Establishing and Enhancing the Alignment of Research Data Management Practices at Kyrgyz and Austrian Universities

Abstract

With research in virtually all disciplines becoming increasingly data-driven, universities are establishing research data management (RDM) policies and developing/providing services to their researchers to increase the Findability, Accessibility, Interoperability and Reusability of the data. While requirements are highly similar, developing these in isolation may lead to incompatible systems and soes not allow to benefit from sharing the effort required to establish these services.

This project aims at bringing these initiatives together, understanding the requirements and leveraging the best practices and services established at the partner universities, to create a flexible and scalable RDM framework that can be adapted to the individual needs of the universities while benefiting from joint efforts invested and enhancing interoperability. Activities will include workshops and pilot service development/roll-out to share information on the existing infrastructure ranging from RDM policies to system set-up and service operations, analysing differences and gaps, and providing a basis for informed decision making on the path forward at each institution. The project will constitute an important step towards a comprehensive RDM framework, improved data management practices, enhanced researcher capabilities, and stronger international collaborations, ultimately driving research quality and competitiveness in the cooperation between universities in Europe and Kyrgyz.

Keywords: Research Data Management (RDM), Confidentiality, Trust in Research Data and Processes, Data Processing Pipelines, RDM Policy, Data Repositories

1 Background and Problem Statement

Research data management (RDM) has become a critical aspect of academic research, impacting data integrity, accessibility, and the overall quality of research outcomes. However, universities face several challenges in managing research data effectively, particularly in addressing issues related to:

1. **Data Location/Storage**: Decentralized and fragmented data storage systems spread across individual labs often make it difficult for researchers to locate, access, and manage datasets efficiently.

2. **Data Ownership**: Ambiguity around who owns the data—whether it is the institution, the researcher, or external stakeholders—leads to complications in accountability, accessibility, and long-term preservation.

3. **Data Consistency**: Maintaining consistency across various datasets is a challenge, especially when dealing with multiple sources and research projects accumulated over prolonged periods of time.

4. **Big Data Management**: With the growing scale of research data, universities often struggle to store, process, and analyse large datasets in a cost-effective and scalable manner.

5. **Data Confidentiality and Sharing**: Balancing the need for open data sharing while ensuring data confidentiality and compliance with ethical standards and regulations is a pressing issue.

6. **Data Analysis and Use of AI**: Many researchers lack access to advanced tools and skills for data analysis and AI-driven insights, limiting the potential of data-driven research.

7. **Data Visualization**: The effective presentation of complex data in a comprehensible and visually appealing format remains a challenge, hindering research communication and decision-making.

The National Development Strategy of the Kyrgyz Republic for 2018–2040 provides for the development of a digital society. The strategy sets the goal of creating an open digital society. To achieve this goal, a number of activities are envisaged, including:

- Provision of digital public services covering digital government in all regions of the country;
- Implementation of digital social services in healthcare and education;
- Creation of a national electronic infrastructure;
- Development of digital content in the online space.

Based on this, one of the strategic goals is a complete digital reform of public administration, which should significantly simplify, and sometimes completely eliminate bureaucracy, reduce the costs of performing government functions. Digital data generated in all areas, and especially in education and healthcare, should become a tool for measuring the effectiveness of budget expenditures (taxpayers' payments) and assessing the real situation in these areas. To this end, government institutions must completely restructure internal processes to formulate public policy based on data and monitor the execution of tasks through digital tools.

In this regard, it is necessary to solve some specific problems that contribute to the development of information resources in the educational and scientific sphere, allow the formation and management of the quality of educational processes, contribute to the formation and development of open educational resources for the system of higher education institutions, create an electronic platform for students and faculty, and form a modern information environment based on new digital technologies. Thus, there is a need to create a digital platform for the research process aimed at making it as efficient as possible in terms of creating, storing, reusing, organizing, and distributing scientific and research data. Research data management is associated with the organization of data from the very beginning of the research cycle to archiving the results obtained and publishing them.

The universities partnering in this project have already taken steps to address these issues:

TU Wien has established an RDM policy in 2018, followed by the establishment of a dedicated Center for Research Data Management (ZFDM). While the policy provides a clear guideline for the goals, roles and responsibilities of the various stakeholders, the ZFDM coordinates the stakeholders and helps with establishing and operating the services. Open source solutions were adapted to establish repository solutions for papers (DSpace), software (GitLab), and file-based research data (Invenio). Research projects were launched to develop novel repository solutions for structured data in relational databases (DBRepo) and for automating data management planning processes (DAMAP).

Kyrgyz State Technical University is actively developing the infrastructure for research data management (RDM) and strives to implement modern approaches in the organization, storage, access and reuse of data in accordance with international standards. The university is interested in creating

an effective ecosystem for research data management, which includes the development of policies, the implementation of specialized tools and the training of researchers.

K. Tynystanov Issyk-Kul State University seeks to improve data organization, storage, access, sharing, and reuse. It aims to align its research and education with international standards, enhancing scientific output, data integrity, and open science initiatives, which it is closely following and establishing solutions for.

Naryn State University, relies predominantly on local data storage within the university. A lot of research data are stored on paper, in some cases we use Google Drive and the university server. In the Kyrgyz Republic, at present, a specialized national policy directly regulating research data management exists only at the initial level and its implementation in universities is very important. Our university does not have data warehouses, servers or repositories accessible to researchers. We are very interested in developing infrastructure, training and expanding access to repositories, developing joint standards or improving storage, identifying requirements for the development of data management plans (DMP) for researchers, as well as international platforms in this area.

We can highlight the significant challenges in research data management within universities, such as:
1. Problem Statement
• Lack of infrastructure: There are no repositories, servers or centralized platforms for storing and sharing research data.
• Low level of regulation: Specific national policies are in their infancy, creating uncertainty for universities in terms of regulation and standards.
• Disjointed approaches: The use of paper-based media and various external platforms such as Google Drive leads to data inconsistency.
• Limited access to training and resources: Lack of knowledge and tools for effective data management, especially in terms of data management plans (DMPs), international standards and platforms.
• Lack of collaboration: Lack of integration with international platforms and standards limits the ability to share and use data.

2 Project Aim

This project aims to address the pressing issues in research data management (RDM) at the partnering universities by leveraging best practices already established at other institutions. By examining and comparing RDM practices and systems/tools deployed, setting up new instances of repositories and performing pilot studies with researchers, we aim to develop a comprehensive framework that can be adapted and further developed for use in universities, ensuring better data management, higher research productivity, and compliance with international standards.

It will aim to define a specific implementation plan and address data management issues focusing on:
Data location/storage

- Creating local data repositories or clusters integrated with cloud systems.
- Increasing the maturity levels of RDM with respect to the FAIR principles (Findable, Accessible, Interoperable, Reusable).
- Investing in server hardware and ensuring its protection.

Data ownership

- Developing an internal university policy regulating data ownership.
- Taking into account national and international legal norms, including data protection laws.

Data consistency

- Establishing uniform standards for data formats and tools for their management.
- Training researchers in the correct structuring and classification of data.

Data privacy and sharing

- Implementation of a privacy policy that complies with GDPR (or other national laws).
- Use of encryption and access management to enhance security.

3 Proposed Solutions

The project will propose a multi-tiered solution focusing on the following key areas:

1. **RDM Policy:** Evaluate the existing RDM Policy of TU Wien for suitability for the partner universities, compare commonalities and differences between the RDM policies at TU Wien and Naryn State University, helping with the establishment of such a policy / the harmonization of policies as a basis for a structured and sustainable approach to establishing advanced RDM services.
2. **Data Management Plans (DMPs)**: Devise a comprehensive Data Management Planning process that ensures proper handling, preservation, and sharing of research data. This process will integrate best practices from Austrian universities, adapted for the specific context of the partner institutions following an investigation of the differences in research processes and requirements of pilot researchers.
3. **Research Data Management Framework**: Develop a flexible and scalable framework tailored to Kyrgyz universities, based on insights gained from TU Wien and the experiences gained with RDM services at the partner institutions. The framework will address data storage, ownership, confidentiality, big data handling, and advanced analysis tools.
4. **Establish and Evaluate Pilots**: Pilot installations of data repositories will be set-up and evaluated with data from selected domains. This will allow us to obtain a shared understanding of the requirements, needs for adaptations to existing open source systems, and have a solid basis for devising the processes towards establishing solid data management services.
5. **Define an implementation plan** for moving the pilots into sustainable services that can be rolled out on an institutional scale and considering cooperation (e.g. data mirroring services) opportunities

4 Project Activities / Methodologies

1. **Study Visits/Comparative Study**: Identify the scope of RDM needs and requirements at Kyrgyz universities and presenting the current best-practice solutions being deployed in Austria. Organize study visits and presentations to understand the existing RDM frameworks, technology infrastructure, and governance models. These visits will involve key stakeholders from the EPU partner universities, including IT experts and researchers from selected pilot domains. This will form the basis for defining policy documents and planning the set-up of pilots and demonstrators.
2. **Perform Focus Group Discussions:** We will conduct Focus Group Discussions (FGDs) with various stakeholders (e.g., researchers, IT staff, administrators) at the partner universities in Kyrgyz to gather insights on current challenges and tailor the RDM framework to local needs, eliciting technical and legal requirements as well as understanding the impact of ethical

aspects and challenges to trustworthiness issues. FGDs will also be held with Austrian partners to ensure alignment with international best practices and identify potentially differing needs.

3. **Set-up Pilots**: Following the study visits, pilot instances of repositories will be set up at Kyrgyz universities, complementing existing services. Research data from selected disciplines will be fed into the pilot repositories. Their suitability to meet the needs of researchers will be evaluated by setting up specific research / data analysis processes, visualizations etc.
4. **Devising a Strategy:** Based on the experiences from the pilots, a strategy will be devised how these services can be integrated into a comprehensive research data management framework.
5. **Project sustainability.** Training of personnel and employees of the scientific library named after Zh.Saalaev, IT department of NSU, the RDM team at KSTU, and continuous professional development of researchers and technical specialists will be prepared and maintained after the project finishes. The partners will further progress with the development of internal regulations supporting the use and updating of infrastructure. The Department of Information Technologies of NSU will be responsible for data management after training.

5 Future cooperation perspectives

Assuming that the pilots will be perceived as a success, the goal for subsequent cooperation will be to cooperate on turning the pilots into sustainable services. This will include finetuning the services and tools, integrating them with the local service infrastructure. This will, by their interoperability and aligned design, also offer potential compatibility with other global developments, specifically the emerging European Open Science Cloud (EOSC), offering potential for further collaborations in the framework of EU projects.

6 References

1) Flicker, K., Rauber, A., Kern, B., Ekaputra, F. (2024). Factors Influencing Perceptions of Trust in Data Infrastructures. International Journal on Digital Curation, 18(1), 2024. DOI: 10.2218/ijdc.v18i1.921
2) Miksa, T., Oblasser, S., Rauber, A. (2021). Automating research data management using machine-actionable data management plans. *ACM Trans. Manag. Inf. Syst.*, vol. 13, no. 2, pp. 1–22, 2021. DOI: 10.1145/349039
3) Miksa, T., Rauber, A., Ganguly, R., Budroni P. (2017). Information integration for machine actionable data management plans. *International Journal of Digital Curation*, 12(1). DOI: 10.2218/ijdc.v12i1.529
4) Moser, M., Weise, M., Tsepelakis, S., Miksa, T. & Rauber, A. (2024). Repositories and Computation: Crossover Episode. *Open Repositories Conference*, June 04-06th (Gothenburg, Sweden). DOI: 10.5281/zenodo.12542572
5) Weise, M., Rauber, A. (2024). Trusted Research Environments: Analysis of Characteristics and Data Availability. *International Journal of Digital Curation*, 18(1). DOI: 10.2218/ijdc.v18i1.939.
6) Weise, M., Staudinger, M., Michlits, C., Gergely, E., Stytsenko, K., Ganguly, R., & Rauber, A. (2022). DBRepo: a Semantic Digital Repository for Relational Databases. *International Journal of Digital Curation*, 17(1):11. DOI: 10.2218/ijdc.v17i1.825