Deliverable D7.1

Data Management Plan, Knowledge Management Resources and Quality & Risk Management Plan

Work Package	Project Management
Task(s)	T7.1 Project Data, Content, and Knowledge Management Infrastructure T7.2 General operational and financial management T7.3 Technology and scientific management T7.4 Open Science and Data Management T7.5 Quality and Risk Management
Author(s)	Binh Vu
Document ID	D7.1_ Data Management Plan, Knowledge Management Resources and Quality & Risk Management Plan _M6
Approver	Matthias Hemmje
Creation date	10/02/2025
Last update	20/08/2025
Version	V8
Supersedes	NA
Dissemination Level	CO

*All dates in dd/mm/yyyy format

This project received funding from the European Union
MSCA SE, Grant Agreement No 101182801

Document Log

Date	Author	Changes	Version	Status
10/02/2025	Binh Vu	Chapter 1	V1	draft
11/02/2025	Binh Vu	Chapter 2	V2	draft
12/02/2025	Binh Vu	Chapter 3	V3	draft
13/02/2025	Binh Vu	Chapter 4	V4	draft
13/02/2025	Thomas Krause	All chapters	V5	draft
14/02/2025	Jana Mertens	All chapters	V6	draft
19/02/2025	Ruben Riestra	Template	V7	draft
04/07/2025	Binh Vu	Executive summary	V8	final

This project received funding from the European Union
MSCA SE, Grant Agreement No 101182801

Contents

This project received funding from the European Union
MSCA SE, Grant Agreement No 101182801

This project received funding from the European Union
MSCA SE, Grant Agreement No 101182801

Executive Summary

Overall Assessment

This deliverable, D7.1, presents the integrated Data Management Plan (DMP), Knowledge Management (KM) strategy, and Quality & Risk Management Plan (QRM) for the GenDAI project. These three plans establish the comprehensive operational, technical, and governance framework that will guide all project activities. The main results include the FAIR-aligned and GDPR-compliant DMP, the architectural design of the centralized Knowledge Management Ecosystem Portal (KM-EP), and a detailed QRM plan with a proactive Risk Register.

The value added by these integrated results is the creation of a robust and coherent framework that ensures the integrity, security, and reproducibility of the project's scientific work. This framework directly supports all subsequent work packages by providing clear, actionable procedures and a central platform for collaboration. Furthermore, its structured approach to data, knowledge, and risk ensures that project assets are primed for future scientific reuse and for the eventual exploitation of the innovative GenDAI diagnostics platform.

Outputs and Project Progress

This deliverable is a cornerstone for achieving the Research and Innovation Objectives (RIO) and Milestones detailed in the Description of Action (DoA), directly addressing the overarching management objectives of Work Package 7. The Data Management Plan serves as a critical prerequisite for milestones in data-centric work packages like WP1 (Data Collection), WP2 (Diagnostics Workflow), and WP4 (AI Discovery) by ensuring all data is handled in a FAIR and GDPR-compliant manner from the outset. Concurrently, the Knowledge Management strategy provides the central collaborative hub essential for partner integration and the seamless exchange of data and results. The Quality & Risk Management Plan is fundamental to meeting the project's timeline by proactively mitigating critical risks identified in the DoA, thus safeguarding the achievement of all milestones. Therefore, the delivery of this integrated plan at month 6 (M6) marks the successful completion of a foundational project milestone, establishing the necessary governance for all subsequent work.

Contributions to Impacts

This deliverable lays the groundwork for realizing the project's long-term impacts as outlined in the DoA. The established framework is instrumental in achieving the primary impact of delivering a novel, AI-driven diagnostic tool for Inflammatory Bowel Disease (IBD) by ensuring the project's data is of high quality, reliable, and ethically managed. By mandating adherence to FAIR principles, this plan also guarantees that GenDAI's datasets and software will contribute to the wider scientific community, advancing research in metagenomics and creating clear paths for future exploitation. Scientifically, the meticulously managed data will become a valuable asset for high-impact publications and reuse in future research projects. Commercially, this governance framework is essential for the future development of the GenDAI platform, establishing the quality, security, and traceability standards required for a market-ready medical diagnostic tool while de-risking the innovation process.

1 Introduction

This chapter introduces Deliverable D7.1, outlining the overall goals and scope of the GenDAI project and explaining the purpose and structure of this integrated Data Management Plan (DMP), Knowledge Management (KM) strategy, and Quality & Risk Management Plan (QRM). It provides essential context for understanding the document and its importance to the project's success.

1.1 Project Overview

The GenDAI project (Genomic applications for laboratory Diagnostics supported by Artificial Intelligence) aims to develop a novel medical diagnostics platform that leverages metagenomic data and Artificial Intelligence (AI) to improve the diagnosis and monitoring of inflammatory bowel disease (IBD). The platform will provide clinicians with a comprehensive, personalized view of a patient's microbiome, enabling more targeted and effective treatment strategies.

1.2 Purpose of this Document

This document serves as an integrated Data Management Plan (DMP), Knowledge Management (KM) strategy, and Quality & Risk Management Plan (QRM) for the GenDAI project. It outlines the procedures and policies that will be implemented to ensure:

- Data Management: The responsible collection, processing, storage, sharing, and preservation of all project data in accordance with FAIR (Findable, Accessible, Interoperable, Reusable) principles and relevant ethical and legal regulations (e.g., GDPR).
- Knowledge Management: The effective capture, organization, sharing, and utilization of project knowledge, including data, documentation, software, and expertise.
- Quality & Risk Management: The proactive identification, assessment, mitigation, and monitoring of risks to the project's success, ensuring the delivery of high-quality outputs.

This project received funding from the European Union
MSCA SE, Grant Agreement No 101182801

2 Data Management Plan (DMP)

This chapter presents the Data Management Plan (DMP) for the GenDAI project. It details the strategies and procedures for managing all project data throughout its lifecycle, from collection and processing to storage, sharing, and long-term preservation. The DMP is firmly grounded in the FAIR principles (Findable, Accessible, Interoperable, Reusable) and ensures full compliance with ethical and legal requirements, particularly the General Data Protection Regulation (GDPR). This section is critical for ensuring responsible and effective data handling, maximizing the value and impact of project data.

2.1 Data Summary

2.1.1 Types of Data

The GenDAI project will generate and utilize several types of data, including:

Type	**Sub-Type**	**Description**
Raw Genomic Data		Raw sequence data (FASTQ files) obtained from metagenomic sequencing of stool samples from IBD patients
Processed Genomic Data		Processed sequence data (e.g., taxonomic profiles, abundance tables) generated by the GenDAI Diagnostics Workflow (WP2)
Metadata	Sample Metadata	Sample collection date, storage conditions, etc.
	Patient Metadata	Anonymized patient demographics, clinical information (diagnosis, disease activity, treatment), and consent information. *Note: This will be handled with extreme care to ensure GDPR compliance.*
	Sequencing Metadata	Sequencing platform, library preparation method, sequencing depth, etc.
	Processing Metadata	Software versions, parameters used, and provenance information
AI Model Data	Training Data	Subsets of genomic and metadata used to train AI models (WP4)
	Model Weights	Parameters of trained AI models
	Model Performance Metrics	Evaluation results of AI models
Software and Code	Source Code	Code for the GenDAI Diagnostics Workflow, GenDAI Safe, GenDAI Discovery, and GenDAI Interactive Reporting components
	Docker Images	Containerized versions of software components for reproducibility
	Workflow Definitions	Definitions of data processing pipelines (e.g., using a Workflow Management System)
Documentation	User Manuals	Documentation for using the GenDAI platform and its components
	Technical Reports	Reports on the development and evaluation of the platform
	Publications	Scientific publications resulting from the project
Clinical Reports		Visualized and analyzed data provided to aid clinicians

This project received funding from the European Union
MSCA SE, Grant Agreement No 101182801

2.1.2 Data Volume and Scale

The estimated data volume and scale are as follows:

Data Type	Estimation
Raw Genomic Data	Approximately 50 GB per sample. Assuming 200 samples, the total raw data volume will be around 10 TB
Processed Genomic Data	Approximately 1 GB per sample, totalling around 200 GB
Metadata	Relatively small in size compared to genomic data (estimated at a few MBs per sample)
AI Model Data	Model weights can vary in size, but are expected to be in the range of a few GBs
Software and Code	Relatively small in size (a few GBs)

These are initial estimates and will be refined as the project progresses.

2.1.3 Data Origin and Collection Methods

Data Type	Collection Methods
Raw Genomic Data	Anonymized data sets provided by a participating laboratory, obtained from residual sample material of IBD patients under informed consent
Processed Genomic Data	Generated by the GenDAI Diagnostics Workflow (WP2) running on the GenDAI platform
Metadata	Collected and managed through a combination of manual entry (e.g., clinical data) and automated capture (e.g., sequencing and processing metadata)
AI Model Data	Generated through the GenDAI Discovery component (WP4)
Software and Code	Developed by project partners (ICT, MTU, UNIBA, SAP, OKK)
Clinical Data	Collected by clinical partners

2.1.4 Data Formats

Data Type	Formats
Raw Genomic Data	FASTQ
Processed Genomic Data	FASTA, CSV, TSV, JSON
Metadata	CSV, JSON, XML (using appropriate metadata standards)
AI Model Data	HDF5, TensorFlow SavedModel, ONNX
Software and Code	Programming language-specific formats (e.g., PHP, JavaScript, Python, Java), Dockerfile
Workflow Definitions	CWL, WDL, Nextflow DSL
Documentation	Word, PDF, HTML

2.2 FAIR Data Principles

2.2.1 Findability

Metadata Standards and Schemas

The project will adopt and adapt relevant metadata standards, as shown below:

Standard/Schema	Description
Dublin Core	For general resource description

This project received funding from the European Union
MSCA SE, Grant Agreement No 101182801

MIxS (Minimum Information about any (x) Sequence)	For describing sequence data
Custom Metadata Schema	Developed specifically for GenDAI to capture project-specific information, including elements related to AI model training and performance. This schema will be documented and made available

The metadata schema will be designed in close consultation with all partners to capture essential information for all data types.

Persistent Identifiers (PIDs)

The project will utilize the Entity Name System (ENS), as developed by OKK, to assign persistent identifiers (PIDs) to all project entities:

Entity Type	Description
Data	Raw data, processed data, and datasets
Software	Software components and workflows
Publications	Scientific publications
Other entities	As needed (e.g., AI models, clinical reports)

The ENS will ensure long-term identifiability and linkability of project resources.

Data Repositories and Registries

- Project-Specific Registry: The GenDAI Knowledge Management Ecosystem (see Section 3) will serve as a central registry for all project data and resources, providing metadata and links to storage locations.
- External Repositories: Where appropriate and in compliance with ethical and legal requirements, data will be deposited in publicly accessible repositories such as:
    - Zenodo: For general-purpose data and software archiving.
    - European Genome-phenome Archive (EGA): For controlled-access human genomic data (if applicable and after appropriate ethical approvals).
    - GitHub/GitLab: For source code and software documentation.

Other domain-specific repositories will be considered as needed.

2.2.2 Accessibility

Data Access Levels

Data access will be managed according to the following levels:

Access Level	Description
Open Access	Data freely accessible and usable by anyone, subject to appropriate licensing (e.g., anonymized metadata, software code)
Restricted Access	Data accessible only by authorized users (e.g., project partners) due to ethical or legal constraints (e.g., patient-identifiable data)

This project received funding from the European Union
MSCA SE, Grant Agreement No 101182801

Embargoed Access	Data initially restricted but made open access after a specified embargo period (e.g., to allow for publication)

Authentication and Authorization

System/Component	Description
GenDAI Safe	The GenDAI Safe component (WP3) will provide authentication and authorization services
ASBAC	The ASBAC system (T3.6) will provide fine-grained access control
External Repositories	Access control mechanisms of external repositories will be used where data is deposited

Data Use Agreements

For restricted-access data, Data Use Agreements (DUAs) will be established between the data providers (FTK and clinical partners) and data users. These DUAs will specify the terms and conditions of data access and use, including:

- Purpose of data use.
- Restrictions on data sharing and redistribution.
- Data security requirements.
- Attribution and acknowledgement requirements.
- Compliance with ethical and legal regulations.

2.2.3 Interoperability

Data Standards and Vocabularies

The project will utilize controlled vocabularies and ontologies to ensure semantic interoperability:

Vocabulary/Ontology	Description
NCBI Taxonomy	For describing microbial taxa
Human Phenotype Ontology (HPO)	For describing clinical phenotypes
SNOMED CT	For clinical terminology
Other ontologies	As needed

The project will also contribute to developing a formal, extensible ontology for the microbiome research domain (WP3).

Data Exchange Formats

Standard data exchange formats will be used to facilitate data sharing and integration between project partners and systems. These formats are listed in Section 2.1.4.

API Specifications

If APIs are developed for accessing data or services, they will be documented using standard API specification languages (e.g., OpenAPI/Swagger).

2.2.4 Reusability

Data Documentation

Comprehensive documentation will be provided for all data types.

Documentation Type	**Description**
Data Dictionaries	Defining data fields, data types, and units of measurement
README Files	Describing the contents and structure of datasets
Provenance Information	Tracking the origin, processing steps, and versions of data (linked to GenDAI Diagnostics Workflow and GenDAI Safe)
Codebooks	Explaining the meaning of codes and abbreviations used in data
Metadata Records	Complete and accurate metadata records conforming to the chosen metadata standards

Data Licensing

- Data: Anonymized data will be released under a Creative Commons Attribution (CC BY) license, or a similar open license, where possible. Restricted-access data will be governed by DUAs.
- Software: Software will be released under an open-source license (e.g., MIT License, Apache License 2.0).

Data Provenance

The GenDAI Diagnostics Workflow (WP2) and GenDAI Safe (WP3) will capture and maintain detailed provenance information:

Provenance Information	**Description**
Input Data	Sources of input data
Processing Steps	Software versions, parameters, and scripts used
Output Data	Generated data files and their relationships to input data
User Actions	Who performed the processing steps
Timestamps	When the processing steps were performed
PIDs	Persistent identifiers for each step

This provenance information will be essential for ensuring reproducibility and traceability of results.

2.3 Data Storage, Backup, and Preservation

2.3.1 Short-term Storage

During the project, data will be stored using a combination of:

- Google Cloud Resources: Provided by SAP (WP3), including storage (Google Cloud Storage), computing (Compute Engine, Kubernetes Engine), and AI/ML services (Vertex AI).
- Local Servers: Used by individual partners for local data processing and analysis (if necessary).

The GenDAI Knowledge Management Ecosystem will provide a unified interface for accessing data stored in different locations.

2.3.2 Backup Procedures

Procedure	Description
Regular Backups	Data on Google Cloud backed up using Google Cloud's built-in services
Multiple Copies	Multiple copies of data maintained in different geographic locations
Backup Verification	Backups regularly verified for integrity and recoverability
Disaster Recovery Plan	Detailed plan to address data loss scenarios (e.g., hardware failure, natural disasters)

2.3.3 Long-term Archiving

Archiving Aspect	Description
System	GenDAI Safe (WP3) based on OAIS (ISO 14721)
Data Formats	Data archived in formats suitable for long-term preservation (open, well-documented formats)
Storage Location	Secure and reliable infrastructure (to be determined)
Access Policies	Managed according to data access levels (Section 2.2.2)

2.4 Data Sharing and Publication

The GenDAI project is committed to open science principles and will share data as openly as possible, while respecting ethical and legal constraints.

- Anonymized Data: Anonymized data will be made publicly available through appropriate repositories (see Section 2.2.1).
- Restricted-Access Data: Access to patient-identifiable data will be strictly controlled and governed by DUAs.
- Embargo Periods: Data may be subject to embargo periods to allow for publication.

A data publication plan will align data release with scientific publications, detailing timelines and appropriate repositories. All shared data will receive Persistent Identifiers (PIDs) via the project's Entity Name System (ENS) and should be cited using standard formats (e.g., DataCite) to ensure proper attribution and discoverability. This promotes reproducibility and wider scientific use.

2.5 Ethical and Legal Compliance

2.5.1 GDPR Compliance

The GenDAI project will adhere strictly to the General Data Protection Regulation (GDPR). Key measures include:

This project received funding from the European Union
MSCA SE, Grant Agreement No 101182801

Measure	Description
Data Protection Officer (DPO)	FTK will be responsible to ensure compliance
DPIA	A Data Protection Impact Assessment will be conducted
Privacy by Design/Default	Privacy considerations integrated into project systems and processes
Data Minimization	Only the minimum necessary personal data will be collected and processed
Data Security	Appropriate measures to protect personal data (Section 2.5.4)
Data Subject Rights	Procedures to ensure data subjects can exercise their rights (access, rectification, erasure)

2.5.2 Informed Consent

The consortium is planning to use anonymized data sets obtained from residual sample material for which consent was already obtained in the past. Should the procedure nevertheless need to be adapted, because further data sets are required, informed consent would be be obtained from all patients participating in the study, in accordance with ethical guidelines and regulations (WP1). The informed consent process would cover:

- The purpose of the study.
- The types of data that will be collected.
- How the data will be used and shared.
- The risks and benefits of participation.
- The data subject's rights.
- Contact information for the DPO.

Consent forms and information sheets would be reviewed and approved by relevant ethics committees.

2.5.3 Data Anonymization and Pseudonymization

- Anonymization: Where possible, data will be anonymized to remove all direct and indirect identifiers.
- Pseudonymization: If anonymization is not feasible, data will be pseudonymized by replacing identifying information with pseudonyms. The mapping between pseudonyms and identifiers will be stored securely and separately from the pseudonymized data.

Specific anonymization and pseudonymization techniques will be documented and reviewed to ensure their effectiveness.

2.5.4 Data Security

The following data security measures will be implemented:

- Access Control: Strict access control policies will be enforced (see Section 2.2.2).
- Encryption: Data will be encrypted both in transit and at rest.

This project received funding from the European Union
MSCA SE, Grant Agreement No 101182801

- Regular Security Audits: Regular security audits will be conducted to identify and address vulnerabilities.
- Intrusion Detection and Prevention: Intrusion detection and prevention systems will be deployed.
- Data Breach Response Plan: A data breach response plan will be developed to address potential data breaches.

2.6 Roles and Responsibilities

Partner	**Responsibilities**
FTK	Overall project coordination, data management oversight, ethical and legal compliance. DPO responsibility.
MTU	AI model development, feature selection, NLP techniques
UNIBA	Visual data analysis, interactive reporting, usability and UX
SAP	Cloud infrastructure, data management, long-term archiving, GenDAI Safe development
ICT	Diagnostics workflow, platform integration, piloting and validation
OKK	ENS implementation, reproducibility, ASBAC support
All Partners	Contribute to data collection, processing, analysis, and documentation. Adherence to the DMP

This project received funding from the European Union
MSCA SE, Grant Agreement No 101182801

3 Knowledge Management Resources

This chapter describes the Knowledge Management (KM) strategy for the GenDAI project, focusing on the establishment and utilization of a centralized, secure Knowledge Management Ecosystem Portal (KM-EP). This platform will serve as the central hub for all project-related knowledge, including data, documentation, software, and expertise. The chapter details the platform's features, functionalities, and integration with other project systems, emphasizing its role in facilitating collaboration, knowledge sharing, and efficient project execution. It specifically addresses how the KM-EP will support the unique requirements of managing genomic data, analysis results, and associated scientific content.

3.1 Knowledge Management Ecosystem Portal

3.1.1 Platform Description

The GenDAI Knowledge Management Ecosystem will be a centralized, on-premise, secure platform built using a combination of open-source and custom-developed components. The platform will be designed to:

- Centralized Repository: A unified platform for storing and managing all project-related data, documents, software, and knowledge objects.
- Secure Access: Robust security measures, including user authentication and authorization, protect sensitive data and ensure compliance with GDPR and other regulations. This is tightly integrated with the GenDAI Safe component (WP3) and the Attribute-Stream Based Access Control (ASBAC) system (T3.6).
- FAIR Data Principles Support: The platform is designed to support the FAIR principles for data management, facilitating data findability, accessibility, interoperability, and reusability.
- Workflow Integration: Seamless integration with the GenDAI Diagnostics Workflow (WP2) and other project systems enables automated data capture, processing, and analysis.
- Audit Trail: A comprehensive audit trail tracks all user actions and data modifications, ensuring accountability and traceability.
- User-Friendly Interface: An intuitive and responsive interface simplifies data access, management, and collaboration.
- Extensibility: The modular architecture of the KM-EP allows for easy extension and customization to meet evolving project needs.

Technology Stack (Detailed):

Component	Technology	Description
Framework	Symfony Framework	Provides the core structure and functionality of the platform
Database	MySQL	Stores project data, metadata, user information, and other persistent information
Frontend	jQuery, AngularJS, Bootstrap	Used to build the user interface, providing a responsive and interactive experience
Backend	PHP	Implements the application logic and handles interactions between the frontend and the database

This project received funding from the European Union
MSCA SE, Grant Agreement No 101182801

Version Control	Git	Manages source code, documentation, and other project assets
Search	Apache Solr	Provides fast and efficient search capabilities for data, documents, and other resources within the platform. Supports faceted search for enhanced filtering and discovery
Authentication	OpenID, OAuth	Supports standard authentication and authorization protocols for secure user access. Integration with GenDAI Safe and ASBAC
Containerization	Docker	Enables packaging and deployment of platform components as portable, self-sufficient containers. Facilitates scalability and reproducibility
API	RESTful API	Provides a standardized way for external systems and applications to interact with the KM-EP. Will be documented using OpenAPI/Swagger specification

3.1.2 Workflow Integration

The KM-EP is tightly integrated with other key components of the GenDAI platform, particularly the GenDAI Diagnostics Workflow (WP2). This integration enables:

- Automated Data Ingestion: Processed data and metadata from the GenDAI Diagnostics Workflow can be automatically ingested into the KM-EP. This eliminates manual data entry and ensures data consistency.
- Provenance Tracking: The KM-EP captures detailed provenance information from the workflow, tracking the origin, processing steps, and versions of data. This is crucial for reproducibility and data quality assurance.
- Data Access for Analysis: Data stored within the KM-EP can be easily accessed by the GenDAI Discovery (WP4) and GenDAI Interactive Reporting (WP5) components for AI model training, analysis, and visualization.

3.1.3 Audit Trail

The platform will maintain a complete audit trail of all user actions, including:

- Data access (who accessed what data, when).
- Data modifications (who modified what data, when, what changes were made).
- User logins and logouts.
- System events (e.g., software updates, configuration changes).

The audit trail will be used for security monitoring, troubleshooting, and compliance reporting.

3.2 Knowledge Sharing and Collaboration

The GenDAI KM-EP is designed to foster seamless knowledge sharing and collaboration, both internally among project partners and externally with the wider scientific community. Internally, the platform provides a suite of integrated communication and collaboration tools. A key feature enables direct content sharing between users, allowing for efficient exchange of data, documents, and other resources within the project team. Annotation capabilities are integrated, and email notifications keep

all users informed of relevant project activities and updates. The KM-EP supports the creation of custom pages to share project-related information. A competence-based learning is also integrated into the system.

Externally, the KM-EP facilitates the dissemination of project results and knowledge through several key mechanisms. A robust import functionality allows for the ingestion of content from a variety of external sources and formats, consolidating existing knowledge within the central platform. Conversely, an export capability enables the seamless sharing of project outputs with external services and platforms.

Furthermore, the KM-EP incorporates features to advertise content and generate shareable links for social media platforms, promoting wider visibility and engagement with GenDAI's research. The system also includes functions for organizing and classifying knowledge using collaboratively developed taxonomies, making it easier to find and access relevant information. Statistics and reports are generated automatically to facilitate monitoring.

3.3 Document Management

The GenDAI KM-EP incorporates a robust and comprehensive document management system, crucial for organizing, securing, and providing access to the wide range of project-related documents. These documents encompass various types, including technical reports, scientific publications, user manuals, meeting minutes, and other essential project artifacts. To facilitate efficient organization, the KM-EP employs a hierarchical classification, allowing users to create a logical and intuitive arrangement of documents that mirrors the project's structure and workflows.

Beyond simple folder organization, the KM-EP offers powerful search capabilities. Documents can be searched and filtered based on comprehensive metadata, including title, author, keywords, creation date, and other relevant attributes. This metadata-driven search allows users to quickly pinpoint specific documents even within a large and growing collection.

Critically, access to documents within the KM-EP is strictly controlled. Access permissions are managed based on user roles and responsibilities, and these permissions are tightly integrated with the overall User Management System. This ensures that sensitive documents are only accessible to authorized personnel, maintaining confidentiality and compliance with data protection regulations. The access control mechanisms are directly aligned with the data access levels defined in the Data Management Plan (DMP) (Section 2.2.2), ensuring a consistent and unified approach to data and document security across the entire GenDAI platform.

3.4 Training and Support

To ensure the effective utilization of the GenDAI KM-EP and to maximize its benefits for all project partners, a comprehensive training and support program will be implemented. This program recognizes that users will have varying levels of technical expertise and aims to provide accessible and tailored support.

The foundation of the training program is comprehensive documentation. Detailed user manuals and step-by-step tutorials will be created, covering all aspects of the KM-EP's functionality, from basic navigation and document management to advanced features like taxonomy management and workflow integration. These materials will be readily available within the KM-EP itself, providing users with on-demand access to information.

In addition to written documentation, structured training sessions will be conducted. These sessions may be delivered online via webinars or, where feasible, in-person workshops. The training sessions will provide hands-on experience with the KM-EP, allowing users to practice using its features and ask

This project received funding from the European Union
MSCA SE, Grant Agreement No 101182801

questions in a supportive environment. The training curriculum will be tailored to the specific needs and roles of different user groups within the project.

Beyond formal training, ongoing technical support will be a crucial component of the program. A dedicated support channel, such as an email helpdesk, will be established to provide prompt and effective assistance to users. This support channel will be staffed by knowledgeable personnel who can address user questions, troubleshoot technical issues, and provide guidance on using the KM-EP effectively. This support will also encompass ongoing maintenance and updates for the KM-EP platform, ensuring its continued stability, security, and performance througho ut the project lifecycle.

4 Quality & Risk Management Plan

This chapter outlines the Quality & Risk Management Plan (QRM) for the GenDAI project. It defines the procedures and mechanisms for ensuring the quality of project outputs and proactively mitigating potential risks. The chapter covers quality assurance measures, including data validation, software testing, and deliverable review processes. It also presents a detailed risk management framework, encompassing risk identification, assessment, mitigation strategies, and monitoring procedures. This section is essential for ensuring project success and the delivery of high-quality, reliable results.

4.1 Quality Assurance

4.1.1 Quality Standards and Procedures

The GenDAI project will adhere to the following quality standards and procedures:

Area	**Standard/Procedure**
Data Quality	Data will be validated for accuracy, completeness, consistency, and timeliness
Software Quality	Software will be developed using best practices for software engineering, including code reviews, unit testing, and integration testing
Process Quality	Project processes will be documented and followed consistently
Ethical/Legal	All project activities will comply with relevant ethical and legal regulations (see Section 2.5)

4.1.2 Data Validation

Data validation procedures will be implemented to ensure data quality:

Procedure	**Description**
Automated Checks	Automated checks will be performed on data during data entry and processing to identify errors and inconsistencies
Manual Review	Data will be manually reviewed by project partners to identify and correct errors
Data Cleaning	Procedures will be in place to clean and correct data errors

4.1.3 Software Testing

Software testing will be conducted throughout the software development lifecycle:

Testing Type	**Description**
Unit Testing	Individual software components will be tested to ensure they function correctly
Integration Testing	Software components will be tested together to ensure they interact correctly
System Testing	The integrated platform will be tested to ensure it meets functional and non-functional requirements
User Acceptance Testing (UAT)	End-users will test the platform to ensure it meets their needs (WP6)

This project received funding from the European Union
MSCA SE, Grant Agreement No 101182801

4.1.4 Deliverable Review Process

All project deliverables will be reviewed and approved by designated reviewers before submission. The review process will ensure that deliverables:

- Meet the requirements specified in the project proposal.
- Are of high quality (accurate, complete, clear, and well-written).
- Comply with relevant standards and guidelines.

4.1.5 Key Performance Indicators (KPIs)

The following KPIs will be used to measure project progress and success:

KPI	Description
Data Quality	Percentage of data that meets quality standards
Software Quality	Number of defects identified during testing
Project Timeline	Adherence to project milestones and deadlines
Data Sharing	Number of datasets made publicly available
Publications	Number of scientific publications
User Satisfaction	Feedback from end-users on the platform's usability and usefulness

These KPIs will be tracked and reported on regularly.

4.2 Risk Management

4.2.1 Risk Identification

The following risks have been identified as potential threats to the project's success (based on the "Critical Risks" section of the project description):

1. Delay in the implementation of secondments.
2. Project deliverable delayed.
3. Not able to focus critical mass of developers on technical problems.
4. Data and development code loss.
5. Technological show-stopper identified due to software/hardware/cloud-based issues.
6. Disputes on Intellectual Property Ownership.
7. The project may be 'overtaken by events', devaluing its relevance.
8. Secondees leave or change in their personal circumstances.
9. New AI/ML models for classification not performing as well as expected.
10. Withdrawal of participants.

Additional risks may be identified during the project lifecycle.

4.2.2 Risk Assessment

Each identified risk will be assessed in terms of its:

This project received funding from the European Union
MSCA SE, Grant Agreement No 101182801

- Probability: The likelihood of the risk occurring (High, Medium, Low).
- Impact: The potential consequences of the risk occurring (High, Medium, Low).
  A risk matrix will be used to prioritize risks based on their probability and impact.

4.2.3 Risk Mitigation Strategies

Detailed mitigation strategies will be developed for each identified risk. These strategies will aim to:

- Reduce the Probability: Take actions to reduce the likelihood of the risk occurring.
- Reduce the Impact: Take actions to minimize the consequences of the risk if it occurs.
- Contingency Planning: Develop contingency plans to address the risk if it occurs.

Specific mitigation strategies for the identified risks are provided in the Risk Register (Section 4.2.5).

4.2.4 Risk Monitoring and Reporting

Risks will be monitored throughout the project lifecycle. The Risk Register will be updated regularly to reflect the current status of risks and mitigation efforts. Risk status will be reported on at project meetings and in progress reports.

4.2.5 Risk Register

Risk ID	Risk Description	WP(s) Affected	Probability	Impact	Mitigation Measures	Responsible Partner	Status	Contingency Plan
1	Delay in the implementation of secondments	1-6	M	H	1.1 Monitor secondments; 1.2 Action plans for delays; 1.3 Visa support; 1.4 Include in CA; 1.5 Identify backup personnel.	FTK, All	O	Re-allocate tasks; adjust timelines; involve additional personnel.
2	Project deliverable delayed	All	M	H	a) Balance participation; b) Ensure overall picture is clear; c) Regular meetings and management.	All	O	Re-prioritize tasks; allocate additional resources; adjust timelines.
3	Not able to focus critical mass of developers	2-5	L	M	Phasing development; issue tracking and escalation.	SAP, ICT, MTU, UNIBA	O	Re-allocate developers; seek external expertise.

This project received funding from the European Union
MSCA SE, Grant Agreement No 101182801

4	Data and development code loss	All	L	H	Comprehensive data backup and disaster recovery plan (see Section 2.3.2).	SAP, All	O	Restore data from backups; re-create code from documentation/version control.
5	Technological show-stopper	1-6	L	H	Engineer workarounds; leave platform positioned for upgrading.	SAP, ICT	O	Seek alternative technologies; adjust project scope.
6	Disputes on Intellectual Property Ownership	2-5	L	M	IP ownership covered by CA using DESCA template (see Section T8.4 of the proposal description).	All	O	Mediation; legal consultation.
7	Project 'overtaken by events'	6-8	L	M	Monitor external developments; engage with stakeholders; disseminate results regularly.	FTK, ICT	O	Adapt project goals and objectives; seek new collaborations.
8	Secondees leave or change circumstances	1-6	M	M	Substitute with existing staff; consult with REA.	All	O	Re-allocate tasks; adjust timelines.
9	New AI/ML models not performing as expected	4	M	H	Trial and compare two different ML approaches (BiGAMI and NLP-based).	MTU	O	Refine models; explore alternative algorithms; adjust performance expectations.
10	Withdrawal of participants	All	L	H	Re-allocate tasks to other partners; seek replacement partner with REA approval.	FTK, All	O	Adjust project scope and budget; re-negotiate with REA.

Note: L: Low – M: Medium – H: High – O: Open

This project received funding from the European Union
MSCA SE, Grant Agreement No 101182801

4.3 Change Management

4.3.1 Change Request Process

Any changes to the project scope, objectives, timelines, budget, or deliverables must be formally requested through a Change Request Form. The form will include:

1. Description of the Change: A clear and concise explanation of the proposed change.
2. Justification: The reason for the change and its benefits.
3. Impact Assessment: How the change will affect the project.
4. Proposed by and date.
5. Approvals sections.

4.3.2 Impact Assessment

The project coordinator (FTK), in consultation with the relevant partners, will assess the impact of each proposed change on:

- Project Objectives and Scope
- Timelines and Milestones
- Budget
- Resources
- Deliverables
- Risks
- Other Relevant Factors.

4.3.3 Communication and Documentation

All approved changes will be communicated to all project partners and stakeholders through appropriate channels (e.g., email, project meetings, updated project documentation). The Change Request Form, impact assessment, and approval documentation will be stored in the Knowledge Management Ecosystem.

This project received funding from the European Union
MSCA SE, Grant Agreement No 101182801

5 Appendix

Glossary of Terms

- **ASBAC:** Attribute-Stream Based Access Control
- **DMP:** Data Management Plan
- **DPO:** Data Protection Officer
- **ENS:** Entity Name System
- **FAIR:** Findable, Accessible, Interoperable, Reusable
- **GDPR:** General Data Protection Regulation
- **IBD:** Inflammatory Bowel Disease
- **KPI:** Key Performance Indicator
- **OAIS:** Open Archival Information System
- **PID:** Persistent Identifier

This project received funding from the European Union
MSCA SE, Grant Agreement No 101182801