‌های آماری می‌توانند به محققان و ارزیابان کمک کنند تا ببینند که داده‌ها چگونه مورد استفاده قرار گرفته و چه تأثیری بر پژوهش‌های دیگر گذاشته است. این امر به‌ویژه در تعیین ارزش و قابلیت استفاده از داده‌ها برای پژوهش‌های آینده مفید است. فرآیند داوری نیز نقش مهمی در اعتبارسنجی داده‌ها و افزایش رویت‌پذیری آن‌ها دارد. پیشنهاد شده است که داده‌های پژوهشی، مشابه مقالات علمی، تحت کنترل کیفیت و داوری مستقل قرار گیرند. این روند می‌تواند به تضمین دقت و کیفیت داده‌ها کمک کند و باعث افزایش اعتماد به داده‌های منتشر شده شود. داوری داده‌ها می‌تواند انگیزه‌ای برای نویسندگان ایجاد کند تا با دقت بیشتری داده‌های خود را مدیریت کرده و به بررسی‌های هم‌تا و فرآیندهای ویرایشی توجه بیشتری داشته باشند (کاستلو و همکاران، ۲۰۱۳).

همانگی با اصول FAIR در ارزیابی و گزینش داده‌های پژوهشی برای مدیریت و ذخیره‌سازی به چهار ویژگی کلیدی توجه دارد: قابلیت یافتن، قابلیت دسترسی، قابلیت تعامل و همکاری، و قابلیت استفاده مجدد. این اصول تضمین می‌کنند که داده‌ها به‌راحتی قابل‌شناسایی و دسترسی باشند، با دیگر سیستم‌ها و نرم‌افزارها یکپارچه شوند، و به‌طور مؤثر برای استفاده مجدد مستندسازی شوند (فوستر و همکاران، ۲۰۲۲). با رعایت این ویژگی‌ها، داده‌ها نه تنها در فرآیند تحقیقاتی آسان‌تر قابل استفاده خواهند بود، بلکه ارزش و کارایی آن‌ها به‌طور قابل توجهی افزایش می‌یابد.

خط‌مشی و مسائل مرتبط با مخازن داده‌ها تأثیر زیادی بر فرآیند ارزیابی و گزینش داده‌های پژوهشی دارند. نیازهای کاربران یکی از عوامل کلیدی در حفظ داده‌هاست، اما مخازن با محدودیت‌هایی در نگهداری و ذخیره‌سازی داده‌ها مواجه‌اند و همه اسناد قابل ذخیره‌سازی نیستند. خط‌مشی‌های هر مخزن مستقیماً بر راهبردهای ارزیابی و گزینش داده‌ها تأثیر می‌گذارد و

هزینه‌های نگهداری داده‌ها با توجه به پیچیدگی و حجم اطلاعات متغیر است (نیو، ۲۰۱۴). علاوه بر این، ارزش و کاربرد داده‌ها نیز متفاوت است؛ برخی داده‌ها به دلیل ویژگی‌های منحصر به فردشان اهمیت بیشتری دارند و انتظار می‌رود که بیشتر مورد استفاده قرار گیرند. بنابراین، داده‌های گزینش شده باید از قابلیت استفاده پژوهشی بالایی برخوردار باشند تا به‌طور مؤثر در تحقیقات آینده مورد بهره‌برداری قرار گیرند.

پژوهش حاضر باهدف بررسی معیارهای ارزیابی و گزینش داده‌های پژوهشی برای ذخیره‌سازی و نگهداری در مخازن داده، به شناسایی و تحلیل عوامل کلیدی مؤثر بر این فرآیند پرداخته است. نتایج نشان داد که ارزیابی داده‌های پژوهشی نیازمند توجه دقیق به مؤلفه‌های متعددی نظیر آماده‌سازی داده‌ها، کیفیت و شرایط فیزیکی داده‌ها، مستندسازی فراداده‌ها، اصول اخلاقی و سیاست‌های مخازن داده است. این پژوهش تأکید می‌کند که کیفیت داده‌ها، از جمله دقت، صحت، کامل بودن و به‌روز بودن، نقش بسزایی در تسهیل استفاده مجدد و ارزش‌گذاری داده‌ها دارد. همچنین، مستندسازی دقیق فراداده‌ها و رعایت اصول اخلاقی، مانند حفظ حریم خصوصی شرکت‌کنندگان و شفافیت در مالکیت داده‌ها، به اعتمادسازی و افزایش قابلیت استفاده داده‌ها کمک می‌کند. به‌علاوه، هماهنگی با اصول FAIR و استانداردهای فنی، تضمین می‌کند که داده‌ها قابلیت یافتن، دسترسی، تعامل و استفاده مجدد را داشته باشند. درنهایت، یافته‌های این پژوهش بر اهمیت سیاست‌گذاری مؤثر مخازن داده برای ذخیره‌سازی و مدیریت داده‌های پژوهشی تأکید دارد. این سیاست‌ها باید نیازهای کاربران، محدودیت‌های فنی و هزینه‌های مرتبط با نگهداری داده‌ها را در نظر گرفته و به نحوی طراحی شوند که داده‌های باارزش پژوهشی بالا را برای استفاده در آینده حفظ کنند. از این‌رو، استفاده از معیارهای جامع و دقیق در ارزیابی و گزینش داده‌های پژوهشی می‌تواند به افزایش کیفیت داده‌ها و تقویت زیرساخت‌های علمی و پژوهشی کمک کند.

در پایان پیشنهادهای کاربردی و اجرایی پژوهش حاضر به شرح زیر است:

۱. توسعه دستورالعمل‌های ارزیابی داده‌ها: ایجاد چک‌لیست‌ها و دستورالعمل‌های استاندارد برای ارزیابی کیفیت داده‌ها که توسط سازمان‌ها و مراکز پژوهشی قابل استفاده باشد.
۲. آموزش متخصصان داده: برگزاری دوره‌های آموزشی برای پژوهشگران و متخصصان به‌منظور آشنایی با معیارهای ارزیابی داده‌ها و شیوه‌های بهینه‌سازی کیفیت آن‌ها.
۳. پیاده‌سازی ابزارهای فناوری: طراحی و استفاده از ابزارهای نرم‌افزاری برای خودکارسازی فرآیند ارزیابی و گزینش داده‌ها.
۴. ایجاد پایگاه‌های داده استاندارد: ایجاد مخازن داده‌های پژوهشی که معیارهای ارزیابی و کیفیت داده‌ها را به‌صورت شفاف در فرآیند ذخیره‌سازی لحاظ کنند.

منابع

- مجیدی، اکبر؛ نقشینه، نادر؛ اسمعیلی‌گیوی، محمدرضا؛ و هاشمی، محمود رضا (۱۳۹۶). مطالعه مبانی، مدل‌ها و مسائل گزینش و مدیریت داده‌های پژوهشی در محیط‌های علمی و دانشگاهی. *تعامل انسان و اطلاعات*، ۴(۲)، ۳۱-۵۷.
- وزیری، اسماعیل؛ نقشینه، نادر؛ و نوروزی‌چاکلی، عبدالرضا (۱۳۹۸). موانع و چالش‌های اشتراک‌گذاری داده‌های پژوهشی. *پژوهشنامه کتابداری و اطلاع‌رسانی*، ۹(۲)، ۲۳-۵. <http://doi.org/10.22067/riis.v0i0.60594>

References

- Alvarez-Romero, C., Martínez-García, A., Bernabeu-Wittel, M., & Parra-Calderón, C. L. (2023). Health data hubs: an analysis of existing data governance features for research. *Health Research Policy and Systems*, 21(1), 70.
- Assante, M., Candela, L., Castelli, D., & Tani, A. (2016). Are scientific data repositories coping with research data publishing?. *Data Science Journal*, 15, 6-6.
- Batini, C & Scannapieco, M. (2016). Data quality dimensions. In: Batini, C and Scannapieco, M (eds.), *Data and information quality: Dimensions, principles and techniques*. Berlin: Springer. pp. 21–51. DOI: https://doi.org/10.1007/978-3-319-24106-7_2
- Boté-Vericad, J. J., & Térmens i Graells, M. (2019). Reusing data: Technical and ethical challenges. *DESIDOC Journal of Library & Information Technology*, 2019, vol. 39, num. 6, p. 329-337.
- Bradshaw, A., Hughes, N., Vallez-Garcia, D., Chokoshvili, D., Owens, A., Hansen, C., ... & Diaz, C. (2023). Data sharing in neurodegenerative disease research: challenges and learnings from the innovative medicines initiative public-private partnership model. *Frontiers in neurology*, 14, 1187095.
- Busetto, L., Wick, W., & Gumbinger, C. (2020). How to use and assess qualitative research methods. *Neurological Research and practice*, 2(1), 14.
- Cai, L., & Zhu, Y. (2015). The challenges of data quality and data quality assessment in the big data era. *Data science journal*, 14, 2-2.
- Corti, L., & Backhouse, G. (2005, May). Acquiring qualitative data for secondary analysis. In *Forum Qualitative Sozialforschung*. 6(2). FQS.
- Corti, L., Woollard, M., Bishop, L., & Van den Eynden, V. (2019). Managing and sharing research data: A guide to good practice.
- Costello, M. J., Michener, W. K., Gahegan, M., Zhang, Z. Q., & Bourne, P. E. (2013). Biodiversity data should be published, cited, and peer reviewed. *Trends in Ecology & Evolution*, 28(8), 454-461.
- Costello, M. J., Michener, W. K., Gahegan, M., Zhang, Z. Q., & Bourne, P. E. (2013). Biodiversity data should be published, cited, and peer reviewed. *Trends in Ecology & Evolution*, 28(8), 454-461.
- Downs, R. R. (2021). Improving opportunities for new value of open data: Assessing and certifying research data repositories. *Data Science Journal*, 20, 1-1.
- El Mestari, S. Z., Doğan, F. S., & Maria Botes, W. (2023, April). Technical and Legal Aspects Relating to the (Re) Use of Health Data When Repurposing Machine Learning Models in the EU. In *Privacy Symposium: Data Protection Law International Convergence and Compliance with Innovative Technologies* (pp. 33-48). Cham: Springer International Publishing.
- Fecher, B., Friesike, S., & Hebing, M. (2015). What drives academic data sharing?. *PloS one*, 10(2), e0118053.
- Filip, I. D., Ionite, C., González-Cebrián, A., Balanescu, M., Dobre, C., Chis, A. E., ... & González-Vélez, H. (2022, December). SMARDY: Zero-trust FAIR marketplace for research data. In *2022 IEEE International Conference on Big Data (Big Data)* (pp. 1535-1541). IEEE.
- Grady, C. (2015). Enduring and emerging challenges of informed consent. *New England Journal of Medicine*, 372(9), 855-862.
- Gregory, K. M., Cousijn, H., Groth, P., Scharnhorst, A., & Wyatt, S. (2020). Understanding data search as a socio-technical practice. *Journal of Information Science*, 46(4), 459-475.
- Harper, L. M. (2023). *Data Reuse Among Digital Humanities Scholars: a Qualitative Study of Practices, Challenges and Opportunities* (Doctoral dissertation, Université d'Ottawa/University of Ottawa).
- Harper, L. M. (2023). *Data Reuse Among Digital Humanities Scholars: a Qualitative Study of Practices, Challenges and Opportunities* (Doctoral dissertation, Université d'Ottawa/University of Ottawa).

- Harris, Howard (2001). Content analysis of secondary data: A study of courage in managerial decision making. *Journal of Business Ethics*, 34, 191-208.
- Kindling, M., & Strecker, D. (2022). Data quality assurance at research data repositories. *Data Science Journal*, 21, 18-18.
- Koltay, T. (2020). Quality of open research data: Values, convergences and vernance. *Information*, 11(4), 175.
- Kumuthini, J., Zass, L., Chaouch, M., Gill, Z., Ras, V., Mungloo-Dilmohamud, Z., ... & Baichoo, S. (2023). Data standardization in the omics field. In *Genomic Data Sharing* (pp. 137-155). Academic Press.
- Kumuthini, J., Zass, L., Chaouch, M., Thompson, M., Olowoyo, P., Mbiyavanga, M., ... & Owolabi, M. (2019). Proposed guideline for minimum information Stroke research and clinical data reporting. *Data Science Journal*, 18, 26-26.
- Majidi, A., Naghshineh, N., Ismaili-ghivi, M.R., Hashemi, M. (2017). Study of the foundation, models and issues of research data curation and management in scientific and academic environments. *Human Information Interaction*, 4(2), 31-57. [In Persian]
- Mannheimer, S. (2022). Data curation for qualitative data reuse and big social research: Connecting communities of practice. Humboldt Universitaet zu Berlin (Germany). Book
- Manu, E., Akotia, J., Sarhan, S., & Mahamadu, A. M. (2021). Identifying and sourcing data for secondary research. In *Secondary research methods in the built environment* (pp. 16-25). Routledge.
- Mckenna-Foster, A., Cotera, M., & Hahnel, M. (2022). Open Science ETDs and Institutional Repositories: Making Research Data FAIRer. *The Journal of Electronic Theses and Dissertations*, 2(1), 5.
- Niu, J. (2014). Appraisal and selection for digital curation. *International Journal of Digital Curation*, 9(2), 65-82.
- Omukuti, J., Megaw, A., Barlow, M., Altink, H., & White, P. (2021). The value of secondary use of data generated by non-governmental organisations for disaster risk management research: Evidence from the Caribbean. *International journal of disaster risk reduction*, 56, 102114.
- Oza, V. H., Whitlock, J. H., Wilk, E. J., Uno-Antonison, A., Wilk, B., Gajapathy, M., ... & Lasseigne, B. N. (2023). Ten simple rules for using public biological data for your research. *PLOS Computational Biology*, 19(1), e1010749.
- Palmer, C. L., Weber, N. M., & Cragin, M. H. (2011). The analytic potential of scientific data: Understanding re-use value. *Proceedings of the American Society for Information Science and Technology*, 48(1), 1-10.
- Panchenko, L., & Samovilova, N. (2020). Secondary data analysis in educational research: opportunities for PhD students. In *Shs web of conferences* (Vol. 75, p. 04005). EDP Sciences.
- Peng, G., Lacagnina, C., Downs, R. R., Ganske, A., Ramapriyan, H. K., Ivánová, I., ... & Moroni, D. F. (2022). Global Community Guidelines for Documenting, Sharing, and Reusing Quality Information of Individual Digital Datasets. *Data Science Journal*, 21:008, 20pp.. <http://doi.org/10.5334/dsj-2022-008>
- Rabianski, J. S. (2003). Primary and secondary data: Concepts, concerns, errors, and issues. *The Appraisal Journal*, 71(1), 43.
- Rantasaari, J. (2021). Doctoral Students' Educational Needs in Research Data Management: Perceived Importance and Current Competencies. *International Journal of Digital Curation*, 16(1), 36-36.
- Rantasaari, J. (2022). Doctoral Students' Research Data Management Competencies Based on the Quality of Their Data Management Plans. *Proceedings of the IATUL Conferences*. Paper 4. <https://docs.lib.purdue.edu/iatul/2022/clr/4>

- Rehnert, M., & Takors, R. (2023). FAIR research data management as community approach in bioengineering. *Engineering in Life Sciences*, 23(1).
- Rolland, B., & Lee, C. P. (2013, February). Beyond trust and reliability: reusing data in collaborative cancer epidemiology research. In *Proceedings of the 2013 conference on Computer supported cooperative work* (pp. 435-444).
- Sakai, Y., Miyata, Y., Yokoi, K., Wang, Y., & Kurata, K. (2023). Initial insight into three modes of data sharing: Prevalence of primary reuse, data integration and dataset release in research articles. *Learned Publishing*, 36(3), 417-425.
- Sandelowski, M., & Barroso, J. (2007). *Handbook for synthesizing qualitative research*. New York: NY: Springer.
- Sherif, V. (2018, March). Evaluating preexisting qualitative research data for secondary analysis. In *Forum qualitative sozialforschung/forum: Qualitative social research* (Vol. 19, No. 2).
- Stratton, S. J. (2015). Assessing the accuracy of survey research. *Prehospital and disaster medicine*, 30(3), 225-226.
- Stvilia, B., & Lee, D. J. (2024). Data quality assurance in research data repositories: a theory-guided exploration and model. *Journal of Documentation*, 80(4), 793-812. <https://doi.org/10.1108/JD-09-2023-0177>
- van der Velde, K. J., Singh, G., Kaliyaperumal, R., Liao, X., de Ridder, S., Rebers, S., ... & Swertz, M. A. (2022). FAIR Genomes metadata schema promoting Next Generation Sequencing data reuse in Dutch healthcare and research. *Scientific data*, 9(1), 169.
- Vuokko, R., Mäkelä-Bengs, P., Hyppönen, H., Lindqvist, M., & Doupi, P. (2017). Impacts of structuring the electronic health record: Results of a systematic literature review from the perspective of secondary use of patient data. *International journal of medical informatics*, 97, 293-303.
- Whiteside, Mary; Mills, Jane & McCalman, Janya (2012). Using secondary data for grounded theory analysis. *Australian Social Work*, 65(4), 504-516.
- Wilkinson, M. D., et al. (2016). The FAIR Guiding Principle for Scientific Data Management and Stewardship. *Scientific Data*, 3(1): 160018. DOI: <https://doi.org/10.1038/sdata.2016.18>
- Woolard, Robert H.; Carty, Kathleen; Wirtz, Philip; Longabaugh, Richard; Nirenberg, Ted D.; Minugh, Allison P.; Becker, Bruce & Clifford, Patrick R. (2004). Research fundamentals: Follow-up with subjects in clinical trials: Addressing subject attrition. *Academic Emergency Medicine*, 11(8), 859-866.
- Yoon, A. (2017). Data reusers' trust development. *Journal of the Association for Information Science and Technology*, 68(4), 946-956.
- Zuiderwijk, A., Türk, B. O., & Brazier, F. (2024). Identifying the most important facilitators of open research data sharing and reuse in Epidemiology: A mixed-methods study. *Plos one*, 19(2), e0297969.
- Vaziri, E., Naghshineh, N., & Noroozi Chakoli, A. (2020). Barriers and Challenges of Research Data Sharing. *Library and Information Science Research*, 9(2), 5-23. <https://doi.org/10.22067/riis.v0i0.60594> [In Persian]