EOSC-Life

EOSC-Life: Building a digital space for the life sciences

D8.2 – Common Framework for Quality, Data and Service Management

WP8 - International Impact, Innovation and Sustainability
Lead Beneficiary: INFRAFRONTIER
WP leader: Michael Raess and Jing Tang
Contributing partner(s): EATRIS, UH, ECRIN, EMBRC, ERINHA, EMBL-HD, EMBL-ELIXIR, FZJ, EU-OS, INFRAFRONTIER, IMTM, IRFMN

Authors of this deliverable: **Michael Raess, WP8 partners**

Contractual delivery date: **31 Aug 2023**
Actual delivery date: **31 Aug 2023**
H2020-INFRAEOSC-2018-2

Grant agreement no. 824087
Horizon 2020
Type of action: RIA

Table of Contents

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

Executive Summary

Building on the existing "Common BMS RI Framework for Quality Management"[1], this document describes the main aspects of quality management in research infrastructures and their role in standardising operational procedures and improving the reliability and reproducibility of results.

Implementing a structured approach to research data management is critical to improving the integrity, accessibility, and reusability of data. The formulation of data management plans is mandatory for EC-funded projects. Data provenance, FAIRification and secure access to sensitive data are essential elements of data management. EOSC-Life has developed a set of tools and resources to help research infrastructures manage these aspects effectively.

For research infrastructures that offer multiple data services, a structured approach to managing IT services can be beneficial. The FitSM methodology is highlighted in this context, since its flexible and lightweight approach is particularly suited to smaller organisations.

By integrating these structured approaches to quality management, data management and IT service management, research infrastructures can significantly improve the overall quality, integrity and reliability of their research activities and services.

Project Objectives

With this deliverable, the project has contributed to the objectives of extending the quality management framework developed in CORBEL towards service management in line with the EOSC developments.

Detailed Report on the Deliverable

1. Quality Management in research infrastructures

1.1. Quality management in a research environment

Quality management in an organisation is a systematic approach to planning, coordinating, and controlling activities to ensure that products or services meet or exceed customer expectations. It involves setting quality standards, implementing procedures to achieve those standards, and

[1] "A framework for quality management in the biomedical research infrastructures (BMS RIs)" https://doi.org/10.5281/zenodo.834333

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

continuously monitoring and improving performance to ensure consistent quality throughout the organisation. The goal of quality management is to improve customer satisfaction, increase operational efficiency and promote the overall success of the organisation.

Quality control in an academic research setting is primarily concerned with ensuring the integrity and reliability of research results. It involves processes and procedures to ensure that research is conducted in a rigorous and accurate manner. Quality control in research institutes usually focuses on compliance with established research protocols, ethical guidelines, and legal requirements. The aim is to ensure the validity and credibility of research findings and to maintain the reputation and trustworthiness of the institute.

Quality management in a research institution can also play a crucial role in the reproducibility of data. By establishing sound quality control processes and procedures, research institutions ensure that research is conducted rigorously and accurately. This can include adherence to established research protocols, ethical guidelines, and legal requirements.

The aim of data reproducibility is to ensure the validity and credibility of research findings. If research results can be reproduced by independent researchers using the same methods and data, this strengthens confidence in the accuracy and reliability of the results. Reproducible research enables verification and validation, which are essential for scientific progress and facilitating further research.

Research data that cannot be reproduced can have significant economic consequences[2]. They lead to a waste of funding and time, hinder innovation, and technological progress, discourage investment and damage the reputation of research institutions and individual researchers. Several research publications have discussed the economic impact of non-reproducible research, highlighting the importance of reproducibility for effective decision-making, resource allocation and economic growth[3].

By emphasising quality control, research institutions can minimise errors, biases and inconsistencies in data collection, analysis, and interpretation. This strengthens confidence in the accuracy and reliability of the results. It ensures that research results can be trusted and used by other researchers, policy makers and the wider scientific community and ultimately helps to remediate the economic consequences of non-reproducible research.

Systematic approaches to quality management in research institutions can include some or all of the following features:

- Standard Operating Procedures (SOPs): Research institutions develop and implement SOPs that outline standardized processes and protocols for various research activities. These SOPs ensure consistency and accuracy in data collection, analysis, and reporting.
- Quality Assurance (QA) Programs: QA programs involve systematic monitoring and evaluation of research activities to ensure compliance with established standards and protocols. This includes regular audits, inspections, and reviews of research processes and documentation.
- Training and Education: Research institutions provide training and education programs to researchers and staff members to enhance their knowledge and skills in research methodologies, data management, and quality control practices. This helps to ensure that research activities are conducted in a rigorous and accurate manner.

[2] Freedman LP, Cockburn IM, Simcoe TS (2015) The Economics of Reproducibility in Preclinical Research. PLoS Biol 13(6): e1002165. https://doi.org/10.1371/journal.pbio.1002165

[3] In the context of the EOSC-Life project's WP8, a Methodology framework to enhance reproducibility within EOSC-Life has been developed: https://doi.org/10.5281/zenodo.4705078.

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

- Document Control: Implementing document control systems ensures that research documentation, including protocols, consent forms, and data records, are properly managed, version controlled, and easily accessible. This helps maintain the integrity and traceability of research data.
- Continuous Improvement: Research institutions focus on continuous improvement by regularly assessing and analysing research processes and outcomes. This involves identifying areas for improvement, implementing corrective actions, and monitoring the effectiveness of these actions to enhance the quality of research.

1.2. Specifics of quality management in research infrastructures

Some features of quality management are specific to research infrastructures:

Infrastructure-specific standards: Research infrastructures have specific standards and requirements that are unique to their specialised facilities and services. Quality management provides means to ensure compliance with these specific standards, e.g., security protocols for high-risk environments or data management standards for big data centres.

Service-oriented approach: research infrastructures usually provide services and resources to a wide range of users, including researchers from different disciplines, industry partners and external collaborators. Quality management in research infrastructures focuses on ensuring the provision of high-quality services, meeting the needs of different user groups, and maintaining customer satisfaction.

Collaborative networks: Due to the distributed nature of many research infrastructures, their quality management requires establishing and maintaining collaborative networks, ensuring effective communication, and harmonising quality standards with partner institutions.

Access and user management: research infrastructures need procedures in place to manage access to their facilities, equipment, and resources[4]. Quality management facilitates the development and implementation of access policies, user selection procedures and user training programmes to ensure the proper use and management of infrastructure resources.

Technological and instrumental considerations: Research infrastructures provide access to advanced technologies, specialised instruments, and complex equipment. Quality management facilitates the optimisation of operation to ensure accurate and reliable research results.

Long-term sustainability: Research infrastructures require long-term planning and sustainability strategies to ensure their continued operation and availability to the research community. Quality management addresses issues of financial stability, resource allocation and strategic planning to support sustainable quality services and infrastructure development.

In the implementation of quality managing systems on a research infrastructure level standardised approaches are often applied:

[4] The Life Science Login allows researchers to access data and services using their home organization credentials or other identities like Google, Linkedin, or LS ID. It also enables service providers to manage user access rights and create different access levels for research groups or international projects, benefiting both academia and industry. https://lifescience-ri.eu/ls-login/

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

- Good laboratory practice (GLP): GLP is a set of guidelines and regulations that set the standards for conducting non-clinical laboratory studies. It provides a framework for quality control in areas such as study design, data collection, documentation, and reporting.
- Good clinical practice (GCP): GCP is an international ethical and scientific quality standard for the design, conduct, recording and reporting of human clinical trials. It ensures the protection of participants' rights, the integrity of data and the credibility of trial results.
- ISO 9001: ISO 9001 is a widely recognised international standard for quality management systems. Research institutions can adopt this standard to implement a comprehensive quality management system covering various aspects of research activities, including protocol development, data collection, analysis, and reporting.
- Accreditation programmes: research infrastructures may seek accreditation from recognised bodies specialising in quality control and standardisation in research. These accreditation programmes assess the institution's compliance with specific quality standards and best practices in research.

1.3. A framework for quality management in the biomedical research infrastructures

All the biomedical research infrastructures participating in EOSC-Life have agreed to a *Common BMS RI Framework for Quality Management*[5]. The 13 biomedical research infrastructures endorsing this document recognize that proper use of tested reference materials, standard operating protocols, study designs, data analysis and data storage is critical in biomedical research to improve quality and reduce waste of resources. Biomedical research infrastructures are actively involved in the development of community standards and references for biomaterials, reagents, technology platforms, operational procedures, data acquisition and processing. To ensure consistent application of these standards, they support quality management methods and systems. These systems ensure that the resources and services provided by the BMS RIs are of the highest quality, deliver reliable and reproducible results, and meet regulatory requirements. This benefits public funders, publishers, journals, academia, and industry, ultimately contributing to more efficient delivery of medicines and addressing societal health challenges.

In the document it is acknowledged that open access to research data is an important quality measure, however research data management and data service management are not specifically addressed.

2. Research data management

2.1. Ensuring integrity, accessibility, and usability of data throughout its lifecycle

Research data management is the process of organising, storing, preserving, and sharing data generated by research activities. It encompasses various tasks and considerations related to the life cycle of research data, from its creation to its eventual archiving or disposal. The aim of research data management is to ensure the integrity, accessibility, and usability of data throughout its lifecycle. The most important aspects of research data management are:

[5] https://doi.org/10.5281/zenodo.834332

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

Data organisation: Research data should be organised and structured in a consistent and logical way to allow easy retrieval and understanding. This refers to naming conventions, file formats, folder structures and metadata documentation.

Data storage and backup: Adequate storage and backup systems should be in place to protect research data from loss or damage. This may involve the use of secure servers, cloud storage or physical backup solutions to ensure data integrity and availability.

Data documentation and metadata: Research data should be accompanied by comprehensive documentation and metadata that includes information about the data collection process, variables, methods, and any relevant contextual details. This helps ensure data reproducibility and supports data sharing and reuse.

Data security and privacy: Appropriate measures must be taken to ensure the security and privacy of research data; particularly where sensitive or personally identifiable information is involved. Usually this requires encryption, access controls and compliance with relevant data protection regulations.

Data sharing and access: Research data should be shared and made available to the research community whenever possible and appropriate. This promotes transparency, collaboration and the potential for further analysis and validation. Data sharing can be done through repositories, data archives or controlled access platforms.

Data retention and archiving: Long-term retention and archiving of research data is essential for future reference and potential reuse. Data should be stored in a format that ensures its long-term accessibility and usability, taking into account technological obsolescence and evolving data standards.

2.2. Data management plans

Data Management Plans (DMPs) are mandatory for effective data management in EC-funded projects. DMPs describe the life cycle of research data, including collection, processing, and preservation. They ensure that data are FAIR (findable, accessible, interoperable, and reusable) by addressing aspects such as data handling, collection details, methodology and standards, data sharing and long-term preservation.

The EC provides a set of guidelines on FAIR data management that includes a DMP template[6]. In order to assist with the setting up of a DMP, the Fairsharing.org website provides information on data standards and the databases that implement them[7].

2.3. The concepts of data provenance and FAIR data in research data management

Data provenance and the FAIR Data Principles are critical components of research data management and contribute to data integrity, reproducibility, and usability.

Data provenance refers to the documentation and tracking of the origin, history, and transformation of research data. It includes information about data sources, collection methods, processing steps and any

[6] These guidelines have been developed for Horizon 2020 but are also valid for Data Management in Horizon Europe: https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

[7] Fairsharing Educational FAQ – Are you a researcher creating a Data Management Plan? https://fairsharing.org/educational#faq6-1

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

changes made to the data. By recording data provenance, researchers can trace and verify the origin and quality of data, ensuring transparency and facilitating the reproducibility of research results.

By adhering to the FAIR principles[8], research data becomes more discoverable through metadata and persistent identifiers, accessible through open access or controlled access mechanisms, interoperable through the use of standard formats and protocols, and reusable for future research purposes.

Both data provenance and the FAIR Data Principles contribute to the governance of research data:

By recording data provenance, researchers can trace and verify the origin and quality of data, ensuring transparency and facilitating the reproducibility of research results. The research data becomes more discoverable through metadata and persistent identifiers, accessible through open access or controlled access mechanisms, interoperable through the use of standard formats and protocols, and reusable for future research purposes.

Data provenance and the FAIR Data Principles facilitate data sharing and collaboration between researchers. Data provenance provides information about the context and quality of shared data, and adherence to FAIR principles ensures that shared data is easily discoverable and usable by others. FAIRness of data increase facilitates its access and use for future studies and data provenance increases the trustworthiness and credibility of reused data by providing information on their origin and processing history.

2.4. EOSC-Life contributions to data provenance and tools for data FAIRification

Within the EOSC-Life project, WP6 focused on creating a common understanding of provenance information in the field of life sciences. The aim was to create interoperable and harmonised provenance information between different organisations involved in research, such as biobanks, research centres, universities, and laboratories. This led to the development of the Common Provenance Model (CPM). The CPM defines the design of distributed provenance components that enable the linking of provenance information generated by different research organisations[9]. The CPM also serves as an open conceptual foundation for the development of the *ISO 23494 - Provenance information model for biological material and data standard* series[10].

WP6 of EOSC-Life has also developed the FAIRsharing Wizard[11] which that helps stakeholders explore and understand the 1641 standards and 2041 databases in the FAIRsharing register[12]. The FAIRsharing Register ensures that the standards and databases are well described, linked, and visualised. It enables their discoverability, accessibility, interoperability, and reusability. The FAIRsharing Wizard helps users filter and select relevant datasets based on subject areas, data types and specific requirements. It is complemented by the FAIRsharing Educational page[13] with infographics and factsheets and the FAIRassist page that guides users to existing tools for assessing FAIRness.

[8] Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

[9] D6.6 — Common provenance model for processing biological material, data generation and computational workflows: https://doi.org/10.5281/zenodo.8279525

[10] ISO/TS 23494-1:2023 - Biotechnology — Provenance information model for biological material and data — Part 1: Design concepts and general requirements: https://www.iso.org/standard/80715.html

[11] D6.4 — FAIR assistance: https://doi.org/10.5281/zenodo.8264604

[12] https://fairsharing.org/

[13] https://fairsharing.org/educational

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

The FAIR Cookbook[14] is a comprehensive resource for the life sciences that provides guidance and examples for implementing FAIR data management. It covers technical aspects and provides recipes for researchers, data managers, software developers and policy makers. The cookbook aims to make data discoverable, accessible, interoperable, and reusable by embedding FAIR principles into data generation. It includes recipes for managing data access, selecting licences, packaging data, providing metadata, and applying FAIRification to different types of data. It supports the development of high-quality data management practices and can be used by anyone involved in the data management lifecycle.

Bioschemas[15] improves data interoperability in the life sciences by promoting the use of Schema.org markup on websites. This ensures that websites and services have machine-readable metadata with a consistent structure, making it easier to find, summarise and analyse distributed data. In addition to adding life science types to Schema.org, Bioschemas has also developed specifications to promote the consistent adoption of Schema.org markup in the field. The Bioschemas collection also lists resources that already use the Bioschemas markup.

RO-Crate[16] is a community-driven initiative that simplifies the packaging of research data with metadata. It makes the formal description of metadata accessible and practical for a range of scenarios, from individual researchers working with data folders to large-scale computational research environments. Importantly, the Common Provenance Model (CPM, see above) was integrated with RO-Crate. Key target users of RO-Crate include computationally and data intensive researchers, digital repository managers, infrastructure providers, individual researchers looking for a user-friendly tool to make their data FAIR, and data managers supporting research projects in the creation and maintenance of datasets.

This section would not be complete without referencing RDMkit as a comprehensive resource for the research data management in the life sciences provided by ELIXIR[17].

2.5. Managing access to sensitive data

Access to sensitive data is a topic of particular importance of those life sciences research infrastructures dealing with patient-derived health data: A detailed report from EOSC-Life WP4[18] focuses on policies, specifications, and tools for the secure management of sensitive data in research. It gathers requirements for controlled access to sensitive data from life sciences research infrastructures and the EOSC-Life work packages dealing with user management, access services and cloud deployment and national requirements for hosting and distributing sensitive data. The report analyses data in clouds and assesses cloud providers' compliance with these requirements. The report also provides an overview on the types of sensitive data handled by the life sciences research infrastructure and collects regulations, best practices, and standards.

Sensitive data in life science research infrastructures constitute a challenge for the categorisation and localisation of the corresponding digital resources. To improve the FAIRification process, a toolbox demonstrator was developed as part of EOSC-Life WP4. This toolbox[19] improves the search for digital

[14] https://faircookbook.elixir-europe.org/content/home.html
[15] https://fairsharing.org/3517
[16] https://www.researchobject.org/ro-crate/
[17] RDMkit (elixir-europe.org)
[18] D4.1 - Requirements for hosting/distributing and access control for sensitive data: https://doi.org/10.5281/zenodo.7037444
[19] Toolbox for sensitive data objects: https://fairsharing.org/3577

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

objects such as regulations, guidelines, best practices, and tools. It is based on a categorisation system that is harmonised across several life science research infrastructures.

3. IT Service Management

3.1. Established standards for IT service management

IT service management (ITSM) is used by organisations in various sectors, including private companies, government agencies, educational institutions, healthcare organisations, financial institutions, and service providers. ITSM helps these organisations to effectively manage their IT services, align them with business objectives and ensure operational efficiency. By implementing ITSM, organisations can improve service quality, increase customer satisfaction, and optimise their IT infrastructure and operations.

IT service management encompasses a set of activities, processes, and practices that organisations use to design, deliver, manage, and improve IT services to meet the needs of their customers. It also facilitates the alignment of IT services with business objectives, the effective and efficient delivery of services, and the continuous improvement of service quality.

There are several standards and frameworks that organisations can use to guide their IT service management practices. Some of the widely accepted standards are:

IT Infrastructure Library (ITIL): ITIL[20] provides a comprehensive set of best practices for IT service management. It covers various aspects such as service strategy, service design, service transition, service operation and continuous service improvement.

ISO/IEC 20000: This ISO standard[21] specifies the requirements for an IT service management system. It outlines the processes, controls, and metrics that organisations must implement to achieve effective IT service management.

COBIT (Control Objectives for Information and Related Technologies): COBIT[22] focuses on the management and control of IT processes. It provides a framework for organisations to ensure that IT services are aligned with business objectives and comply with relevant regulations.

A common characteristic of these frameworks is their comprehensiveness, which comes at the costs of the considerable resources required for their implementation.

3.2. FitSM as a lightweight alternative approach for IT service management

FitSM[23] is a simple and practical standard for IT service management (ITSM) that is particularly suitable for small and medium-sized organisations. It provides a set of guidelines and best practices for implementing effective ITSM processes. It focuses on simplicity and flexibility and provides a streamlined approach to ITSM without the need for extensive documentation or complex procedures.

[20] https://www.axelos.com/certifications/itil-service-management
[21] https://www.iso.org/standard/70636.html
[22] https://www.isaca.org/credentialing/cobit-foundation
[23] https://www.fitsm.eu

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

FitSM is aligned with ITIL, ISO/IEC 20000 and other related standards, making it compatible with established frameworks.

FitSM covers the key areas of ITSM, including service management principles, service lifecycle management, service planning, service operations and continuous service improvement. It emphasises the importance of customer satisfaction, service quality and continuous improvement. The FitSM standard is designed to be accessible and easy to implement, making it a practical choice for organisations that are new to ITSM or have limited resources. It provides a foundation for organisations to implement effective IT service management practices, improve service delivery and increase overall IT performance.

FitSM is particularly well suited to small and medium-sized organisations such as research infrastructures. It offers a streamlined approach to ITSM without complex procedures or extensive documentation. Because of its simplicity and flexibility, it can be tailored to an organisation's specific needs and processes. This adaptability is particularly beneficial for smaller organisations with limited resources. In addition, compared to more comprehensive frameworks, FitSM requires significantly less time to implement and therefore offers rapid adoption and quick results.

4. Conclusion: Research data management, IT service management, and quality management in research infrastructures

Data management plays a crucial role in supporting the overall quality management of research infrastructures. By implementing effective data management procedures, research infrastructures can improve the overall quality of their activities and services. Proper data management ensures that research data are accurate, reliable, and error-free. This promotes the integrity of research results and prevents the dissemination of false or misleading information. Appropriate data management practices such as documentation, version control and data sharing protocols facilitate the reproducibility of research results. Transparent data management processes allow others to review and validate research results, improving the overall quality and reliability of research outputs. Research infrastructures are subject to various legal requirements, such as data protection regulations or ethical guidelines. Effective data management helps to ensure compliance with these requirements, protect the privacy and confidentiality of research participants and secure sensitive data. By implementing robust data security protocols, research infrastructures can mitigate risks of unauthorised access, data loss and corruption and ensure the confidentiality and availability of research data.

In addition to the importance of data management for the quality management of research infrastructures, the concept of IT service management can further enhance data management procedures. ITSM provides a framework for implementing effective IT service management processes that can be of great benefit to research infrastructures' data management. By implementing ITSM, research infrastructures can ensure that their IT services and systems supporting data management comply with best practices and industry standards. By incorporating the concept of IT service management as offered by FitSM, research infrastructures can improve their data management practices and ensure accurate, reliable, and secure research data. This ultimately improves the overall quality, integrity and trustworthiness of research activities and services offered by research infrastructures.

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.

Delivery and Schedule

The delivery is delayed: Yes

The deliverable was delayed to August 2023 to include the results of a final workshop dedicated to this topic that was held in the frame of the 4th and final AGM at the 27/28 of June.

Adjustments

Adjustments made: None

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 824087.