Macrothink Institute™

Journal of Social Science Studies
ISSN 2329-9150
2024, Vol. 11, No. 1

The Construction of University Scientific Research Supervision Platform under the View of Data Middle Platform

Yufan Lin (Corresponding author)

Information Centre, Zhejiang Open University

Hangzhou, China

Tel: 0086-17794561560 E-mail: LYF19930811@126.com

Received: March 1, 2024 Accepted: April 11, 2024 Published: May 6, 2024

doi: 10.5296/jsss.v11i1.21883 URL: https://doi.org/10.5296/jsss.v11i1.21883

Abstract

The system platform of colleges and universities has been developed from the informatization perspective based on data sharing via the reshaping of new content and the creation of technological architecture. The creation of the University Scientific Research Supervision Platform based on the Data Middle Platform attempts to analyze and address the issues above from the data sharing and service level perspective. These issues include Complicated scientific research fund reimbursement, Historical data hoarding, Late supervision & early warning, Slow system data update, Risk of data leakage, and Inconsistent inter-departmental data. The platform's technological architecture has transformed from a straightforward stacking of business processes to one that allows for the sharing and interchange system data. The solutions include:

●Data sharing policies.

●Managing data dictionaries.

●Organizing historical data.

●Integrating platforms.

●Optimizing analytical functions.

●Ensuring data security, among others.

Keywords: data sharing, data middle platform, integrating platforms, optimizing analytical

functions

1. Background

InInterdisciplinary cross-research supported by informatization has increased since the Ministry of Education's 'Education Informatization 2.0 Action Plan' was introduced. Deep use of big data and artificial intelligence technology in university scientific research will also become the norm, and the organization and management of university scientific research will experience significant changes. Universities' informationization also considers the inter-departmental linkages and the implementation of regular scheduling and oversight systems to ensure inter-departmental routine and long-term collaboration. Business management fully utilizes big data and artificial intelligence for decision support, from data collection to the data center and from electronic service to intelligent assistance. The 20th National Congress report suggests enhancing the system for innovation in science & technology, promoting the Internet, big data, and artificial intelligence, speeding up the creation of an innovative nation, a strong nation in science & technology, a strong nation in networks, a digital China, as well as an intelligent society. These recommendations offer favorable policy support for the thoughtful growth of the university system. University research is advancing quickly toward management system reform, building innovative campus systems, and improving research information management, data integration, and exchange of various business systems due to favorable external policies and settings. The foundation for data center construction in universities domestically and abroad has been strengthened by introducing data governance laws, digital service laws, data governance manuals, other data-related support policies, and management methods that keep up with the times in the U.S. and EU countries (Cheng, 2022). Existing platforms and methods used by universities can be concluded from theoretical and practical perspectives. Universities often use Online Research Management Systems software platforms to manage research projects, grants, and publications. These systems allow for tracking research progress, budgeting, and reporting. They also facilitate communication between researchers, administrators, and funding agencies. Senior researchers at Nankai University often supervise junior researchers through mentorship programs or advisory committees. These committees provide guidance, feedback, and support to researchers, helping them navigate the research process and develop their skills(Peuter, 2024_). The postgraduate educational and instructional supervision system is an integral part of the quality assurance system of postgraduate education based on the theory of total quality management (TQM)(Ouyang,2023). Grimmer argues that for automated text methods to become a standard tool for scientists, methodologists must contribute new techniques and methods of validation.

Data is a crucial strategic resource on par with human and natural resources. In the context of big data in the data center industry, we can only better adapt to the iterative updates of different system platforms by mastering the large-scale and flexible use of data (Chen, 2013). The data middle platform, which unifies the above services, was created and is increasingly utilized in the market based on data centers, data governance, and extensive data analysis. For universities, the data center needs to be tailored to the actual application of teaching and learning scenarios, and a lot of data research work is required in the early stage, including

data system research, data table research, business process analysis, data demand confirmation, and manual data input among other things.

2. Case Study Analysis

For instance, based on the integration of technological basis, Zhejiang University, which has achieved significant advancements in information system construction and digital intelligence services in recent years, may develop an adapted customized data service platform, data middle platform, and business platform for other colleges and businesses. The Open University (OU), which has feasibility and quantified improvements in efficiency and data management, manages data such as teacher and student feedback, course statistics, and behavioral trajectories using artificial intelligence, big data, and machine learning technology. As a result, OU is better equipped to serve staff and student needs & concerns and provide better data support. At the same time, it uses the most up-to-date data analysis tools to enhance its educational policies and processes, setting an example for other institutions. Sustainability Research shows us the research project led by Professor Minghui Qian from the School of Information Resources at Renmin University of China focuses on the exploration and practice of public safety data governance with the involvement of public users. The project is based on the "Jinhe Network Wisdom Supervision Platform," aiming to improve the participation and management of public safety data. The project emphasizes the importance of a systematic approach to data governance, suggesting that the platform can significantly enhance how research data is managed and utilized for supervision purposes.

These case studies illustrate the practical application and benefits of implementing a platform for research supervision in universities' feasibility and quantifying improvements in efficiency and data management. They demonstrate the potential for improved efficiency, better data governance, and the facilitation of research supervision activities. While these examples are not directly from university research supervision, the principles and technologies applied can be adapted to create a robust and effective research supervision platform in a university setting.

3. Problems

Due to flaws in the system, technology, and system, the original university study suffers from the following general and particular issues: 1. Complicated scientific research fund reimbursement 2. Historical data hoarding 3. Late supervision & early warning 4. Slow system data update 5. Risk of data leakage 6. Inconsistent inter-departmental data.

Currently, the following challenges are wreaking havoc on university information building: (a) There is a visible pattern of repeated construction and a need for top-level planning and design. The main faults are as follows: first, there is apparent duplication of construction and a lack of top-level planning and design.t. Because no campus-level business data model exists, the data can only serve its business system and cannot manage big data, data analysis, and conversion applications holistically. (c) The user experience must be improved, and separate service entrances for teachers and students must be provided. (d) Gaps in the business system and unfinished work on the security system.

3.1 Complicated Scientific Research Fund Reimbursement

The first procedure for reimbursing research money entails filling out forms for appointment reimbursement and daily reimbursement approval (as well as uploading attachments for vouchers that should contain invoices and payment documentation). The initial procedure calls for the project manager to manually enter duplicate data into the university's OA and financial systems for registration. The Research Office (which issues funding, the Financial Office, which handles reimbursement, and other project-related departments) are all involved in the approval process. As a result, the standard procedure requires at least 3-5 working days to finish and is likely to be rejected. After being rejected, it must be submitted again for another approval. Many scholars and researchers are often in financial straits and need timely financial support for their research projects.

3.2 Historical Data Hoarding

Many academics and researchers struggle financially and frequently need more timely financial assistance for their extensive research undertakings. The primary cause for the discrepancy between the current state of the data centers and data governance and the ideal state is the extensive quantity of inefficient hoarding of historical data (Wang, 2021). The financial data primarily consists of the 12 tables that the economic system requires (such as the labor cost information table, the labor cost detail information table, the consulting cost information table, the reception cost information table, the asset purchase reimbursement information table, the conference cost information table, the training cost information table, the foreign cost information table, the travel cost information table, and the project expenditure information table). Large and burdensome data volumes are the primary characteristics of historical data, and many essential fields are not recorded promptly because of gaps involving handwritten paper records. The university research data in question are historical data templates (including results, funds, and project management). The data under each category show invalid piles and grow year after year. The data are saved in various formats, lacking effective and uniform management. What is more, it needs more judgment in processing invalid data.

3.3 Late Supervision & Early Warning

Universities need a closed-loop management system (due to the lack of a direct method and policy for monitoring and early warning of research management), which leads to sluggish monitoring and covert early warning. For instance, travel expenditures are not reported electronically, and the flow of conference funding needs to be clarified. The supervisory department needs to flag and investigate each node's loopholes, which results in the audit loopholes of the entire research project building up later and wasting the most incredible opportunity to fix the procedure. Contrarily, it is challenging to effectively transform the data gathered and processed in the early research into the early warning module (that can support the supervision department's decision-making) because the data model of the early warning supervision module is not clearly defined. The scope of access to the system and the later platform development is unclear.

3.4 Slow System Data Update

The university's original historical data is scattered and distributed online & offline, and there are several online systems involved. Additionally, the synchronization data mechanism between the systems could be better, which causes three questions: (a) Slow system data updates, (b) Slow real-time, and (c) Poor data quality. Another critical problem is the poor quality of the original data. A significant amount of scientific research data was not processed by the secondary cleaning and classification in the data middle (because of the excessive amount of data & high system load), which prevented the early stage of data governance from having a significant impact. Additionally, there needs to be a shared knowledge of the dictionary structure, meta-data management, and data standards of the source data (Jin, 2021), (as well as preparatory data research for the system's scientific research data). Due to the factors above, the collected system research data were disorganized & delayed, lacked unified middleware planning, and the annotations of the iteratively updated data versions were not displayed on each node, making it difficult for the data users & managers to manage the versions on their behalf.

3.5 Risk of Data Leakage

(Web link security flaws, Trojan horse infections, hacking, firewall flaws, etc.) are the primary data leakage security issues(Wei & Liu, 2017). The important data for the oversight and administration of academic research includes (money, bank account information, identification information, and other private information) that may be readily stolen and utilized. There are also hazards of data leaking via web link attacks and hacker penetration. Consider a web link attack as an illustration; untrustworthy organizations will frequently send spam emails to university teachers' and students' school mailboxes to persuade them to click on the links in the emails, allowing them to illegally obtain (IP addresses, web passwords, and other private information). They will then break into teachers' and students' computers and steal important data. In recent years, with the widespread usage of VPNs, college and university VPN accounts have been able to go over school firewall control and enter into databases directly, increasing the danger of data management. Another method of data leaking that requires attention is the sale of VPN accounts on dubious websites that contain university password information.

3.6 Inconsistent Inter-departmental Data

There needed to be more consistency in the data between departments due to the prior data management in universities not identifying the source departments and linked departments for each type of data. Different departments employ various data definitions and formats, which makes it challenging for decision-makers to collect correct & complete data and consequently impacts the standard of decision-making. Each university business department tends to run and keep its data, making it difficult for data managers to request and leaving them with limited options and a time-consuming sharing process. The university's business database architecture and access rights for research-related systems are under the system construction company's control, making it challenging for university system administrators to operate and maintain the research database thoroughly and more difficult to build later.

Different business systems' fundamental architectural architecture and data dictionaries are inconsistent, and the pertinent data may be stored across various systems, departments, and databases. The primary cause of conflicting data and the challenge of data operation & maintenance is the lower involvement of university technological departments in the original database design.

4. Recommendations

4.1 Data Sharing Policy

In recent years, institutional regulations governing university scientific research supervision have evolved incrementally in the direction of informationalization, improvement, cooperation, service, and data. The data sharing policy is one of the particular management policies universities have, covering standard specification, database development, result management, financing process oversight, multidimensional query, statistical analysis, etc. The policies, which may be categorized into three categories: (a) Data management policies, (b) Data sharing plans, and (c) Research data management plans, are primarily directed at the university's business divisions and mainly address the issues of inconsistent data across departments and sluggish system updates. The data management policy is the first policy that offers institutional backing for later data sharing and usage. The pertinent rules and action guidelines are provided, along with the appropriate data source and supervisory departments. The data-sharing plan addresses the process management of data, establishes plans and specifications for each distinct node of data transmission, offers comprehensive process advice for data cleaning& circulation, and aids in developing a data center for university research administration.

On the other hand, the scientific research data management strategy is a detailed plan for storing and categorizing the valid data cleaned and processed by the intermediate platform. It simplifies the later results management and helps the pertinent statistical analysis and the multidimensional query of the iconic data display. To accomplish data consistency and knowledge exchange, the process utilizes data integration tools & technologies that combine data from several departments & systems at the level of the intermediate platform.

4.2 Managing Data Dictionaries

To guarantee that the definition, usage, and maintenance techniques of data are uniform across all business divisions of universities, data dictionary management is carried out for the middle office to prepare the global data management strategy. The creation of foundational data (data standard management, data model management, data development, data management, data quality management, and data source management) was a component of the initial data governance for universities, which facilitated the initial standardization and adjustment of source data for academic research (Liu & Zhang, 2019). Based on this, the administration of the data dictionary for the data center of university scientific research supervision is expanded (which can successfully address the issues of complex scientific research fund reimbursement and sluggish system data update). The scientific research data are controlled and analyzed promptly following university requirements for the

standardization of scientific research business data, and an overall university scientific research supervision platform with a single table is established to standardize the accuracy, real-time, and sharing of scientific research data. According to the business, the planning of the associated data dictionary may be separated into horizontal and vertical projects. The early warning model's color-coded supervisory nodes are before, during, and after. The following is the infrastructure of the data dictionary in the middle platform that meets the above requirements:

select *from kyj_dict_item

id	dict	code	name
1	CZZT	0	insert
2	CZZT	1	update
3	CZZT	2	delete
4	XMLX	1	Horizontal
5	XMLX	2	Vertical
6	YJLX	1	Before the event
7	YJLX	2	In the Matter
8	YJLX	3	After the fact
9	YJJB	1	Red
10	YJJB	2	Yellow
11	YJJB	3	Blue
12	SFQKSS	1	Yes
13	SFQKSS	2	No
14	SFLA	1	Yes
15	SFLA	2	No
16	SFJLZD	1	Yes
17	SFJLZD	2	No
18	SFDJZWCF	1	Yes
19	SFDJZWCF	2	No

Make a thorough table design and subsequent data analysis based on the data dictionary's architecture. To improve the administration of the data dictionary even more, use custom code to obtain the lower-level data corresponding to the higher-level id or create a tree structure for the data.

4.3 Organizing Historical Data

Data migration, historical data collation, data governance, data confirmation, and incorporation into the data center (as well as data interaction & sharing) are the primary phases in the historical data cleaning process that successfully address historical data hoarding. The platform's data processing processes support the simultaneous exchange of monitoring, research, and personnel information by adhering to the entire staff, total volume, process, and elements. (a) The first workload preparation work includes data migration,

classification and screening for historical data, classification, standardization, and backup migration of pertinent data already present in each system. Reintegrating the current data by the platform supervision department's management requirements, which include field information adjustment, data dictionary table correspondence, and batch return for correcting any data that does not satisfy management requirements until it satisfies platform data requirements, is the second step in data governance. (b) After formal data migration into the data middle platform, the third phase of data validation is reviewing and validating the data. (c) The final data interaction uses the fundamental data of scientific research as the connecting factor, connecting university business departments to cooperate horizontally in the sharing of scientific research information and realizing the data center of data chain sharing to maximize the scientific research information data service, avoid resource waste, and also provide data assurance for the ensuing platform unified integration & analysis function optimization. It is crucial to perform effective data quality management and set up a data quality management system to routinely monitor & maintain data to guarantee the correctness& integrity of the data.

4.4 Integrating Platforms

The unified and integrated architecture of the University Scientific Research Supervision Platform is necessary to address the two main issues of insufficient supervision and early warning (as well as non-uniform data across departments). This platform fills and represents an advancement compared to the previous platforms and methods, mainly reflected in its Platform Unified Integration Function. It shows the link between the Data middle platform and other data integration systems.

The external must dock with the university's OA system's message center to achieve information pushing, intelligent reminders, and docking of outstanding matters. Unified Identity Authentication has strict data standard requirements for the platform's organizational structure data and necessitates the ability to enter into Numerous Business Systems using a single login. The data connectivity of each module across platforms must be improved, data silos must be eliminated, and real-time data connectivity between Financial System and Personnel System must be prioritized. Along with enhancing management functions for scientific research performance assessment & data reporting, it also enhances instructors' and students' access to and knowledge of the digital world. At the same time, it expands the functionality of the project, financing, and achievement modules, such as (The Management of the Project Setup Process, Budget Template Management, and Funding Management for Scientific Research, etc). The data center serves as the integration interface and includes services (like MySQL, Redis, the Center Web Service, Basic Service, Public Service, Data Governance, and Data Service), among others. The Middle Platform facilitates the building of the Platform Integration, synchronizes the supervisory needs, establishes Early Warning Rules, continually expands the List of Issue Requirements, and builds the High-Quality Early Warning Model Cluster. Establishing the Color Block Early Warning Model is necessary for the first building of the Early Warning Model to determine the unavoidable cause-effect relationship between Academic Research and Behavioral Issues. The names of the Early Warning Models, Early Warning Rules, Data Base Logical Relationships, Matching Disposal

Units, and other data components must be adapted to the specific circumstances at the school. Take A university in Zhejiang Province as an example, build the Platform Unified Integration Function Diagram as following Figure 1:

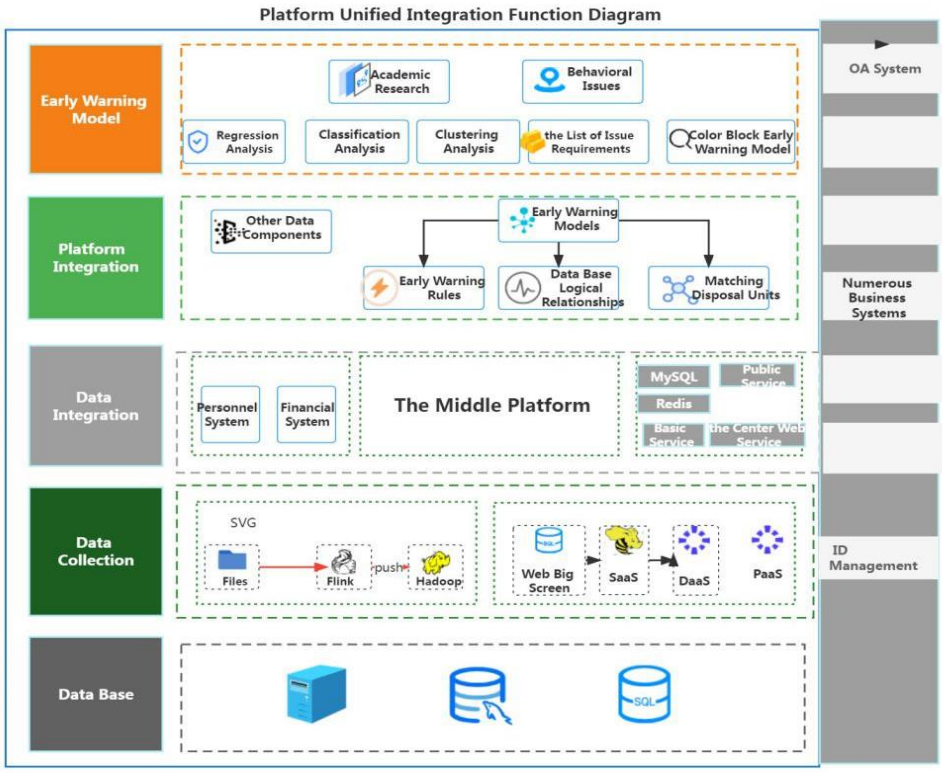

Figure 1. Platform Unified Integration Function Diagram

Universities can tailor the University Scientific Research Supervision Platform to their specific needs by implementing customization strategies that align with their institutional goals, academic environments, and the unique characteristics of their student body. Adapting early warning models and data components to local academic environments is also prominent. The first step in tailoring the platform is profoundly understanding the local educational environment. This includes the university's mission, academic programs, research focus, and the demographic profile of students and faculty. By understanding these factors, the university can identify its unique needs and challenges and customize the platform accordingly. Early warning models are designed to predict student performance and identify those at risk of falling behind. Universities can customize these models by Incorporating institution-specific data points, Adjusting the model's algorithms to reflect the university's grading scale, and Integrating qualitative data. Adapting data components and Integrating institutional knowledge and expertise also enable the platform to be a valuable asset in promoting academic success and advancing research excellence.

4.5 Optimizing Analytical Functions

Business data are all routed to the Data Middle Platform based on the logical framework of platform unification and integration, and the platform application problem of late supervision

& warning is resolved from the standpoint of analysis function optimization. The optimized analysis function design ideas place a strong emphasis on Cross-level collaboration, Process improvement, Platform research data push, Financial data push, Data inter-modulation, Data modeling, and Early warning based on fundamental research data, Fundamental financial data, and data under the conditions of meeting the Early warning model. The technical foundation of (Web Big Screen, SaaS, DaaS, and PaaS) are some of those that are included. The web layer consists of PC and mobile terminals; the SaaS layer displays the well-known big data screen and data cockpit; and the DaaS (data service) layer corresponds to the data open interface, data warehouse, and data middle platform. The PaaS (platform service) layer is the most visible layer to the outside world, consisting of the big data computing platform, the data collection and aggregation platform, the manual filling system, etc. The next stage of huge data modeling and analysis is derived when the technological foundation of the four layers above is appropriately constructed. At that point, the data is sent to the data center in a more regular format. Regression analysis, Classification analysis, and Clustering analysis are some of the more popular modeling techniques that can be chosen during the modeling & analysis process. These techniques can further support scientific decision-making with data. Data model development, quality assurance, meta-data management, data index creation, and data life cycle management are among the Modeling and analytic tasks.

Utilizing a Data Middle Platform within the University Scientific Research Supervision Platform can yield many expected outcomes and measurable benefits for various stakeholders within the university ecosystem, including researchers, administrators, and funding bodies. Researchers can make more informed decisions about their research directions and methodologies by leveraging the data analytics capabilities of the Data Middle Platform. This can lead to more targeted and effective research efforts. Administrators can use the platform to monitor research progress and outcomes, enabling them to allocate resources more effectively based on data-driven insights. This ensures funds and facilities are directed towards the most promising research initiatives. The platform provides funding bodies with a clear view of how their funds are used and the outcomes achieved, fostering trust and accountability in the research process.

4.6 Ensuring Data Security

Scientific research data are transmitted daily through the extensive data link system chosen by the superior department to mitigate the risk of data leakage & provide data security, and ideal data encryption mechanisms like asymmetric encryption are offered. The platform's management logic adheres to the data center's hierarchical decentralization, granting managers and users with various levels of authority access to the platform's operating interface and data. Users must submit online data access requests to upper-level administrators for data access requests that are outside the area of their power. The platform will then give temporary data access authorization following the administrators' evaluation and confirmation of the user's authority. The platform's technological design can link to most external systems, is highly expandable and flexible, and offers a uniform security interface to reduce security concerns brought on by data transmission efficiently.

5. Contributions

The study's original contributions to the university scientific research supervision field, as well as the possible benefits of applying the suggested platform.

6. Conclusion

The development of various application-based system platforms has benefited from a fresh research perspective offered by the Data Middle Platform that emerged in the significant data era. The technical reconfiguration of the platform has changed from the stacking of early business processes to the exchange & circulation of system data. The development of university research supervision platforms is primarily one of them, and it is one of the practical situations in universities explored in this article. Additionally, data-sharing systems like Financial and Information public service platforms will take up a sizable portion of the industry in the next era of data services. While waiting for further investigation, it is essential to consider how to transition from theoretical research to system technical architecture actively and how to configure suitable data governance & planning solutions for each relevant element and dimension in the construction of the University data middle platform.

Acknowledgments

We greatly appreciate the valuable contributions of our community advisory committee members. We would also like to thank the Zhejiang Open University and every team member who took the time to participate in this study.

Authors contributions

All authors read and approved the final manuscript.

Funding

This work was supported by Zhejiang Open University General Project “Research on the Construction and Operation Mechanism of Inclusive Education Public Service System from the Perspective of Human Capital”[project number XKT2023Y17]; 312 Talent Training Project of Zhejiang Open University.

Competing interests

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

Informed consent

Obtained.

Ethics approval

The Publication Ethics Committee of the Macrothink Institute.

The journal’s policies adhere to the Core Practices established by the Committee on Publication Ethics (COPE).

Provenance and peer review

Not commissioned; externally double-blind peer reviewed.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Data sharing statement

No additional data are available.

References

Chen, D. (2013). A Survey of Data Management and Sharing Policies in UK Research Funding Agencies and Insights. *Library and Information Work, 57*(08), 5-11. (In Chinese) https://doi.cnki:sun:tsqb.0.2013-08-000

Cheng, Y. (2022). Facing the Future: Analysis of Elements and Empirical Insights of the EU Data Governance Framework. *Digital Library Forum, 12,* 47-53. (In Chinese) https://doi.10.3772/j.issn.1673-2286.2022.12.006

Grimmer, J., Brandon, M. (2024). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis, 21*(3), 267-97. https://doi.org/10.1093/pan/mps028

Jin, Y. (2021). Current Situation and Outlook of Research Management Information System Development in Open Universities. *Inner Mongolia Electric University Journal, 3,* 109-112. (In Chinese) https://doi.10.16162/j.issn.1672-3473.2021.03.028

Liu, L., & Liu, W. (2019). A Study on Research Data Management and Sharing Policies in the UK. *Intelligence Data Work, 40*(5), 46-53. (In Chinese) https://doi.10.12154/j.qbzlgz.2019.05.009

Ouyang, X., & Chen, H. (2023). Necessity and Reality: Research on the Construction of Management System for Postgraduate Educational and Instructional Supervision in Universities— Based on the Analysis of the Supervision Texts in 70 Chinese Universities. *Journal of Educational Research and Policies*, *5*(6). https://doi.org/10.53469/jerp.2023.05(06).11

Peuter, D., & Dierickx, K. (2024). Mismatch in Perceptions of the Quality of Supervision and Research Data Management as an Area of Concern: Results from a University-wide Survey of the Research Integrity Culture at a Belgian University. *Accountability in Research*, 31-32. https://doi.10.1080/08989621.2024.2318245

Wang, L. (2021). The Construction of an Integrated Information Platform for University Research and Finance under the "Management and Service". *Administrative Assets and Finance, 15,* 16-18. https://doi.10.1109/IPEC51340.2021.9421117

Wei, Y., & Liu, G. (2017). Analysis of Research Data Management and Sharing Policy in Foreign Universities Based on Data Lifecycle. *Journal of Intelligence, 60.* https://doi.10.1241/johokanri.60.119