Next Generation Integrated Sensing and Analytical System for Monitoring and Assessing Radiofrequency Electromagnetic Field Exposure and Health

D2.5: NextGEM architectural framework – Final version

Document Summary Information

Start Date	01/07/2022	**Duration**	48 months
Project URL	https://www.nextgem.eu/		
Deliverable	D2.5: NextGEM architectural framework – Final Version		
Work Package	WP2	**Task**	T2.4
Contractual due date	29/02/2024	**Actual submission date**	29/02/2024
Type	Report	**Dissemination Level**	PU-Public
Lead Beneficiary	ICOM	**Deliverable Editor**	Eleni Zarogianni (ICOM)

This project has received funding from the European Union's Horizon Europe research and innovation programme under the Grant Agreement No 101057527

Contributors and Peer Reviewers

Contributors
Eleni Zarogianni, Maria-Eleutheria Vlontzou, Dimitris Laskaratos, Nikolaos Psaromanolakis (ICOM), Nikolaos Petroulakis, Petros Zervoudakis, George Nomikos, Panos Chatziadam (FORTH), Andreas Gavrielides, Kyriaki Psara (eBOS), Myrtill Simko, Mats-Olof Mattsson (SPi), Sam Aerts, John Bolte (THUAS), Stefanos Fafalios, Antonis Paragioudakis (SANL), Maryse Ledent (Sciensano) (SANL)
Peer Reviewers
Eduardo Soudah (CIMNE), Fulvio Schettino (UCAS)

Revision history (including peer-reviewing and quality control)

Version	Issue Date	Changes	Contributor(s)
v0.1	01/11/2023	Table of Contents provided	Eleni Zarogianni (ICOM)
v0.2	08/11/2023	Sections populated with the Task leaders	Eleni Zarogianni (ICOM)
v0.3	09/11/2023	Section defined, assigned, and agreed	Eleni Zarogianni (ICOM)
v0.4	12/12/2023	First contributions	All partners
v0.5	24/01/2024	Integration and harmonization	Eleni Zarogianni (ICOM)
v0.6	31/01/2024	Second contributions and updates	All partners
v0.7	20/02/2024	Complete version ready for peer review	All partners
v0.8	20/02/2024	Peer review	Antonis Paragioudakis (SANL), Fulvio Schettino (UCAS)
v0.9	23/02/2024	Comments addressed from peer review, technical and quality assurance	Mats-Olof Mattsson (SPi), Nicolas Louca (eBOS), Eleni Zarogianni (ICOM)
v1.0	29/02/2024	Final review and submission	Nikolaos Petroulakis (FORTH)

Disclaimer

Copyright message

Table of Contents

Executive Summary 9
1 Introduction 10
1.1 Mapping NextGEM Outputs 10
1.2 Deliverable overview and report structure 11
1.3 Updates from previous Deliverable D2.4 "NextGEM architectural framework – Initial version" .. 11
2 System specifications 12
2.1 Stakeholders requirements and scenarios 12
2.2 Stakeholders and usage scenarios 12
2.2.1 Internal stakeholders 14
2.2.2 External Sources 19
2.2.3 External stakeholders 24
2.3 Technical requirements classification 29
3 Data management and privacy of data 32
3.1 Generic approach to data management 32
3.2 Data management within the NIKH framework 33
4 Architecture overview and components 35
4.1 Envisaged NextGEM Innovation & Knowledge Hub 35
4.2 Conceptual NextGEM Innovation & Knowledge Hub 36
4.3 Components of NextGEM Innovation & Knowledge Hub 36
4.3.1 Infrastructure layer 36
4.3.2 Service layer 37
4.3.3 Application layer 43
4.4 External platform's components 45
4.4.1 Local premises 45
4.4.2 Zenodo 45
4.4.3 EMF-Portal 46
4.4.4 Dataverse 46
4.4.5 Yoda 46
4.4.6 European Health Data Space (EHDS) 47
5 Updated flow component architecture 48
6 Communication diagrams 55
6.1 User initialization workflow 55
6.2 Metadata workflows 55
6.2.1 Upload metadata 56
6.2.2 Search metadata 57

6.3 Data transfer workflows 58
7 Integration strategy 59
7.1 Components integration 59
7.1.1 Containerization 59
7.1.2 Container orchestration 59
7.2 Continuous deployment and integration 60
7.3 GitLab code repository 61
8 Conclusion 62

List of Figures

Figure 1: The NIKH platform with respect to its stakeholders 12
Figure 2: A high-level overview of the NIKH platform 35
Figure 3: System-level view of the NIKH platform. 36
Figure 4: Core IDS components and their interactions [IDSA] 38
Figure 5: The Security Assurance Platform Architecture. 42
Figure 6: The principal steps for risk assessment performed in the RA Tool 44
Figure 7: Risk management activities that will be performed in the NIKH-embedded RA Tool. 44
Figure 8: NIKH Architectural Flow diagram 48
Figure 9: User Initialization workflow. 55
Figure 10: Communication Diagram for uploading metadata records. 56
Figure 11: Communication diagram for searching metadata records. 57
Figure 12: Communication diagram for updating or deleting metadata records. 58
Figure 13: Architecture of a Kubernetes Cluster. 60
Figure 14: The CI/CD Pipeline 60
Figure 15: Snapshot of GitLab. 61

List of Tables

Table 1: Adherence to NextGEM's GA Tasks and Deliverables Descriptions ... 10
Table 2: Usage Scenario template. ... 13
Table 3: Usage Scenario for NextGEM members. ... 14
Table 4: Usage scenario for NextGEM local premises. ... 15
Table 5: Usage Scenario for NextGEM local premises with limited capabilities. ... 17
Table 6: CLUE-H Usage Scenario. ... 18
Table 7: Usage Scenario for Zenodo. ... 20
Table 8: Usage Scenario for EMF-Portal. ... 21
Table 9: Usage Scenario for EMF-Portal. ... 22
Table 10: Usage Scenario for external researchers. ... 25
Table 11: Usage Scenario for Public Authorities and Regulators. ... 26
Table 12: Usage scenario for Standardization Bodies and International Organizations. ... 27
Table 13: Usage Scenario for Citizen Awareness and Risk Communication. ... 28
Table 14: Requirements Classification ... 29
Table 15: NIKH Functional and non-functional requirements. ... 30
Table 16: NIKH Security Requirements. ... 31
Table 17: The Graphical User Interface (GUI) ... 48
Table 18: The Controller ... 49
Table 19: Third Party App for Authentication and authorization ... 50
Table 20: Security Assurance Platform ... 51
Table 21: Log shipper ... 51
Table 22: Data Space Connector ... 52
Table 23: The Risk Assessment tool. ... 52
Table 24: The ERMES Modelling tool ... 53
Table 25: Data Services ... 53
Table 26: External Sources ... 53
Table 27: User roles and admissions. ... 56

Glossary of terms and abbreviations used

Abbreviation / Term	Description
API	Application Program Interface
CI/CD	Continuous Integration/Continuous Development
CLUE-H	European Cluster EMF and Health
CVE	Common Vulnerabilities and Exposures
DMP	Data Management Plan
DOI	Digital Object Identifier
EMF	Electromagnetic Field
ERMES	Electric Regularized Maxwell Equations with Singularities
FAIR (principles)	Findable, Accessible, Interoperable, Reusable
GDPR	General Data Protection Regulation
GUI	Graphical User Interface
IA	Information Architecture
IDSA	International Data Space Association
ICNIRP	International Commission on Non-Ionizing Radiation Protection
JSON	JavaScript Open Notation
NextGEM	Next Generation Integrated Sensing and Analytical System for Monitoring and Assessing Radiofrequency Electromagnetic Field Exposure and Health
NIKH	NextGEM Innovation & Knowledge Hub
OpenAIRE	Open Access Infrastructure for Research in Europe
RA	Risk Assessment
REST	Representation State Transfer
RF	Radio Frequency
SAP	Security Assurance Platform
SLA	Service-Level Agreement

SOP	Standard Operating Procedures
TBD	To Be Defined
UC	Use Case
US	Usage Scenario(s)
UX	User Experience
VM	Virtual Machine
WP	Work Package

Executive Summary

This deliverable reports on the activities of Task 2.4 "System architecture design and technical specifications", which is part of the Work Package (WP) 2 "Requirements Analysis, Specification and Design".

The overall goal of this deliverable is to deliver the final collection of the NextGEM Innovation & Knowledge Hub (NIKH) system requirements and report the final version of the system's functional architecture, highlighting the respective components, and interactions (interfaces) between and among them. The NIKH platform aims to address the challenging objective of creating a hub for Electromagnetic Fields (EMF) and Health, offering a standardized way for European regulatory authorities and the scientific community to store and assess project outcomes and provide access to FAIR data.

Deliverable D2.5 "NextGEM architectural framework – Final version" is based on its predecessor deliverable D2.4 "NextGEM architectural framework – Initial version" and comprises the definition of the following milestones in the architecture design, which also corresponds to the specific objectives of Task 2.4:

1. The translation of stakeholder's needs and requirements from D2.1" EMF value drivers towards stakeholders needs on real case studies", and the elicitation of technical specifications.
2. The definition of data sharing and infrastructure models.
3. The definition of the reference architecture of the NIKH, identifying its main components, their functionalities, and interdependencies.
4. The definition of interdependencies and flow of information among the NIKH components.

The main contribution of the present document to the entire project is to provide the foundational conceptualization for the development of the NIKH platform, by describing its intended use, main functionalities, and the core sub-components that it is comprised of.

1 Introduction

The D2.5 "NextGEM architectural framework – final version" deliverable builds upon the efforts made by Task 2.4 "System architecture design and technical specifications". This deliverable is based on the previous D2.4 "NextGEM architectural framework – initial version" and expands on the description of the components that comprise the NIKH architecture, as well as the interactions among them. Therefore, the structure of D2.5 remains the same, apart from the additions of new sections and subsections and updates in the existing sections.

The first sections of this document cover the description of usage scenarios, aiming to clarify the intended use of the NIKH platform by the relevant stakeholders, as well as the issues related to data management needs and requirements set within the NIKH framework.

The first objective of D2.5 is to translate the user needs and requirements into technical system specifications, considering the overall objectives of the NIKH platform to facilitate the storage and management of information related to EMF measurements, research data and risk assessment, as well as the techniques and components planned to be used by the NIKH platform regarding security and privacy requirements.

The second objective of D2.5 is to expand and finalize the architecture of the NIKH platform with respect to the system requirements and specifications produced earlier in Task 2.4. This activity began with the acquisition of design information about all components brought into the project by technical partners, as it has been already described in D2.4 "NextGEM architectural framework – initial version". A concept architecture describes both already available components and those to be developed and organizes them into a conceptual framework. Thus, this finalized architecture is described in Section 4 of this document and will used as a basis for the development activities in WP6 "Development of NextGEM Innovation and Knowledge Hub".

The third objective of D2.5 is to provide an initial description of the communication diagrams that present the flow of information between the components comprising the NIKH platform and provide a graphical representation of the operations/functionalities offered by NIKH. Lastly, the integration strategy for building a first integrated NIKH prototype is presented, providing input to the development activities in WP6.

1.1 Mapping NextGEM Outputs

The purpose of this section is to map NextGEM's Grant Agreement (GA) commitments, both within the formal Task description and Deliverable, against the project's respective outputs and work performed.

Table 1: Adherence to NextGEM's GA Tasks and Deliverables Descriptions

TASKS	
Task Number & Title	**Respective extract from formal Task Description**
Task 2.4 - System architecture design and technical specifications	The System Architecture of the *NextGEM* platform provides the definition of the data sharing and infrastructure models, involved business processes, and service delivery schemes. The overall architecture embodies EMF data and processes' digital sovereignty, as well as architectural templates and blueprints to encompass the diverse nature of EMF research data collection context and scope, and the EMF stakeholder roles and interactions. The *NextGEM* platform will represent the *NIKH* for EMF data sharing and policymaking support, and will encompass architectural principles of openness and transparency; interoperability, evolvability and integrate-ability of diverse data sources and services; trust, authenticity, and security; research compliance and reproducibility. This task will receive input from the other WP2 tasks and will map requirements for the realization of technical specifications, supporting the complete lifecycle of EMF data in an inclusive and open manner that will initially fit to the requirements posed by existing data collected and knowledge obtained during the lifetime of the project, but will be extensible to provision for future use cases and incremental evolution of the body of knowledge.

DELIVERABLE
Deliverable D2.5: NextGEM architectural framework – Final version (M20) This deliverable produces the final of system specifications and description of the main architectural components, comprising the NIKH platform, specifying usage scenarios and component interactions.

1.2 Deliverable overview and report structure

Based on the objectives and work carried out under Task 2.4, the document starts with the Executive Summary followed by the introduction of the document in Section 1.

Section 2 provides a description of the intended use cases of the NIKH platform, aiming to translate user requirements through usage scenarios into system requirements.

Section 3 presents an overview of the data types that will be handled in the *NIKH* platform and their management in order to ensure FAIR and security requirements.

Section 4 includes the architecture overview of the NIKH platform, detailing the functionalities of its components.

Section 5 provides an updated overview of the NIKH platform architecture, with an emphasis on the sequence of actions among the NIKH components.

Section 6 adds another layer to the architectural flow diagram provided in Section 5, by providing communication diagrams for specific interactions of a user with the NIKH platform.

Section 7 outlines the strategy for seamless development of sub-components and the provisions for ensuring smooth integration.

Finally, Section 8 concludes the deliverable with a summary of outcomes and future actions.

1.3 Updates from previous Deliverable D2.4 "NextGEM architectural framework – Initial version"

As already stated, this deliverable is a continuation and an update on the NIKH platform design, its specifications, and requirements, to what has been detailed in D2.4 "NextGEM architectural framework – initial version". Apart from the updated content in all sections, the main additions to this deliverable can be summarized as follows:

- Addition of Usage Scenarios in Section 2.2 for the Harvard Dataverse and the Yoda repositories (Sections 2.2.2.4 and 2.2.2.5)
- In Section 4, although the main architectural design remains mostly unchanged, the addition of further information on the functionality of its constituent components and the addition of the ERMES Modelling Tool in Section 4.3.4 was done. Also, additional information on the integration of the Dataverse and Yoda repositories as external platform components is also given in Section 4.4 (namely, Sections 4.4.4 and 4.4.5)
- Section 5 is a new addition in this deliverable, detailing an updated flow component architecture of the NIKH platform that aims to emphasize and clarify the communication and relation between NIKH components. This section also provides summary tables for each NIKH component, highlighting provided functionalities, relation to other components, the system requirements each component addresses and its implication on the Usage Scenarios as described in Section 2.
- The addition of Section 6 provides a first-level description of the communication/sequence diagrams for the basic operations/functionalities served by the NIKH platform.

2 System specifications

System specifications allow for the extraction of requirements that further enable the specification of the entire system structure and its constituent components. System specifications are primarily based on the outcome of the user needs, which can be elaborated in the context of Usage Scenarios (US). We provide a detailed descriptions of the users and the associated USs in the following sections.

2.1 Stakeholders requirements and scenarios

The term "system architecture" is used to describe the structure of the system in terms of its substituent components and the interactions between those components. The system is there to serve a series of objectives or functionalities that are met by the specification of the system's subcomponents and their interactions. In the NextGEM project, the specification of system's requirements aims at driving the design of the NIKH architecture, based on a thorough assessment of the stakeholders needs via the use of structured USs, in the context of the identified project goals. In the context of NIKH, the adoption of the US's will ultimately lead to the description of system requirements and the synthesis of formal specifications from system requirements. In the following sections, we summarize the main actors and stakeholders within the NIKH framework and formulate US per each one of them.

2.2 Stakeholders and usage scenarios

The stakeholders of the NIKH platform have been originally described in Deliverable 2.1 "EMF value drivers towards stakeholders needs on real case studies" and include the following, as also depicted in Figure 1:

(1) **Citizens**, meaning all being members of the general public (including subgroups exposed to higher degree due to the nature of their work).
(2) **Telecom Industry**, which include **t**he manufacturers and suppliers of network components and user equipment, as well as the providers of telecommunication services, who introduce the sources of the RF-EMF into society.
(3) **Scientific research institutes** and scientific agencies, which provide the knowledge, including associated uncertainties, necessary to fulfill the role of industry and governments in limiting exposure and protecting people.
(4) **Public authorities and regulators**, who are responsible for protecting the public and workers from potential adverse health effects from exposure to RF-EMF.
(5) **Standardization bodies**. The manufacturers and suppliers of network components and user equipment, as well as the providers of telecommunication services, introduce the sources of the RF-EMF into society.

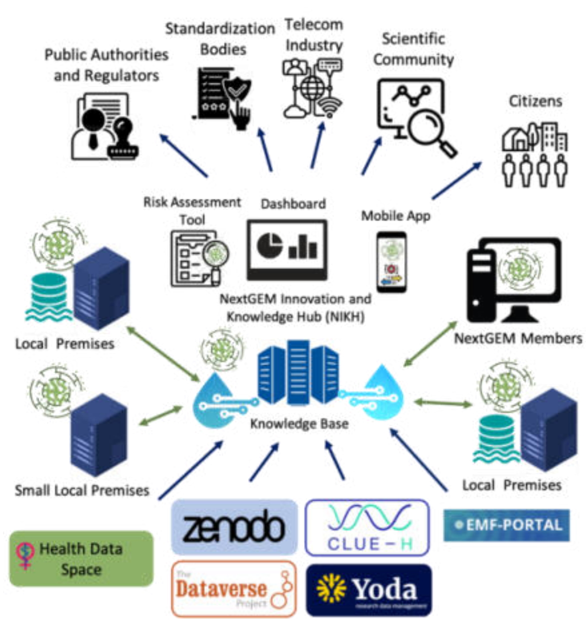


Figure 1: The NIKH platform with respect to its stakeholders

However, apart from these identified stakeholders, we also make the following distinction with regards to the origin of the stakeholders and their relative positioning within NIKH:

1. **Internal stakeholders**: include members of the NextGEM consortium and members of the European Cluster for EMF and Health (CLUE-H) consortium.
2. **External sources**: comprise third-party data providers, such as the EMF-Portal, the Zenodo repository, the Harvard Dataverse platform and the Yoda repository.
3. **External stakeholders**: can be any other type of stakeholder outside the NIKH and the CLUE-H consortia, i.e. public authorities, government bodies, citizens, the media, researchers in other research and academic institutes.

To aid the elicitation of stakeholder requirements, we adopted the approach of constructing **Usage Scenarios (US)**, which capture **end-user requirements** in an agile manner, by describing the role of the stakeholder, and the necessary sequence of actions to be made by the platform in order to achieve an objective.

The US should describe in a practical but precise manner the interactions amongst the various actors of the system that will facilitate the accomplishment of the objectives of the relevant functionalities which are going to be deployed in the system. The basic US-related terminology is presented in Table 2, where each use case has the following uniform structure and the text marked in blue contains the respective US-specific information.

Table 2: Usage Scenario template.

Sub-section	Type	Description
Title	**Usage Scenario X (USX)**	Usage Scenario Title
1	**Scope and Objectives of Usage Scenario**	The scope defines the limits of the usage scenario, and the objectives mention the motive behind the usage scenario
2	**Narrative of Description of Usage Scenario**	Short description is intended to summarize the main idea of use case and A complete description is based on a narrative of the use case from an expert user's point of view, describing what, why, with what expectation, and under what conditions the use case is valid.
3	**Technical Details**	
3.1	**Diagrams of Usage Scenario**	The diagrams show step by step interactions, e.g. Usage Scenario, diagram, sequence diagram, activity diagram, etc.
3.2	**Actors**	In this section, actors which are involved in the usage scenario are listed and described. These can for instance include people, systems, applications, databases, devices, etc.
3.3	**Triggering Event, Preconditions, Assumptions, Success criteria/expected outcome**	Triggering event describes what event(s) trigger(s) this usage scenario (Actor/System/Information / Pre-condition(s) describe(s) what condition(s) should have been met before this usage scenario happens. Assumption may be used to define further general assumptions for this usage scenario. Finally, the success criteria or expected outcome of the usage scenario, if foreseen is mentioned.
3.4	**Information exchange between actors**	It summarizes the information exchanged between two actors with high level concepts.
3.5	**Requirements description**	Requirements coming from high level concepts and technical details written in the usage scenario.

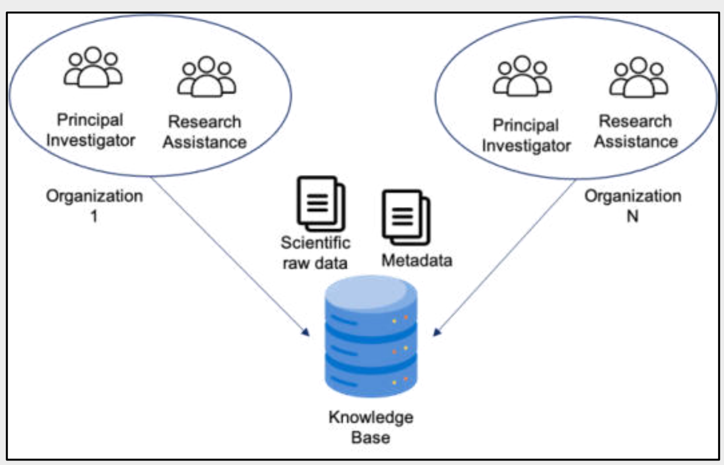
2.2.1 Internal stakeholders

Internal stakeholders are identified as those users, who are part of the NextGEM and the CLUE-H consortia. We identify different role-based access permissions for these users, as defined below.

2.2.1.1 Usage Scenario 1 - NextGEM members

This usage scenario is focused on users within the NextGEM consortium that produce and use data.

Table 3: Usage Scenario for NextGEM members.

Sub-section	Type	Description
Title	**Usage Scenario 1 (US1)**	NextGEM Members
1	**Scope and Objectives of Usage Scenario**	This usage scenario outlines the data produced within an organization and defines the roles of its members, along with the corresponding levels of access permissions for data management.
2	**Narrative of Description of Usage Scenario:**	A **NextGEM Member** is any user who is **registered** and **authenticated** through the NIKH and has the ability to perform certain operations (i.e. edit, add, or delete data), based on his/her access permission rights. The data is divided into three categories: **scientific raw data**, which is collected directly during the experiments and has not yet been processed, **processed data** (either open or closed), and **metadata**, which refers to descriptive information that provides additional context for the scientific raw or processed data. To ensure appropriate data management, this usage scenario identifies different levels of access permissions for the NextGEM Members. All NextGEM members have permission to view metadata records, but the addition, editing and deletion capabilities on metadata and/or processed data (public, or public-after-embargo) will only be permitted to NextGEM users, who are assigned an Administrator role, and are the owners of the data.
3	**Technical Details**	
3.1	**Diagrams of Usage Scenario**	
3.2	**Actors**	**Knowledge Base**: The purpose of Knowledge Base is to provide storage and retrieval capabilities for scientific raw and pre-process data and metadata, as well. **NextGEM Member**: Any authenticated member who belongs to a specific organization from the NextGEM consortium. Members of the two consortia might have distinct roles and permissions assigned to them, such as Principal Investigators (PI) and Research Assistants (RA). As authenticated NextGEM members, we map the above roles and

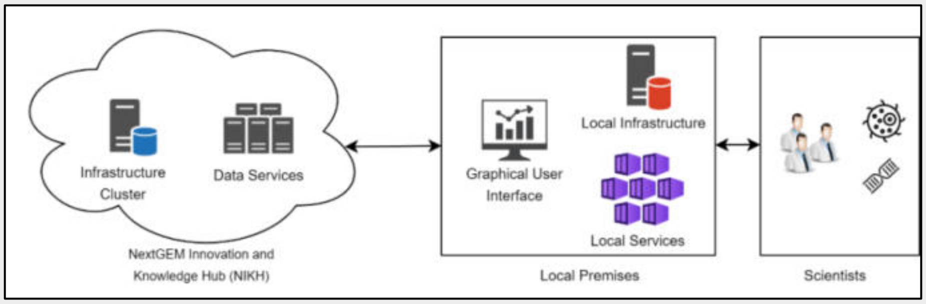
		permissions to a user that acts as an Administrator or a simple user, with the following permissions assigned. ➢ Principal Investigator (PI): The PI has an administrative role, allowing them to perform all actions on any resource within the organization, such as editing, adding and deleting data, and managing access levels to other members. ➢ Research Assistant: The RA's role allows them to view any resource that belongs to the organisation but with limited or no ability to edit or delete them.
3.3	**Triggering Event, Preconditions, Assumptions, Success criteria/expected outcome**	A NextGEM Member can request to edit, add, or delete scientific data or metadata via the appropriate graphical user interface. **Preconditions**: The NextGEM Members must be registered and authenticated through the NIKH platform. **Assumptions**: The NextGEM Members must have appropriate access permissions to edit, add, or delete data. **Success criteria/expected outcome**: The NextGEM Members receive confirmation of the successful completion of their authorized requests, while unauthorized requests are rejected with appropriate error messages.
3.4	**Information exchange between actors**	**NextGEM Member ↔ Knowledge Base**: A NextGEM Member requests access to scientific raw data or metadata which belong to another partner organization. If the NextGEM Member has the permissions to retrieve or modify the resource, a request is sent to Knowledge Base, which retrieves or modifies the requested data.
3.5	**Requirements description**	R1.1: The platform must provide users customizable access levels to resources, allowing for the restriction or permission of access as necessary. This can include the ability to limit access based on factors such as user's role, the type of data being accessed, or the organization to which the user belongs.

2.2.1.2 Usage Scenario 2 - NextGEM local premises

This usage scenario is focused on the NextGEM Local Premises that can interact with a hub to store metadata of experiments.

Table 4: Usage scenario for NextGEM local premises.

Sub-section	Type	Description
Title	**Usage Scenario 2 (US2)**	NextGEM Local Premises
1	**Scope and Objectives of Usage Scenario**	The scope of this usage scenario is restricted to local premises which consist of all necessary entities connected to a communication network, facilitating data exchange between data providers and consumers. The objective of this usage scenario is to **derive requirements related to the semantic interoperability** and collaboration among heterogeneous data storages, achieving secure and sovereign data exchange. Semantic interoperability ensures that data can be comprehended unambiguously by all involved parties across the platform layers.
2	**Narrative of Description of Usage Scenario**	To ensure the security criteria, storage capabilities, and effective data exchange mechanism among all participants in data sharing, it is necessary for every data provider to utilize a specific endpoint, known as Local Premise. This usage scenario mandates the use of this endpoint to guarantee the protection of sensitive data, as well as to facilitate seamless **data transfer and sharing** between all involved parties. By complying with this requirement, all

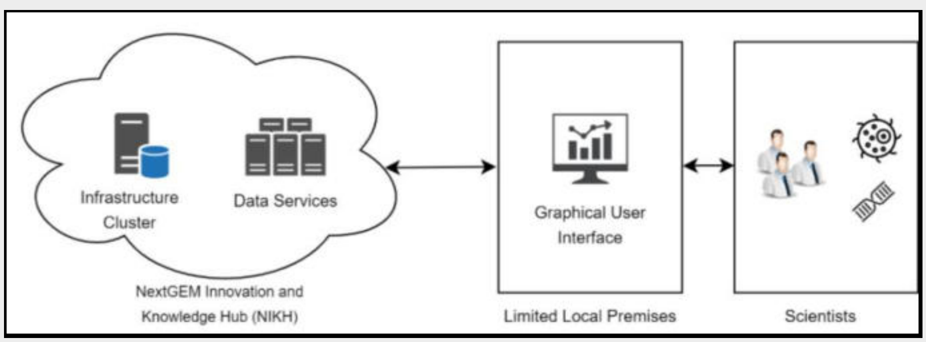
		stakeholders can ensure that their data remains secure and protected, while also fostering efficient collaboration and sharing practices amongst all involved parties.
3	**Technical Details**	
3.1	**Diagrams of Usage Scenario**	
3.2	**Actors**	**Infrastructure Cluster**: Infrastructure Cluster refers to a physical infrastructure including all hardware resources, such as servers, storage, and networking required to support the delivery of cloud services to local premises. **Data Services**: The Data Services include the Knowledge Base, which acts as a centralized repository for data provisioning of the generated scientific knowledge. Its main purpose is to facilitate the storage and retrieval of metadata produced from each local premise. **Local Infrastructure**: It is the server infrastructure as well as the physical resources (e.g., local data storage) running on it owned by a local premise operator/owner. **Local Services**: These services are typically located in a local premise operated by a local premise owner enabling the exchange and sharing of scientific raw data. **Graphical User Interface**: This component must have the capability to effectively manage all experimental laboratory study findings. Furthermore, it must be easily accessible to scientists to enable continuous monitoring of scientific studies and public health data **Scientists**: Any trained personnel, scientists and other professionals, who produce scientific knowledge and contribute to reusable data sets that will be further used in the future.
3.3	**Triggering Event, Preconditions, Assumptions, Success criteria/expected outcome**	**Triggering Events:** Internal stakeholders could have the need to share sensitive scientific raw data within a communication network and/or the need to store metadata generated during experimental processes. **Preconditions**: Preconditions for this usage scenario include the availability of a Local Premise endpoint, which would enable secure and sovereign data exchange, as well as the presence of a stable and secure communication network. **Success criteria/expected outcome**: Success criteria would include the ability to protect sensitive data, facilitate seamless data transfer and sharing, and foster efficient collaboration among stakeholders.
3.4	**Information exchange between actors**	**Scientists ↔ Local Infrastructure ↔ Local Services:** The process of scientific experimentation involves the generation of scientific raw data. This data is stored locally on Local Infrastructure. Local Services enable the exchange and sharing of this data to other stakeholders in a secure and sovereign manner. **Scientists ↔ Infrastructure Cluster ↔ Data Services:** During the scientific experimentation process, scientists generate metadata, which is centrally stored on the Cluster Infrastructure. The Data Services facilitate the storage and retrieval of this metadata by internal and external stakeholders.
3.5	**Requirements description**	R2.1 Local premises and centralized hub infrastructure must be able to store data persistently, with the ability to retrieve and store data at any time. R2.2: Local premises must be able to facilitate the interoperability and semantic integration among different datasets.

		R2.3: The platform's services must be capable to support multiple organizations with their own isolated data and configurations. This requires ensuring adequate isolation between organizations to maintain security and protect sensitive data. R2.4: The Graphical Interface must provide an intuitive design that allows users to easily add, delete, and edit scientific experiments data with minimal effort.

2.2.1.3 Usage Scenario 3 - NextGEM limited local premises

This usage scenario focuses on the NextGEM Small Local Premises that can interact with the NIKH Hub to store data and metadata of conducted experiments. NextGEM limited local premises could include any research institute, or facility that might not have local infrastructure in place to manage their data accordingly.

Table 5: Usage Scenario for NextGEM local premises with limited capabilities.

Sub-section	Type	Description
Title	**Usage Scenario 3 (US3)**	NextGEM Limited Local Premises
1	**Scope and Objectives of Usage Scenario**	The scope of this usage scenario is restricted to limited local premises, which consists of all necessary entities facilitating the transfer of experimental data from a data producer to hub. The objective of this usage scenario is to identify requirements related to both, the user interface that collects and transfers data and the hub that facilitates the storage, share and exchange of these data, achieving secure and sovereign data exchange.
2	**Narrative of Description of Usage Scenario**	This usage scenario requires that limited local premise must be able to collect and transfer their data to a hub, even if its own physical resources are unavailable to support a local premise deployment. Additionally, the limited local premises must maintain continuous interaction with the centralized hub to ensure reliable and uninterrupted data transfer.
3	**Technical Details**	
3.1	**Diagrams of Usage Scenario**	
3.2	**Actors**	**Infrastructure Cluster**: Centralized infrastructure refers to a physical infrastructure including all hardware resources, such as servers, storage, and networking required to support the delivery of cloud services to limited local premises. **Data Services**: The knowledge database is a centralized repository for data storage and provisioning of the generated scientific knowledge. Its main purpose is to facilitate the storage and retrieval of both scientific raw data and metadata produced from each limited local premise. This allows the limited premises to conduct experiments and collect their data without concern about storage limitation. **Graphical User Interface**: This component must be capable of managing all the research results for the experimental laboratory studies. More importantly, the accessibility of such

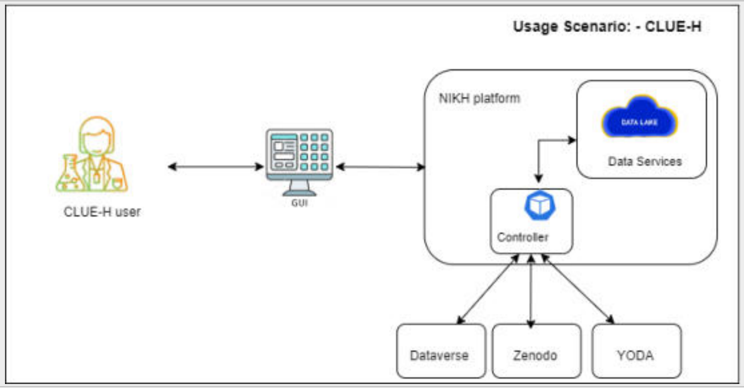
		component must enable scientists to continually monitor scientific studies and public health data. **Scientists**: Any trained personnel, scientists and other professionals which produce scientific knowledge and contribute to reusable data sets that will be further used in the future.
3.3	**Triggering Event, Preconditions, Assumptions, Success criteria/expected outcome**	**Triggering Events:** Internal stakeholders could have the need to store scientific raw, processed data and metadata generated during experimental processes. **Preconditions**: A precondition for this usage scenario is the availability of a platform that can enable persistent data storage. **Success criteria/expected outcome**: The platform's ability to store, share and exchange data among internal and external stakeholders in a secure manner.
3.4	**Information exchange between actors**	**Scientists ↔ Infrastructure Cluster ↔ Data Services:** During the scientific experimentation process, scientists generate scientific raw data, as well as metadata, which is centrally stored in the Infrastructure Cluster. The Data Services facilitate the storage, retrieval and exchange of both scientific raw data and metadata.
3.5	**Requirements description**	R3.1: Centralized Services must be capable of efficiently handling all data (i.e. scientific raw data, processed data and metadata) generated from limited local premises. R3.2: Centralized Services must facilitate seamless interoperability and semantic integration between the data produced by the limited local premises and other parties. R3.3: Graphical Interface must have an intuitive and user-friendly interface that enables scientists to easily navigate and access platform services, including data storage, analysis, and collaboration.

2.2.1.4 Usage Scenario 4 - EMF and Health Cluster (CLUE-H): interconnect with other EU Projects

This usage scenario provides the vision for enabling the exchange of data between the CLUE-H project and NIKH.

Table 6: CLUE-H Usage Scenario.

Sub-section	Type	Description
Title	**Usage Scenario 4 (US4)**	EMF and Health Cluster (CLUE-H) exchange of data among EU projects.
1	**Scope and Objectives of Usage Scenario**	The main scope is to share data and research outputs produced by each of the 4 research projects that are part of the CLUE-H network.
2	**Narrative of Description of Usage Scenario**	The EMF Health Cluster (CLUE-H) is a European Research Cluster established to answer questions related to the exposure and risks of radiofrequency electromagnetic fields, specifically in 5G technologies. The CLUE-H network involves more than 70 European research organizations in four research consortia (ETAIN, GOLIAT, NextGEM, SEAWave), with additional contribution from scientists in the USA, Korea and Japan. The total funding will amount to more than €29 million from Horizon Europe 2021-2027. Each consortium within the CLUE-H establishes its own means for managing, sharing and exchanging data, as per described in its own consortium specific Data Management Plans (DMP) and also elaborated in the "CLUE-H Deliverable – Shared Data Management Plans between cluster partners".
3	**Technical Details**	

3.1	**Diagrams of Usage Scenario**	
3.2	**Actors**	**CLUE-H members** are all members belonging to the following consortia**:** - **NextGEM members**: all members of the NextGEM consortium. NextGEM members will predominately use their local premises, the NIKH platform and also Zenodo for the storage of their (meta) data records, as per described in Usage Scenarios 1, 2 and 3. - **GOLIAT members**: members of the GOLIAT consortium will use Dataverse (http://dataverse.org) to share and archive data. - **SEAWave members**: members of the SEAWave consortium will be primarily using the Zenodo (https://zenodo.org/) repository as a means for data sharing. - **ETAIN members**: members of the ETAIN consortium will be primarily using Yoda (https://www.uu.nl/en/research/yoda). **Application GUI/MobileApp:** The user makes a request and receives the results via the GUI Web interface or the MobileApp. **Controller**: handles all incoming requests from the GUI and relays them to the appropriate components that handles the specific request.
3.3	**Triggering Event, Preconditions, Assumptions, Success criteria/expected outcome**	**Preconditions**: A precondition for this usage scenario is the capability of the platform to make research outputs searchable, findable and accessible through different external sources. **Success criteria/expected outcome**: The NIKH's platform's ability to integrate all the afore-mentioned repositories (Dataverse, Zenodo and Yoda) and act as a central Hub for all EMF-related research outputs.
3.4	**Information exchange between actors**	**CLUE-H members ↔ NIKH platform ↔ External Sources:** A member of the CLUE-H consortia utilizes the NIKH platform to search for interesting research outputs. Through the GUI's search functionalities he/she is able to retrieve appropriate results based on his/her search criteria. The NIKH platform, via the Controller, integrates with all external repositories, such as the Zenodo, Dataverse, etc. and is able to handle the user's request and serve it back to the GUI.
3.5	**Requirements description**	R4.1 The NIKH platform should integrate efficiently with 3rd party repositories. R.4.2 Resources exposed by 3rd party repositories should be findable, accessible, interoperable, reusable and made available to the NIKH platform. R.4.3 The NIKH GUI should be provide user-friendly query/search options.

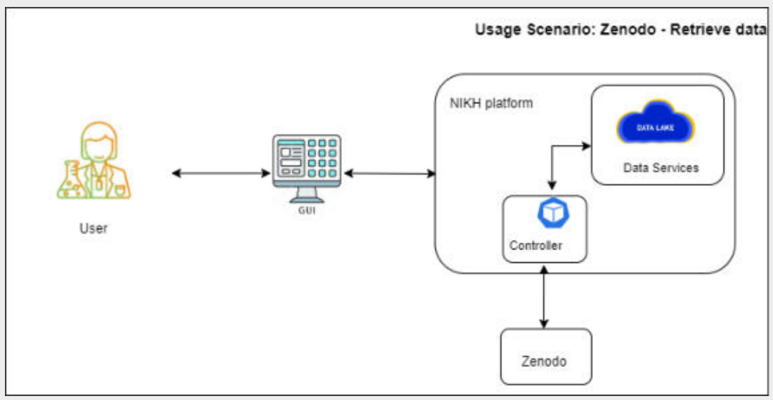
2.2.2 External Sources

Extremal sources are considered any 3rd party repository or data source that will be linked to the NIKH platform. The external sources include the Zenodo, the EMF-Portal repositories and other initiatives such as the EMF and Health Cluster and the European Health Data Space, which collectively handle EMF-related data and publications. In what follows, we present a US per each repository and initiative.

2.2.2.1 Usage Scenario 5 – Zenodo: upload/retrieve public data

This usage scenario showcases the integration of Zenodo with the NIKH platform for uploading and retrieving public available data.

Table 7: Usage Scenario for Zenodo.

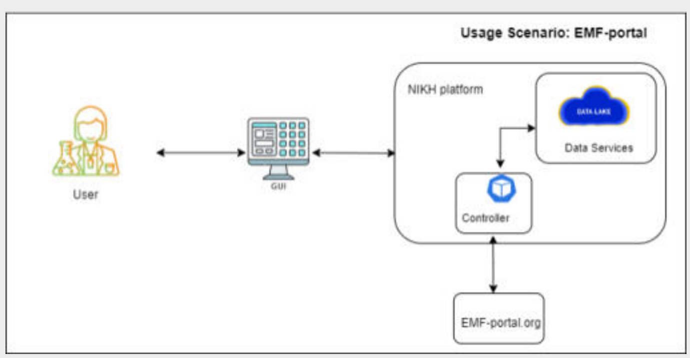
Sub-section	Type	Description
Title	**Usage Scenario 5 (US5)**	Connection to Zenodo
1	**Scope and Objectives of Usage Scenario**	This usage scenario outlines the interconnection of the NIKH with the Zenodo (via its API) to allow NextGEM members to perform key functionalities, based on their permission levels.
2	**Narrative of Description of Usage Scenario**	Zenodo is a general-purpose data repository built on open-source software that accepts all forms of research output from data files to presentation files. The Zenodo REST-API supports all functionalities related to the upload (or deposit), search and download of data.
3	**Technical Details**	
3.1	**Diagrams of Usage Scenario**	
3.2	**Actors**	**NextGEM and non-NextGEM members**: users that are registered and authenticated through the NIKH platform and also those who are not registered. Via the NIKH platform, users can make a request to search, view, upload and download data, via the NIKH platform to Zenodo. **Application GUI/MobileApp:** The user makes a request and receives the results via the GUI Web interface or the MobileApp. **Controller**: handles all incoming requests from the GUI and relays them to the appropriate components that handles the specific request. Here, the Controller is connected to Zenodo's API, and performs the action indicated by the NextGEM member (i.e., search and view data).
3.3	**Triggering Event, Preconditions, Assumptions, Success criteria/expected outcome**	A user makes a request, via the NIKH GUI, to access data based on specific search criteria. The GUI connects to the Controller component of the Central NIKH services, which handles the request and relays it to the Zenodo API. **Success criteria/expected outcome**: The NextGEM Members successfully receives the results of his/her request, upon satisfaction of authorization criteria, unauthorized requests are rejected with appropriate error messages.

3.4	**Information exchange between actors**	➢ **Retrieval: NextGEM/non-NextGEM Member ↔ GUI ↔ Controller ↔ Zenodo**: If a NextGEM member requires data that are located on Zenodo, the controller should retrieve the data and combine it with the data stored in NIKH's repositories and premises. The resulting requested data then is returned to the NextGEM Member ➢ **Storage**: A data storage flow can enable direct storage of data in ZENODO without the need for mediation by the NIKH platform.
3.5	**Requirements description:**	R5.1 The GUI must provide user friendly and attractive interface and an intuitive design that allows users to easily add, delete, and edit data and metadata records as they see fit. R5.2 The user interfaces shall minimize the need for interaction for accessing the required info. In that, the GUI should make information accessible using "least-clicks" rule; a double-menu (top and side) to be used for navigation. R5.3 The Controller should handle the request and relay it to the appropriate API endpoint.

2.2.2.2 Usage Scenario 6 - EMF-Portal.org: link with EMF research knowledge

This usage scenario presents the connection to the EMF-Portal for retrieving EMF-related literature.

Table 8: Usage Scenario for EMF-Portal.

Sub-section	Type	Description
Title	**Usage Scenario 6 (US6)**	Connection to EMF-Portal.org
1	**Scope and Objectives of Usage Scenario**	This usage scenario explores the ability of the user to search and retrieve EMF-related bibliography from the EMF-Portal. We further explore the possibility to connect the EMF-Portal.org with the NIKH platform, in order to provide a holistic and inclusive view of the EMF bibliography to all users.
2	**Narrative of Description of Usage Scenario**	The EMF-Portal is an internet information platform that systematically summarizes scientific research data on the effects of electromagnetic fields (EMF). The current EMF database contains an inventory of more than 38.000 publications and nearly 7.000 summaries of individual scientific studies.
3	**Technical Details**	
3.1	**Diagrams of Usage Scenario**	
3.2	**Actors**	**NextGEM and non-NextGEM members**: users that are registered and authenticated through the NIKH platform and also those who are not registered. Via the NIKH platform, users will be able to search and view data, coming from the EMF-Portal.

		Application GUI/Mobile App: The results of the request will be monitored back to the user via the GUI Web interface or the Mobile App. **Controller:** handles all incoming requests from the GUI and relays them to the appropriate components that handles the specific request.
3.3	**Triggering Event, Preconditions, Assumptions, Success criteria/expected outcome**	A user makes a request to the NIKH platform, via the NIKH GUI, to access bibliography based on specific search criteria. Due to the fact that EMF-portal does not offer any publicly available API to directly retrieve up-to-date data from its repositories, a web scraping mechanism takes over harvesting and extracting content from the EMF-portal in a daily basis. In that way, the NIKH platform after processing the retrieved data can provide the requested information from the EMF-portal to the end-users.
3.4	**Information exchange between actors**	**NextGEM/non-NextGEM Member ↔ GUI ↔ Controller ↔ EMF-Portal**: If a NextGEM member requires data that is possible located on the EMF-Portal, the controller should retrieve the data and combine it with the data stored in NIKH's repositories and premises. The resulting requested data then is returned to the NextGEM Member
3.5	**Requirements description**	R6.1 Identify the application programming interface to integrate the EMF portal literature collection in the NIKH platform. R6.2 The NIKH platform integrates efficiently with the EMF-portal. R6.3 The NIKH controller handles efficiently requests from and to the EMF-portal. R6.4 The GUI provides a seamless interface where the user can select the EMF-portal as the desired 3rd party repository to be searched into.

2.2.2.3 Usage Scenario 7 – European Health Data Space: connect with other data sources

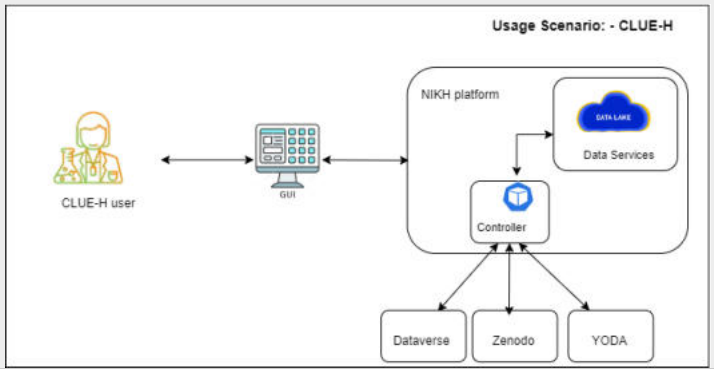
The potential interaction of European Health Data Space under NextGEM is currently under investigation. A detailed description of the EHDS initiative can be found in Section 4.4.4 of this document. At this stage, it is not possible to extract a Usage Scenario, due to the limited functionalities that EHDS currently holds.

2.2.2.4 Usage Scenario 8 – Harvard Dataverse Repository

This usage scenario showcases the integration of Dataverse with the NIKH platform for uploading and retrieving public available data.

Table 9: Usage Scenario for EMF-Portal.

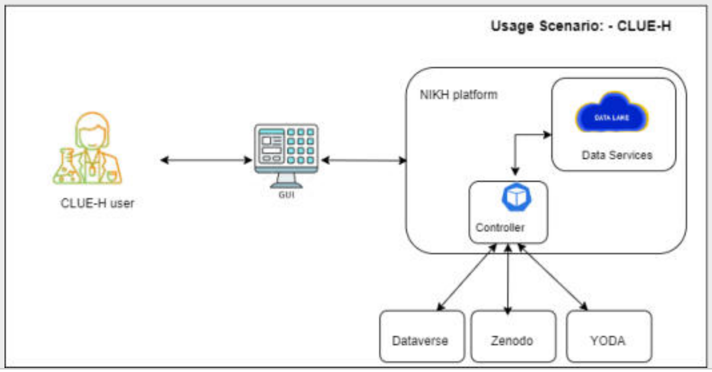
Sub-section	Type	Description
Title	**Usage Scenario (US8)**	Harvard Dataverse.
1	**Scope and Objectives of Usage Scenario**	The main scope is to share data and research outputs produced by each of the 4 research projects that are part of the network under CLUE-H. Harvard Dataverse is the open repository chosen by GOLIAT consortium.
2	**Narrative of Description of Usage Scenario**	The Harvard Dataverse Repository is a free data repository open to all researchers from any discipline, both inside and outside of the Harvard community, where you can share, archive, cite, access, and explore research data. Each individual Dataverse collection is a customizable collection of datasets (or a virtual repository) for organizing, managing, and showcasing datasets. Each individual Dataverse collection is a virtual archive to store your data. Each dataverse collection consists of a dataset and metadata files, documentation and/or code files.

		When data are published Dataverse automatically assigns a Digital Object Identifier (DOI) (a standard data citation) and related metadata is open and findable via search engines, even when the data are restricted. Types of access: open data to the general public, or restricted access and customized access (by defining tailor-made terms of use.)
3	**Technical Details**	
3.1	**Diagrams of Usage Scenario**	
3.2	**Actors**	The actors involved are all stakeholders, both internal and external, as envisioned by NextGEM (as described in Section 2.1) and all other stakeholders involved in the other 3 EU-, research projects.
3.3	**Triggering Event, Preconditions, Assumptions, Success criteria/expected outcome**	A user interacting with the NIKH platform wants to search and retrieve resources from the Dataverse repository.
3.4	**Information exchange between actors**	**CLUE-H user ↔ GUI ↔ Controller ↔ Dataverse**: A user in the CLUE-H consortia searches for assets coming from Dataverse. The controller receives the request and pass it on to Dataverse, which gives back the requested assets, which are then served to the user via the GUI.
3.5	**Requirements description**	R8.1 The NIKH platform integrates efficiently with the Dataverse repository. R8.2 The GUI provides a seamless interface where the user can select the Dataverse as the desired 3rd party repository to be searched into.

2.2.2.5 Usage Scenario 9 – Yoda by Utrecht University

This usage scenario showcases the integration of Yoda with the NIKH platform for uploading and retrieving public available data.

Sub-section	Type	Description
Title	**Usage Scenario (US9)**	Yoda – a research data management service
1	**Scope and Objectives of Usage Scenario**	Yoda is a research data management service that enables researchers from Utrecht University and their partners to securely deposit, share, publish and preserve large amounts of research data during all stages of a research project. Yoda is the repository of choice for the ETAIN consortium, which is part of the CLUE-H described previously.

2	**Narrative of Description of Usage Scenario**	Yoda allows you to store, share, publish and archive your data, and make your data findable. Yoda is also suitable for (privacy) sensitive data. Yoda complies with Utrecht University's Information Security policy for data classified as public, internal use, sensitive or critical. Yoda charges for TB of stored data per month. For research data safety reasons, Yoda automatically replicates your data files to a secondary UU datacenter location
3	**Technical Details**	
3.1	**Diagrams of Usage Scenario**	
3.2	**Actors**	The actors involved are all stakeholders, both internal and external, as envisioned by NextGEM (as described in Section 2.1) and all other stakeholders involved in the other 3 EU-, research projects.
3.3	**Triggering Event, Preconditions, Assumptions, Success criteria/expected outcome**	Yoda is handled by University of Utrecht and its intended use is not to act as an open access repo per se. Instructions on how to use Yoda are given in the official site of the Utrecht University (https://www.uu.nl/en/research/yoda/guide-to-yoda/i-want-to-start-using-yoda). At the time of writing this deliverable, we are still exploring possible ways to integrate with Yoda, in liaison with our partners of the ETAIN consortium.
3.4	**Information exchange between actors**	We envision that integration with Yoda will be handled by the Controller component of the NIKH platform. More information will follow in the upcoming deliverables of WP6.
3.5	**Requirements description**	R9.1 The NIKH platform may be able to integrate efficiently with Yoda.

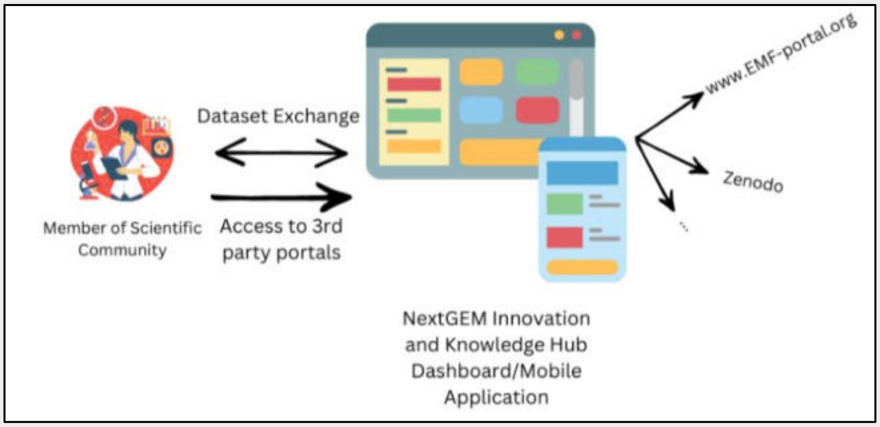
2.2.3 External stakeholders

External stakeholders are considered all stakeholders, who are not currently part of the NextGEM or the CLUE-H consortia, but will take benefit from accessing the NIKH platform, in the context of data and knowledge sharing. These include any member of the scientific community, any regulatory or public authority, standardization bodies and organizations and any member of the general public and/or media agency. Below, we present their distinct usage scenarios.

2.2.3.1 Usage Scenario 10 - Scientific community

The scientific and research results from the experimental laboratory studies (*in vitro, ex vivo, in vivo*) and human studies on EMF exposure and supportive horizontal activities will be made available via the NIKH platform to the scientific community.

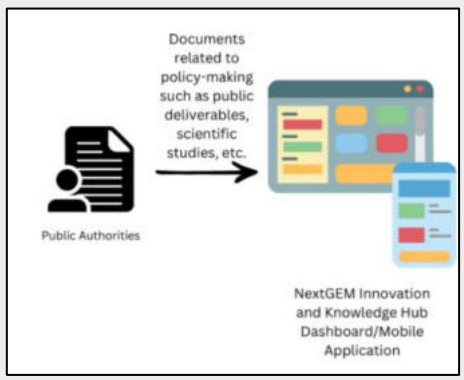
Table 10: Usage Scenario for external researchers.

Sub-section	Type	Description
Title	**Usage Scenario 10 (US10)**	Scientific Community
1	**Scope and Objectives of Usage Scenario**	This usage scenario analyses how the scientific community will benefit from the workflows defined for NIKH and how it will maximise the impact of the platform on the ongoing scientific work in the fields related to NextGEM.
2	**Narrative of Description of Usage Scenario**	The use case is based on the ability of members of the scientific community to be able to login and exchange information through the NextGEM platform. This is a mutual exchange of datasets and other media, as well as access to 3rd party portals and external servers hosted on partners' premises.
3	**Technical Details**	
3.1	**Diagrams of Usage Scenario**	
3.2	**Actors**	In this case the actors are all the members of the external scientific community that are interested in uploading their metadata of any acceptable format and a variety of additional material (datasets, publications, etc.) to the NIKH platform.
3.3	**Triggering Event, Preconditions, Assumptions, Success criteria/expected outcome**	The usage scenario will be marked as successful if members selected from various scientific fields are able to login and use the NextGEM dashboard. In this way, scientists will gain access to the intended NIKH functions quickly and reliably, meaning that they will be able to execute simple functions such as searching and downloading data relevant to their research, as well as uploading their own data.
3.4	**Information exchange between actors**	Via the NIKH platform the involved actors will be able to view metadata and be able to contact each other outside the platform sparking new collaborations inter- and intra-community wise, as well as connect and integrate with other 3rd party repositories.
3.5	**Requirements description**	R10.1: Ability to access the dashboard by inputting username and password. R10.2: Visualisation of the dashboard modules. R10.3: Facilitate access to 3rd parties (e.g., Zenodo) and ability to interact where possible (e.g., upload) through NIKH. R10.4 Search outputs through the metadata catalogue.

2.2.3.2 Usage Scenario 11 - Public authorities and regulators

This usage scenario depicts how the public authorities and regulators can have access to publicly available results of the NIKH platform.

Table 11: Usage Scenario for Public Authorities and Regulators.

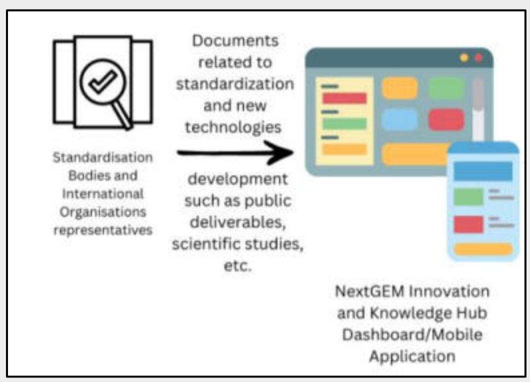
Sub-section	Type	Description
Title	**Usage Scenario 11 (US11)**	Public Authorities and Regulators
1	**Scope and Objectives of Usage Scenario**	This usage scenario analyses how public authorities and regulators can benefit from the workflows defined for NIKH. This scenario can maximise the impact of the platform in creating/updating policies related to the health effects of EMF technologies. Additionally, it can drive innovation forward keeping in perspective how such technologies might affect the well-being of EU citizens.
2	**Narrative of Description of Usage Scenario**	This usage scenario is dedicated for members of public authorities and regulators on a local, national and European level. Through the NIKH platform these members will be informed about relevant documentation such as deliverables, policy briefs, CLUE-H documents and results of the scientific studies conducted by NextGEM and the rest of the CLUE-H projects. This will help the relevant authorities take informed decisions based on scientific results and verified studies conducted by specialists in the relevant fields.
3	**Technical Details**	
3.1	**Diagrams of Usage Scenario**	
3.2	**Actors**	In this case, the actors are all the members of regulatory authorities and the public sector interested in forging or updating new policies concerning EMF-related technologies and require input from scientific studies.
3.3	**Triggering Event, Preconditions, Assumptions, Success criteria/expected outcome**	The usage scenario will be marked as successful if members from the relevant authorities are able to use the NextGEM dashboard and access the intended NIKH functions quickly and reliably. This implies authorities and regulators being able to execute simple functions such as searching and viewing information relevant to policies in which they are involved.
3.4	**Information exchange between actors**	The actors will not be able to interact directly with each other. However, through the NIKH website they will be able to view important information related to regulations and policies which they can then use to contact each other outside the platform sparking new collaborations between regulators/public servants and the scientific community.

3.5	**Requirements description**	R11.1: Intuitive access to all relevant documents with easy-to-understand language. R11.2: Data visualisation with maximum information extraction in a minimal timeframe.

2.2.3.3 Usage Scenario 12 - Standardization bodies and international organizations

This usage scenario demonstrates how the standardization bodies and international organization can interact with the results of NextGEM through the NIKH platform.

Table 12: Usage scenario for Standardization Bodies and International Organizations.

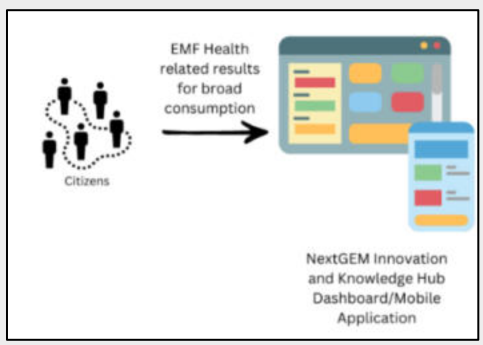
Sub-section	Type	Description
Title	**Usage Scenario 12 (US12)**	Standardization Bodies and International Organizations
1	**Scope and Objectives of Usage Scenario**	This usage scenario analyses how relevant standardization bodies and international organisations can benefit from the workflows defined for NIKH platform. This will enable them to maximise the impact of developing new standards concerning health effects of EMF technologies and how this can drive innovation forward in developing new technologies relevant to EMF radiation.
2	**Narrative of Description of Usage Scenario**	This usage scenario is dedicated for members of standardisation bodies and international organisations. Such members will be able to be informed through the NextGEM platform about relevant documentation such as deliverables, CLUE-H documents and results of the scientific studies conducted by NextGEM. This will help these organisations to develop new standards based on scientific results and verified studies conducted by specialists in the relevant fields.
3	**Technical Details**	
3.1	**Diagrams of Usage Scenario**	
3.2	**Actors**	In this case the actors are all the members of standardisation bodies and international organisations that are interested in updating or developing new standards concerning EMF-related technologies and thus require input from scientific studies.
3.3	**Triggering Event, Preconditions, Assumptions, Success criteria/expected outcome**	This usage scenario will be marked as successful if members from the relevant standardisation bodies and international organisations are able to use the NextGEM dashboard and access the intended NIKH's functions quickly and reliably. This also entails being able to execute simple functions such as searching and viewing information related to standards in which they are involved.

3.4	**Information exchange between actors**	The actors will not be able to interact directly with each other. However, through the NIKH platform they will be able to view information related to standardisations and thus be able to contact each other outside the platform sparking new collaborations between standardisation bodies/international organisations.
	3.5 Requirements description	R12.1: Visualisation of relevant data.

2.2.3.4 Usage Scenario 13 - Citizen awareness and risk communication

Finally, the last usage scenario is focused on how the citizen awareness and risk communication can be improved through the NIKH.

Table 13: Usage Scenario for Citizen Awareness and Risk Communication.

Sub-section	Type	Description
Title	**Usage Scenario 13 (US13)**	Citizen Awareness and Risk Communication
1	**Scope and Objectives of Usage Scenario:**	This usage scenario analyses how EU citizens will benefit from the workflows defined for NIKH. Specifically, through the NIKH platform citizens can take evidence-based decisions concerning their living conditions and therefore improve their quality of life.
2	**Narrative of Description of Usage Scenario:**	This usage scenario is based on EU citizens being able to view information through the NextGEM platform concerning the potential effects of EMF exposure from technologies that are used in everyday life.
3	**Technical Details**	
3.1	**Diagrams of Usage Scenario**	
3.2	**Actors**	In this case the actors are all EU citizens that want to be informed about the latest research on health effects from EMF radiation exposure that is sourced through modern technologies.
3.3	**Triggering Event, Preconditions, Assumptions, Success criteria/expected outcome**	The usage scenario will be marked as successful if EU citizens are able to use the NextGEM website, without being registered, and access to the intended NIKH's functions quickly and reliably. The citizens should be able to execute simple functions such as searching and viewing information relevant to quality-of-life parameters.
3.4	**Information exchange between actors**	The actors will not be able to interact directly with each other. However, through the NIKH platform they will be able to view information and be able to contact each other outside the platform sparking new collaborations inter- and intra-community wise.

3.5	**Requirements description**	R13.1: Visualisation of relevant data.

2.3 Technical requirements classification

System requirements are usually divided into two main categories:

- **Functional** requirements, which are defined by the data, the involved stakeholders and the interactions between them, thus implying what the system will do.
- **Non-functional** requirements, which additionally define constraints and rules in the system operation. These can be further divided into subcategories as defined in Table 13.

The requirements identified in this document are classified based on their focus and scope as described below:

Table 14: Requirements Classification

Classification	Description
Security and Privacy	The NIKH platform will combine standardized technologies to provide lightweight and reliable mechanisms for end-to-end authentication and authorization for its entities (users, services, premises, external parties, etc.)
Semantic Interoperability	The NIKH platform will integrate an extremely large amount of heterogeneous data. These data need to be consistently and formally represented and managed. The platform will be able to exchange data with unambiguous, shared meaning.
Scalability	The platform will be scalable to support a large number of users and local premises, as well as the increased demands to store massive volumes of data.
Availability	The platform should be resilient to continuous operate without interruption when one or more of its components fail, on both cloud services and premises.
Quality of Service (QoS)	QoS should be as high as possible in terms of interaction with the cloud services, such as performance, availability and reliability. (e.g., low latency of response in third party requests, users and premises operate continuously even if one or more components fail)
Multi-tenancy	The platform should provide isolated access to its resources for multiple organizations (tenants). Maintaining adequate isolation between tenants is crucial for ensuring security and protecting sensitive data.
Usability	Usability is referred to User eXperience (UX), user-friendliness, intuitiveness, accessibility, and user satisfaction
Requirement Type (F/NF)	Functional requirements (F) define what the system should do, while non-functional requirements (NF) define how well it should do it. These requirements coming from specific use cases, the involved stakeholders and the interactions between them.

The above classification criteria are then applied to different use cases identified in this document as shown in the following table. In this last version of the System Design and Architecture, we revisited all requirement attributes and finalized them ensuring their description is precise and the associated set of uses cases have been correctly identified.

Table 15: NIKH Functional and non-functional requirements.

No.	Req.-ID	Security and Privacy	Interoperability	Scalability	Fault tolerance	Quality of Service (QoS)	Multi-tenancy	Usability	Functional/ Non-Functional
1	R1.1	•					•		F
2	R2.1					•			F
3	R2.2		•						F
4	R2.3	•					•		NF
5	R2.4							•	NF
6	R3.1			•					NF
7	R3.2		•						NF
8	R3.3							•	NF
9	R4.1							•	F
10	R4.2		•					•	F
11	R4.3							•	NF
12	R5.1							•	NF
13	R5.2							•	NF
14	R5.3			•					F
15	R6.1		•					•	F
16	R6.2		•					•	F
17	R6.3		•						F
18	R6.4							•	NF
19	R8.1		•						F
21	R8.2							•	NF
22	R9.1		•						F
23	R10.1	•		•				•	F
24	R10.2			•	•	•		•	F
25	R10.3		•						F
26	R10.4		•					•	F
27	R11.1		•	•				•	NF
28	R11.2			•	•	•	•	•	F
30	R12.1	•		•				•	F
33	R13.1	•		•				•	F

Due to the intricate nature of the NIKH platform, in the context of satisfying various needs related to access and data across various stakeholders (as per described in the Usage Scenarios), we would like to emphasize on explicitly stating Security Requirements that safeguard privacy and security constraints at a user-level, at an infrastructure level and a data-usage level. Specifically, the NIKH platform would ensure the following requirements:

Table 16: NIKH Security Requirements.

No.	Req.-ID	Category	Importance	Description	Applicability	Usage Scenarios
1	S1	Security	MUST	The NIKH platform MUST offer mechanisms for authentication and authorization of each involved actor/entity/stakeholder.	User	US1 to US9
2	S2	Security	MUST	The NIKH platform MUST support role-based access to ensure data security and privacy	User	US1 to US9
3	S3	Security	SHOULD	The NIKH platform SHOULD perform vulnerability assessments of its constituent parts to ensure proper functionality and availability of services	Infrastructure	All USs.
4	S4	Security	SHOULD	The NIKH platform SHOULD perform an assessment of the impact and risk of the detected vulnerabilities.	Infrastructure	All USs.
5	S5	Security	SHOULD	The NIKH platform SHOULD adopt a sovereign and trustworthy process to perform data exchanges and transfers among interested parties.	Data	US1 to US9
6	S6	Security	SHOULD	The NIKH platform SHOULD be able to encompass and enforce policy restrictions related to the usage of data.	Data	US1 to US9

3 Data management and privacy of data

3.1 Generic approach to data management

As already stated above, the NIKH platform will handle a plethora of data, which according to the DMP v1.0 will include:

i. Experimental data (coming from in vivo, ex vivo, in vitro, and human studies)

ii. Open-source computational tools

iii. Device specifications

iv. Standard Operating Procedures (SOPs),

v. Project deliverables and reports

vi. Journal-type data, such as handbooks, literature reviews, results papers, articles etc.

As stated in the DMP-initial version (D1.1: Data management plan, initial version), "the management of data and results will be based on practices and procedures that ensure that used data and results are: a) stored securely and preserved in order to ensure its continuing utility, b) appropriately identifiable, retrievable, and available when needed, and c) kept in a manner that is compliant with legal obligations, including the Data Protection Act 1998 / The General Data Protection Regulation (GDPR) (Regulation (EU) 2016/679) and the Freedom of Information Act 2000 and d) able to be made available to others in line with appropriate ethical, data sharing and open access principles, especially when the data underpins published research."

Moreover, as "data and outcomes shared in open domain can be very beneficial to society, NextGEM needs (and aims) to carefully balance openness and protection of sensitive data. As stated in the Guidelines on FAIR Data Management data should be "as open as possible and as close as necessary". All partners and especially data providers that participate in the consortium should comply with all applicable data protection or similar laws regulating the processing of any personal data." The data will either be raw or processed; the latter is particularly important in the case of outputs related to personal data; concerned datasets will be anonymized.

As an EU funded project, NextGEM will participate in the Open Research Data Pilot of the European Commission (openAIRE), which enables open access and reuse of research data. *NextGEM* will follow an openness approach with regards to its generated/gathered data and research outputs.

However, the project will employ the policy of providing data as open as possible and as closed as necessary following rules, regulations and suggestions that protect the provision of data to make results replicable and in turn protect the interests of the members of the NextGEM consortium.

As stated in the Grant agreement of NextGEM:

The beneficiaries must manage the digital research data generated in the action ('data') responsibly, in line with the FAIR principles and by taking all of the following actions:

- as soon as possible and within the deadlines set out in the DMP, ensure open access — via the repository — to the deposited data, under the latest available version of the Creative Commons Attribution International Public License (CC BY) or Creative Commons Public Domain Dedication (CC 0) or a license with equivalent rights, following the principle 'as open as possible as closed as necessary', unless providing open access would in particular:
  - be against the beneficiary's legitimate interests, including regarding commercial exploitation, or
  - be contrary to any other constraints, in particular the EU competitive interests or the beneficiary's obligations under this Agreement; if open access is not provided (to some or all data), this must be justified in the DMP.

Research outputs as specified above will be either:

- Set as openly available for public use. Research outputs will be uploaded on an **open repository** with relevant metadata. Zenodo has been chosen as the repository currently used by most partners in the consortium. that will be used for the project datasets has to be defined by the consortium.

- Set as open after publication, or restricted under a specific framework that will be documented.
- Set as confidential due to internal regulations and/or legal reasons that data providers ought to comply with. Confidential research outputs will either be shared within the consortium or become accessible in-house after a proper agreement is signed. Data produced by the project that cannot be published as open data or open after publication will be listed and an explanation will be given of the reasons that restricted or prohibit open access. Nevertheless, information about the restricted data along with the related metadata will be published in the open repository and NIKH.

As a matter of complying with the FAIR process, **metadata** will be assigned to each research outputs. European metadata standards will be used to make data interoperable with other data sets of similar type. Proprietary data formats will be avoided as much as possible. Research outputs will be made available in platform and software independent data formats.

3.2 Data management within the NIKH framework

The DMP provides with sufficient information pertinent to the types of data and their format within the NIKH framework (Section 3 of the DMP v1.0). However, there are other issues yet to be defined and elaborated with regards to the following issues:

- ***Where will data be stored?***
  Research organizations and institutes that are part of the NextGEM consortium maintain their own procedures and policies for the secure storage of their data and research outputs. Local premises of research organizations and institutes constitute what has been described in Section 2 as "NextGEM Local Premises", comprising of different levels of available infrastructure and possibly different needs with regards to storage.
  It is up to the organization's administration to decide what data and how data are stored locally and define access permissions to their data accordingly. In all cases, **the NextGEM Local Premises are the hosts of the data.**
  Usually data that will be hosted in local premises will be either raw data (e.g. pseudonymized/non anonymized data, very large files, etc.) or data with specific legal restriction as set by the partners. Other outputs set as sensitive, but accessuble to the NextGEM partners could also be stored on the NextGEM repository for use at the level of the consortium. Other outputs could be shared either on open repositories (e.g. Zenodo, GitHub, etc.) or NIKH. This will have to be set based on different characteristics of the outputs, such as the privacy level, the size, their uses in specific tasks (e.g. for risk assessment), needs for long term preservation, etc.

- ***Which data will be stored in the core NIKH platform?***
  Based on the discussions that continually take place among project partners, the current status quo defines that predominantly metadata will be also made available centrally, at the core part of the NIKH platform.
  In specific cases, where research institutes do not have sufficient local infrastructure, the storage of raw or processed data might be considered for central storage. Discussions are still underway among SPi, THUAS, SANL, CNR, SC and UZH to define what constitutes **raw** data and what the processing needs of these data entail.

- ***Are there metadata standards in place?***
  Relevant metadata standards will be used to make metadata interoperable. A metadata template has already been drafted and is included in D1.1: Data management plan, initial version. The template is mapped to Data Catalog Vocabulary (DCAT https://www.w3.org/TR/vocab-dcat-3/) and Schema.org (needed for the indexation of web pages) metadata standards. A metadata record can be delivered in many formats, such as a pdf, html document, an exchange format as a .ttl, RDF-XML, JSON-LD, etc.
  Discussions concerning the format and the structure of the metadata template are still ongoing.

- ***Will the data in the central data storage be available after the completion of the project for research purposes?***
  Conceptually, it is also of paramount importance that data in the central data storage are available after the completion of the task, given the vision of the NIKH platform to become a central hub of knowledge for all

EMF-related research. Technically, it would be possible to maintain the data generated within NextGEM beyond the project's lifetime for an unlimited time, thus fully supporting future studies and research activities. Discussions are, however, still on-going within the NextGEM consortium on the benefits and risks of maintaining data generated within NextGEM beyond the project's lifetime.

- **Who has access to the data?**
  In short, the user to whom data belongs and other users that he/she authorises. Within NextGEM, the Principal Investigator and/or leading scientists are responsible for identifying who has access to experimental and research data and assign access permissions and roles accordingly. The core NIKH platform will develop the capabilities to assign roles with regards to data, as well, in terms of authorization mechanisms.

4 Architecture overview and components

4.1 Envisaged NextGEM Innovation & Knowledge Hub

The NextGEM's overall vision is to ensure EU citizens' safety when employing existing and future EMF-based telecommunication technologies. In this context, NextGEM will provide a framework for the generation of health-relevant scientific knowledge and data on new scenarios of exposure to EMF in multiple frequency bands, and develop and validate tools for evidence-based risk assessment. On the technical side, NextGEM will also develop the NIKH platform, which will handle information on EMF measurements, research data and risk assessment.

Specifically, the overarching aim of the **NIKH platform** is to monitor, store, share and access EMF exposure and biological data along with SOPs for lab experiments, and engineering solutions to maintain compliance with safety standards, minimize exposure levels in set environments and contexts, and increase citizen's awareness on EMF information and research. To enable validation and exploitation, a benchmarked proof-of-concept reference platform will be developed, which will encompass the NIKH features and will be open and extendable, to ignite the incremental evolution of NIKH into a central Innovation and knowledge base for EMF data.

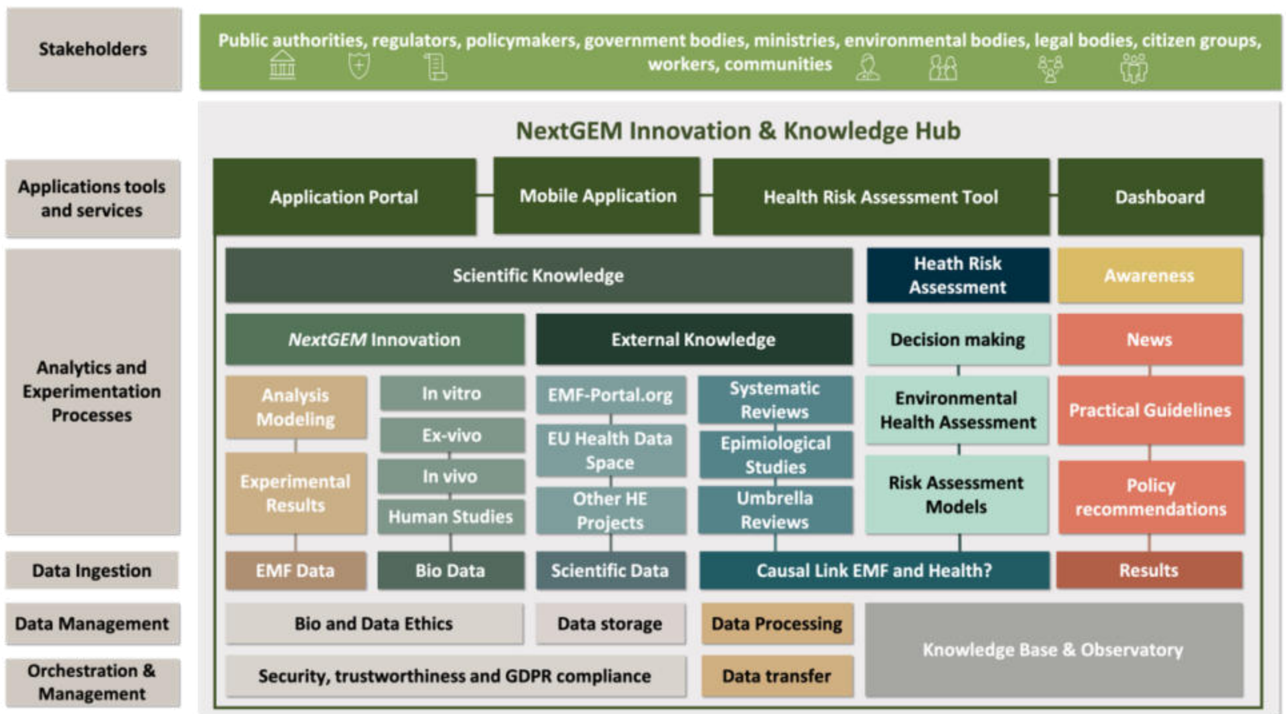


Figure 2: A high-level overview of the NIKH platform

The identification of system specifications for the development of the *NIKH* platform comprises of the following main pillars encompassing all aspects of the platform (as also depicted in Figure 2):

i) The design of the platform to store the innovations and research outputs produced from experimental and human studies performed within the *NextGEM* project.

ii) The inclusion of external scientific knowledge, either from online research portals, such as the Zenodo or the EMF-Portal.org or through synergies with projects funded under other clusters and pillars of Horizon Europe, or other EU programs (i.e., the CLUE-H Working Group).

iii) The development of applications tools and services, which will:
   i. Ensure security and GDPR compliance, such as the Security Assurance Platform (SAP) component.
   ii. Provide a Risk Assessment tool for the assessment of health impact of exposures.

iv) The identification of all relevant stakeholders, with a shared interest to interact with the NIKH platform and the development of an interface or front-end dashboard to facilitate such interaction with the core NIKH platform.

The validation of the developed platform and mechanisms will be performed in **three** different **case studies** i) Potential effects of indoor levels of RF exposure of vulnerable people on reproduction and development, ii) optimised outdoor urban planning and 5Gdesign architecture and investigations for public awareness of cancer-related health hazards and iii) health effects of exposure to mmWave EMF in indoor and outdoor environments.

Finally, the NIKH platform will be based on an open-source framework integrating detailed codebook and metadata for biological data in a micro-services architecture containing the computational tools required for the envisioned functionalities and allowing flexibility and interoperability to embody diverse business processes and scenarios.

4.2 Conceptual NextGEM Innovation & Knowledge Hub

The generic NIKH conceptual architecture shown in Figure 3 depicts its main organizational layers and functionalities to be supported in order to satisfy various stakeholders' needs.

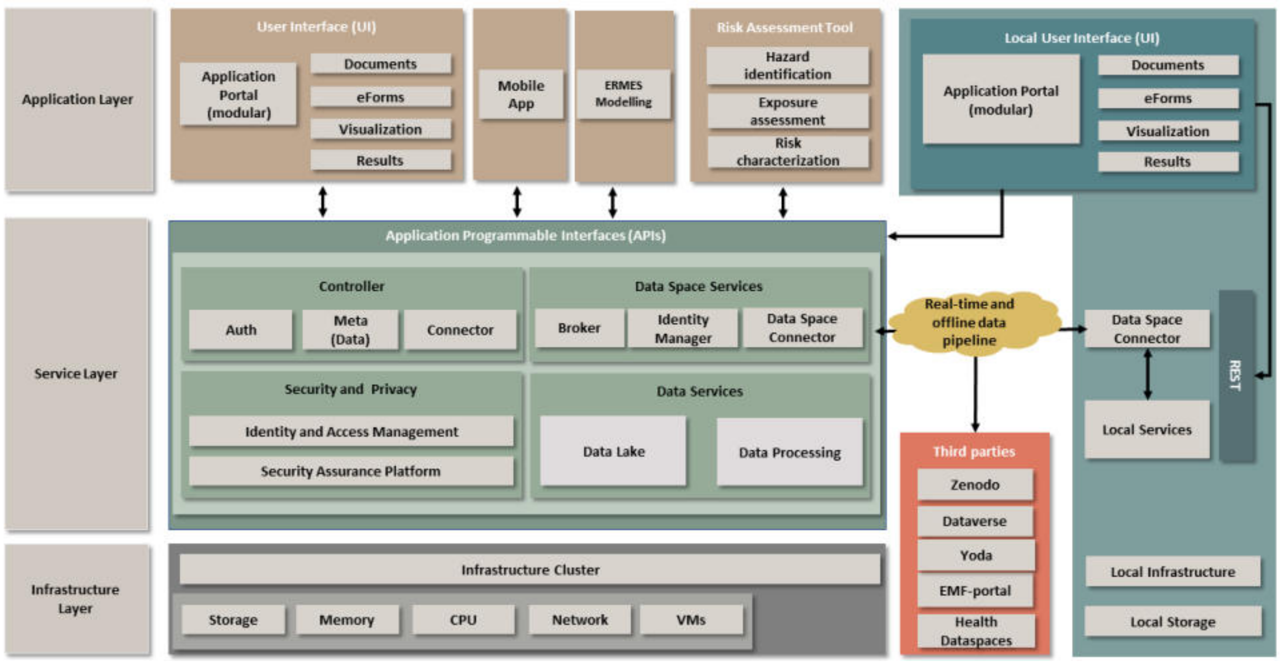


Figure 3: System-level view of the NIKH platform.

The concept in this figure differentiates between Infrastructure, Service and Application layers and the Local premises, which comprise their own local network, infrastructure, storage, and application capabilities. Each of these layers comprises different components and sub-components, with distinct functionalities and capabilities.

4.3 Components of NextGEM Innovation & Knowledge Hub

The main components of the NIKH architecture along with their description are detailed in the following sections.

4.3.1 Infrastructure layer

The infrastructure layer is the nuts and bolts of the NIKH platform and consists of a physical infrastructure and a virtual infrastructure layer.

4.3.1.1 Physical infrastructure

The physical infrastructure of the NIKH platform has been designed to meet its demanding technical requirements. The hardware elements, including the main servers, storage devices, internal and external network connections have been selected to provide high-performance computing and storage capabilities, while also ensuring availability and reliability, for application storage, providing adequate processing power and storage capacity to support the platform's various applications and services. The internal network within NextGEM has been designed with security and resilience aspects, incorporating multiple layers of security, utilizing a range of technologies and protocols to ensure multiple layers of protection. These include firewalls, intrusion detection and prevention systems, and data encryption technologies, as well as regular security audits and vulnerability assessments. Finally, the physical infrastructure of NextGEM will be deployed at FORTH's main data center, which provides a secure and well-equipped environment for the platform's hardware components. The deployment will involve meticulous planning and testing to ensure that hardware is installed and configured appropriately. Once deployed, the physical infrastructure will support the backbone of the NIKH platform, providing the essential technical framework to support its diverse range of applications and services.

4.3.1.2 Virtual infrastructure

The virtual infrastructure within NextGEM is a critical layer that enables the platform to support multiple virtual machines (VMs) simultaneously. This layer is based on a hypervisor, which is a software layer that enables the creation and management of virtual machines. Fully isolated VMs should be created enabling to ensure downtime, redundancy, incremental backups, and built-in security using integrated firewalls, security groups and IP sets.

In particular, VMs can be used to host a wide range of components, including application portal, risk assessment tool and security assurance tool. In addition to these components, VMs can also facilitate the management of containerized services by hosting a virtualized such as Kubernetes cluster. Kubernetes can abstract away the underlying virtual infrastructure, providing a consistent set of tools for managing the application lifecycle regardless of the underlying infrastructure.

4.3.2 Service layer

The service layer is the agglomeration of all applications and services that will serve the functionalities of the NIKH platform. We further, conceptually and pragmatically, distinguish between different components and, thus, services that will handle access control and privacy requirements, data ingestion, storage and data exchange and efficient communication and workflow between and among all substituent components. In the following sections, we describe each component individually, detailing initial requirements and functionalities.

4.3.2.1 Controller

The Controller of the NIKH Platform is the component responsible for orchestrating and managing received requests. The Controller contains all of the platform's interfaces, through which a user interacts with the platform, offering the REST API endpoints. Every request is handled by the Controller, which performs the corresponding actions and generates the appropriate response.

Regarding its architecture, Controller implements various software components which are responsible for managing the various heterogeneous system resources that are available through other services on the platform. The main functionalities of the Controller are summarized below.

- The Controller provides access to the available resources based on the authorization/authentication token of the user (operated by the **Authentication** subcomponent).
- When a non-certified user wants to access available data from a NextGEM member, the Controller handles this request by initiating the communication between the NIKH Platform Connector and the respective data provider's Connector. Additionally, the Controller provides access to the open-access data of the NIKH Platform or third-party repositories (operated by the **Authentication** subcomponent).
- The Controller provides the available data catalogue (by storing the respective metadata in its database) to the NIKH UI (**Metadata** subcomponent).
- The Controller provides functionalities for performing operations on the records comprising the metadata catalogue (specifically for the creation, edit and deletion of records- **Metadata** subcomponent)
- When a user wants to upload a new data resource to the NIKH Platform, the Controller can automate the procedure by making the appropriate API calls to the user's Data space Connector (operated by the **Connector** sub-component).
- The Controller can additionally ensure that the data resource is registered, as an available resource, at the Broker (**Connector** subcomponent).
- Finally, the Controller enables access to existing knowledge bases such as Zenodo and other 3rd sources to provide an integration to the latest public information and results (**Connector** subcomponent).

4.3.2.2 Data Space Services

The core purpose of Data Spaces is to enable controlled, sovereign, and secure exchange and sharing of data between stakeholders. The International Data Spaces Association (IDSA) introduces a decentralized data sharing architectural concept, in which data physically remain at their source and are only transferred to another participant when data exchange requests are instantiated. In International Data Spaces (IDS), data sovereignty and trust are established, since each participant can attach usage restrictions to their data and monitor data transactions through continuous monitoring. Additionally, security is ensured, through the identity evaluation of each participant. Furthermore, IDS offers metadata storage, as well as metadata-query functionalities that enable participants to search for the appropriate data sources and request access to the respective data.

NIKH aims to encompass the architectural principles of trust, data sovereignty, transparency; interoperability, and integrate-ability of diverse data sources and services. In this scope, International Data Spaces are a suitable solution, since their technical components have been developed based on those same values and implement the technical environment where those principles can be realized.

4.3.2.2.1 Data Space Ecosystem

A data space ecosystem consists of Connectors that link data providers and data consumers and enable them to securely exchange data assets, while preserving sovereignty and control over their data.

As per the December 2023 Report [1], the Data Connector term has been introduced to distinguish connectors based on the IDS Standard (IDS Connectors) from connectors interacting on a protocol agnostic standard (Data Connectors), which is the foundation for interoperable data spaces. Towards this protocol agnostic standard, the Dataspace Protocol is currently being developed by the IDSA, comprising a set of specifications designed to facilitate interoperable data sharing between entities governed by usage control and based on Web technologies.

A central component of the data space ecosystem is the Data Space Connector, which enables the data exchange among participants. Each participant is represented by a connector, which allows the registration of offered data resources, along with the metadata that describe them. A data connector executes the **complete data exchange process** and ensures the bidirectional communication and enforcement of usage policies upon data, which ultimately leads to the bilateral agreement between involved parties.

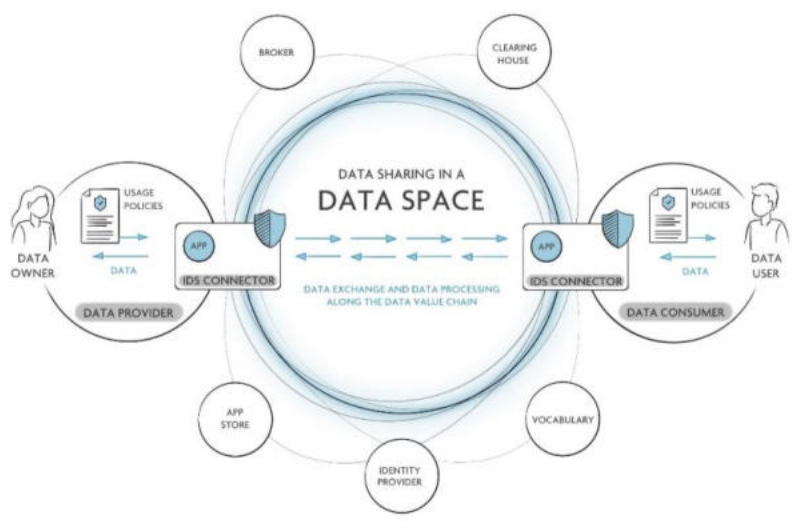


Figure 4: Core IDS components and their interactions [IDSA]

In the context of NIKH, the chosen implementation of the Dataspace Connector is the Eclipse Dataspace Connector

1 https://internationaldataspaces.org/wp-content/uploads/dlm_uploads/IDSA-Data-Connector-Report-92-No-8-September-2023.pdf

[2]. It is an open-source, IDS-compliant, modular and highly extensible framework, in which the control plane and data plane are separated. This solution can be extended to support different types of data storage options, identity and access management and can be adapted to cover various requirements, for example, in terms of access rights policies.

The Eclipse Dataspace Connector allows users to create metadata records of their data and link them to specific usage policies through contract definitions. The access policies supported by the NextGEM Members connectors are mapped to the defined policies within NIKH, such as

- Public,
- Public after embargo,
- Restricted access and
- Closed.

The respective contract definitions can be negotiated with other dataspace members, who request access to these data. If the negotiation is accepted by both parties, a contract is agreed upon and the data asset can be securely exchanged.

Description of Component Functionality within the NIKH platform:

- In the NIKH architecture, Connectors are located either at the user's local premises for NextGEM members (e.g., scientific research institutes), offering data resources or within the Data Spaces Services of the NIKH Service Layer for NextGEM members, non-certified users and Mobile App users (e.g., citizens) requesting access to specific data or intending to upload data to the NIKH Platform central storage.
    - For a data provider, the Connector enables the registration of the metadata pertinent to an offered data resource, along with its usage policy at the Connector and subsequently at the NIKH platform.
    - For a data consumer, the Connector supports requests to the NIKH platform metadata manager, in order for the data consumer to obtain information about available data providers and the description (metadata) of their data. It also enables the contract negotiation (acceptance of usage policy on the consumer's part), in order for the consumer obtain access to the requested data.

Interaction with other components:

- Controller: Each Connectors' access information is stored at the NIKH Platform. Additionally, the metadata of the available data assets are handled by the NIKH Controller and can be queried and retrieved by the users in the form of a metadata catalog.

4.3.2.3 Data services

Data services are responsible for the ingestion, storage and any processing required for the data. Below, we give an introduction to the Data Lake architecture and its data processing capabilities. A detailed elaboration of the architecture, tools and applications that will support the development of the Data Lake will be given in the deliverable D2.5 "NextGEM architectural framework – final version".

4.3.2.3.1 Data lake

The Data Lake is the central storage component of the NIKH Platform. A data lake is a centralized repository that allows the storage of all structured and unstructured data at any scale and in any format. Since it can store structured, semi-structured, or unstructured data, data can be kept in a more flexible format for future use. When storing data, a data lake associates them with identifiers and metadata tags for faster retrieval. Given the large amounts of data in a data lake, tagging objects with metadata is important to make them accessible in the future. The structure of a data lake's software varies, but the objective is to make data easy to locate and use.

The structure of the data (or so-called schema) is not defined when data is captured, but at the point in time when the data scientist wants to work with the data (schema-on-the-read). The goal behind schema-on-read is flexibility (allowing

[2] https://projects.eclipse.org/projects/technology.edc

multiple schemas per data set) and deferring the work to create a schema and run data transformations until it becomes necessary. Different types of analytics on the data like SQL queries, big data analytics, full-text search, real-time analytics, and machine learning can be used to uncover insights. The goal is to derive value from the collected data by exploring the data, training statistical models, and transforming the data to be used in services and applications. Through crawling, cataloguing, and indexing of the data, there is the ability to understand what data is in the data lake. Finally, data must be secured to ensure the data assets are protected.

Within the NIKH framework, all metadata of the NextGEM members' data resources will be centrally stored in the Data Lake. As outlined in Section 3.2, scientific and/or experimental data might also be uploaded in raw format, necessitating a basic processing step. In this regard, the Data processing sub-component will be responsible for providing data processing services, such as filtering, transformation, etc. Data lakes can reside on-premises or in a cloud environment. At the time of drafting this deliverable, we have investigated solutions to the Data Storage that will be reported in Deliverables that will be part of WP6 activities.

4.3.2.3.2 Data processing

Within the NIKH platform, there is a variety of operations and processing steps that might need to be executed, in order to extract useful information from the data stored in the data lake. These could include data profiling, parsing, filtering, format conversion, transformation, schema extraction, data cleansing and de-duplication, data masking, data mining, model training and fitting. In the beginning, these operations might need to be executed in an explorative manner. However, once some way of extracting meaningful information has been developed, a process needs to be established to automatically repeat the steps on new or updated data sets. Therefore, these operations are often arranged in a data flow, where – conceptually – the operations are applied to chunks or streams of the data and intermediate results are stored and forwarded to the next operation.

4.3.2.4 Security and privacy

The Security and Privacy layer consists of two distinct components dedicated to the identification and access management on the one hand and the component security assurance on the other.

4.3.2.4.1 Identity and access management

The Identity and Access Management component of the NIKH Platform is the component responsible for the registration and login of NextGEM Members to the NIKH Platform. It is also responsible for ensuring the correct permissions are applicable for the different types of user roles within the NIKH Platform.

The core functionalities of the authentication/authorization component are:

- **User registration**: Upon registration the user adds details pertaining to his/her information (i.e. name, address, email, etc.), and also information regarding the organization he/she belongs to.
- **User Log in**: The user logins it using his/her credentials (such as username, password, etc.)

This Identification and Access Management module is handled by a third party authentication service, named Keycloak[1]. The Controller, which has been described above, acts as a middle interface between Keycloak [3] and the rest of the NIKH platform's components and services.

The core conditions set by this component are:

- Authentication/authorization should allow a user to log in to the NIKH Platform only with valid credentials (username & password).
- Authentication/authorization should only allow a user to register when being part of an organization comprising a NextGEM member.
- Authentication/authorization should allow a user to have administration rights only when logging in with admin credentials.

[31] https://www.keycloak.org/

4.3.2.4.2 Security Assurance Platform

Security and privacy assurance is crucial for any organization that uses IT infrastructures since it provides an overview of its security posture. Security posture refers to the overall security readiness and resilience of an organization's IT environment. It is a measure of an organization's ability to detect, prevent, and respond to security threats and incidents. A healthy security posture means that an organization has implemented a comprehensive set of security controls and practices to protect its IT assets and data and has the ability to quickly identify and respond to security incidents.

A security posture assessment involves evaluating an organization's IT environment, including its network infrastructure, applications, endpoints, data, and users, to identify vulnerabilities and areas of risk. This assessment can include reviewing security policies and procedures, evaluating the effectiveness of security controls and technologies, and conducting penetration testing and vulnerability assessments. The results of such assessments can help organizations understand their current security state and identify areas for improvement. By identifying vulnerabilities and weaknesses, organizations can take steps to address these issues before they are exploited by attackers. This can help to mitigate the risk of data breaches, cyber-attacks, and other security incidents. Additionally, a security posture assessment can help organizations to comply with regulatory requirements and industry standards.

Sphynx's Security & Privacy Assurance Platform (SAP) [4] supports the continuous assessment of the security and privacy posture of a target infrastructure, and the management of risks arising from this posture. To do so, the platform provides **key capabilities** including modelling enterprise system and business assets, whose security and privacy must continually be assessed; multimodal security and privacy threat and vulnerability assessments of these assets (static vulnerability analysis and dynamic testing, continuous multi-layer runtime monitoring, hybrid analysis); assessments of the impact and risk of detected vulnerabilities.

The operation of the security and privacy assurance platform is based upon an **asset model** that includes the description of several types of assets of the target organization, such as software, hardware, network and data assets. Moreover, the asset model includes a description of the relations between those assets (e.g., an asset controls another asset) allowing the identification of vulnerabilities and risk propagation within lethargy organization assets. For example, there is the case of a server (hardware asset) that provides a database service (software asset) that maintains user data (data asset). We then have the following relations: server **contains** database and database **contains** data. In the case that a vulnerability is identified in the server the induced risk propagates to the contained assets (database and data). The continuous monitoring of the security posture is leveraged by using specific **event log shippers** that Collect information from the target infrastructure and create events upon them. More specifically, lightweight shippers (namely Beats) collect logs and events which are subsequently stored in ElasticSearch [5]. The collected events are then accessed through queries on the ElasticSearch database. This allows the platform's monitoring capability to analyze these events, providing assessments based on specific model driven monitoring rules (assessment criteria). For example, a service-level agreement (SLA) may dictate that the availability of a service should be at a specific percentage and an availability monitor rule can be used to provide evidence that the SLA is not violated.

The goal of a security posture assessment is to identify gaps and weaknesses in an organization's security posture, and to develop a plan to address these issues. This plan may involve implementing additional security controls, improving security policies and procedures, increasing security awareness training for employees, and establishing incident response plans.

The function of the **identification of vulnerabilities** is based on the National Institute of Standards and Technology (NIST) [6] as part of its Common Vulnerabilities and Exposures (CVE) program and the globally-accessible knowledge base of adversary tactics and techniques leveraging real-world observations MITRE ATT&CK [7]. Vulnerability

[4] Sphynx Analytics, The Security and Privacy Platform https://www.sphynx.ch/products/#assurance-platform

[5] https://www.elastic.co

[6] National Institute of Standards and Technology, https://www.nist.gov/

[7] MITRE ATT&CK, https://attack.mitre.org/

assessments can be static based on defined assets within the asset model and dynamic where assets that are not included in the asset model can be identified.

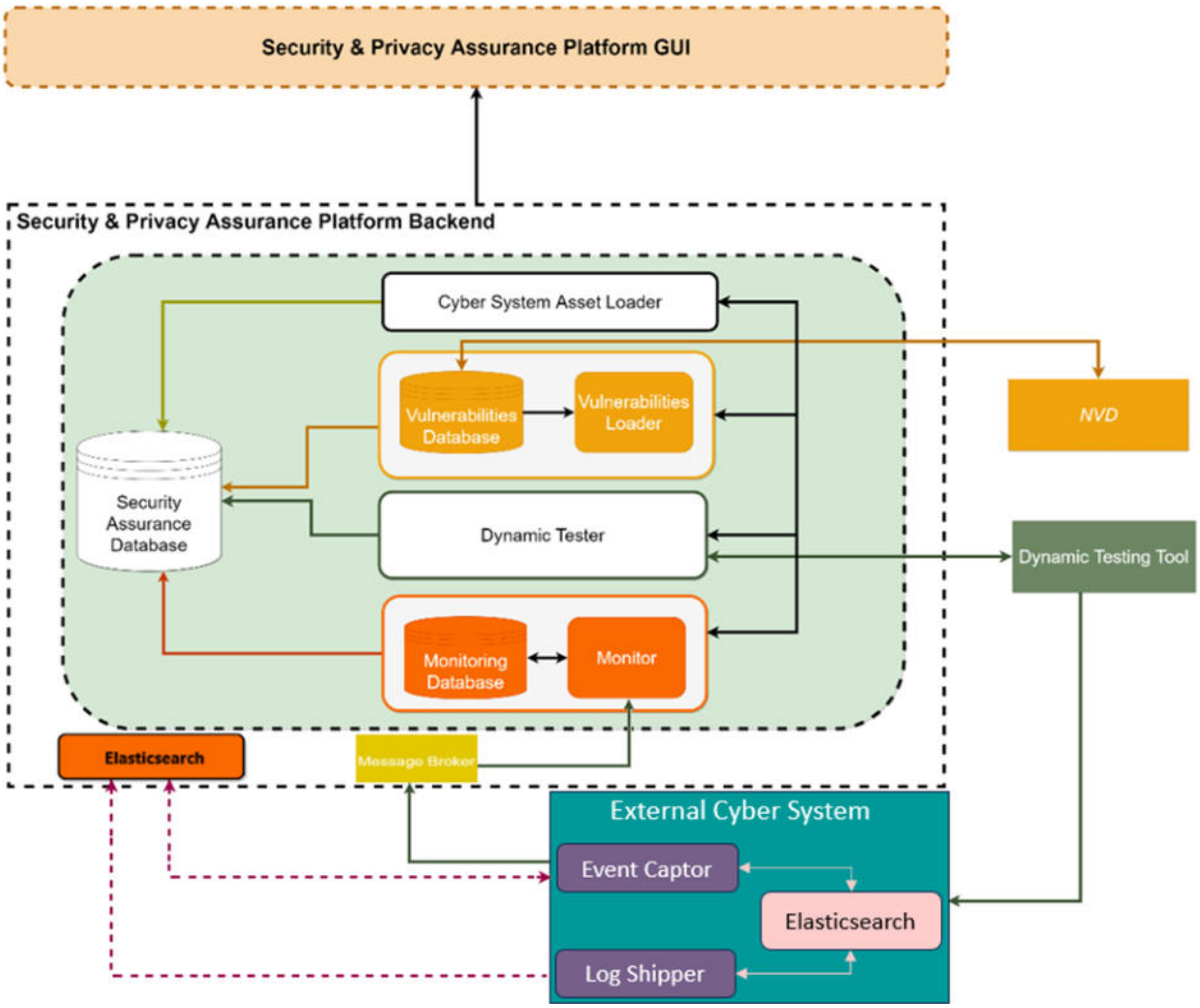


Figure 5: The Security Assurance Platform Architecture.

For the purposes of providing security assurance for the NIKH platform, the SA platform will be deployed on a dedicated VM. This strategy is crucial for **continuous monitoring**, ensuring uninterrupted surveillance even if certain components of the NIKH platform become unresponsive.

During the implementation of other NIKH components, specifications for the assets used will be catalogued (e.g. operating systems, database types and versions, etc), so that this information is tested against the NVD, for identifying possible vulnerability issues with said assets, as part of SPA's vulnerability analysis capabilities.

To facilitate the continuous gathering of relevant information, we will deploy log shippers on each NIKH component. The deployment of the log shippers is essential to ensure continuous monitoring of the diverse components constituting NIKH. For example, one of the most critical assessments is the service availability assessment. It involves checking if a service responds within a specific timeframe. If a service is inaccessible for a period greater than the specified timeframe, a violation flag is triggered indicating the availability issue of the assessed service or component. For this rule to function effectively, the asset needs to be either publicly exposed to the internet or internally visible to a log shipper. To illustrate, external sources like Zenodo, Harvard Dataverse Repository, and Yoda can be directly assessable since they are already accessible online. However, this is not the case for data stored within limited local premises. Evaluating their availability, involves deploying a log shipper on the connector component which will serve as the access point to the aforementioned premises, enabling effective assessment. Overall, it is crucial to monitor the availability of all components. This includes the authentication, metadata and connector manager components, data services, GUI, as well as the risk assessment tool. To ensure constant monitoring each component will be paired with a log shipper. This setup further serves the purpose of utilising other monitoring capabilities as well.

A detailed definition of the assets targeted and their relations populating the NIKH asset model as well as the assessments and monitoring rules applied, will be thoroughly provided on deliverable "D6.3 - Trustworthy data management and compliance with ethics and legal aspects - Initial report" and deliverable D6.8 which is the final report of the prior.

4.3.3 Application layer

The application layer is the layer in closest proximity to the user of the NIKH platform and thus consists of applications and services that the user can directly interface with. Below, we describe the web and mobile applications (portals) of the NIKH platform, the Risk Assessment tool and the ERMES modelling tool..

4.3.3.1 Application Portal and Graphical User Interface (GUI)

The main objective of the application portal and Graphical User Interface (GUI) is to provide a unique access point to the NextGEM ecosystem (internal and external stakeholders) and a visual way for users to interact with the various components of NIKH including the Knowledge Base and 3rd party portals with the provision of role-based access control. A key User Interface (UI) feature contains the adaptivity provided through a configurable NextGEM dashboard for the identified target groups including public authorities, industry players and the scientific community. For the design and implementation of NIKH's Graphical User Interface (GUI), the requirements and technical specifications are followed to develop and produce the detailed low-level design of the front-end dashboard, and to be validated by the end-users. The designs will be described in detail along with the accompanied functionality in deliverable "D6.1 - Development of tools, dashboard and mobile app - First cycle". The appropriate type of UI components required are selected in the implementation of the dashboard. A graphical web-based front-end dashboard will be developed to facilitate the visualisation of data and results from EMF measurements, analytical models and biological investigations, documents, together with practical guidelines, tools and applications, thus supporting public authorities and regulators with scientific evidence to implement exposure directives and improving risk assessment, management and communication. The dashboard will rely on open standards to gather data, adhering to the Findable, Accessible, Interoperable, Reusable (FAIR) principles, from the NextGEM platform thus enabling the interfacing between the back-end system and the dashboard for easy and secure communication and data transfer. This can fulfil the need raised by various national and international organizations since EMF data will be publicly accessible by the NextGEM dashboard. The organizations in question may serve the general public or be connected to special interests such as trade unions, environmental organizations, trade and industry organizations, standard-setting bodies etc.

The NextGEM application portal and GUI are currently in the design phase, which includes conducting research to identify main preferences and expectations, the information architecture, visual design and interaction designs phases that collectively will provide users with an ease of use based on design principles making it intuitive and user friendly. Furthermore, the GUI will be designed to improve productivity by providing easy access to commonly used features and functions as well as improve user engagement by providing visual feedback making the designed applications more engaging and interactive. Another aim of the GUI is to standardize the user experience across different software applications and systems.

4.3.3.2 Mobile App

The external stakeholders of the NIKH platform, including citizens, authorities, and scientific communities are also offered a mobile-friendly experience which extends the capabilities of the NIKH platform. The mobile app allows the NextGEM consortium to further promote EMF related knowledge by facilitating a mobile app offering a user-friendly approach to risk appreciation and protection. This provides a visually intuitive interface, visualizing meaningful EMF-related insights to users such as practical exposure guidelines, location-based information, and dynamic interaction with EMF initiatives. The application is essentially a web-based human machine-interface simplifying complex information for easy comprehension by diverse user groups enriching the app's content with different perspectives and insights.

The app incorporates mobile-specific features that extend the core functionality while catering to the unique needs of mobile users. The mobile app optimizes the existing GUI for diverse screen sizes adapting the GUI for mobile responsiveness and on-the-go usability aiming for an optimized user experience. This approach ensures that the NextGEM Mobile App extends the capabilities of NIKH on a broader audience but also exposes further knowledge related to electromagnetic fields. The development process includes testing and refinement ensuring accessibility to a broader audience. Functional and non-functional requirements are derived from Usage Scenarios and modelled workflows, guiding the creation of a detailed mobile app representation that aligns with the NIKH platform's goals.

4.3.3.3 Risk assessment tool

The process of **Risk governance** applies the principles of good governance to the **identification, assessment, management, and communication of risks**. An ambition of NextGEM is to provide a tool embedded in NIKH, which can be used for risk assessment by different competent stakeholders, but also to apply results from the assessment process in other risk governance activities. The **NextGEM Risk Assessment (RA) Tool** will be designed to perform human health risk assessment of mobile communication technologies with reference to specific health outcomes. Results from the RA Tool will feed into stakeholders needs for the different risk governance activities. The different steps in the RA process are depicted in Fig. 5 below.

Figure 6: The principal steps for risk assessment performed in the RA Tool

Results from the RA Tool will feed into stakeholders needs for the different risk management activities, in addition to risk assessment that includes mitigation and communication of risk (see Fig. 6).

Figure 7: Risk management activities that will be performed in the NIKH-embedded RA Tool.

There are different types of data, from different sources, that will be used by the RA tool. This includes:

- Processed data and metadata from experimental activities (modelling, exposure assessment, various types of biological studies, meta-analyses) within NextGEM,
- Processed data and metadata from corresponding activities within the CLUE-H cluster,
- Metadata from published scientific articles, exposure monitoring campaigns, health registers, and other "grey papers" when appropriate,
- Exposure guidelines from transnational, national, and regional authorities; recommendations from authorities and special interest groups.

Furthermore, the RA Tool will be equipped with appropriate tools for statistical analyses and Risk of Bias assessment. The outputs from the RA Tool will be directed to stakeholders, such as scientists from different fields, competent authorities and regulators, insurers, industry representatives, risk communicators and managers in both public and private settings, special interest group representatives, media professionals and the general public. The outputs will include:

- Qualitative evidence-based risk assessment,
- Quantitative evidence-based risk assessment,
- Exposure guidelines´ compliance,
- Knowledge gap identification,
- Risk communication and risk mitigation advice.

Within the NIKH, the RA Tool will extract data, either from the NIKH platform (via Data Connectors) or external sources, such as Zenodo, EMF-Portal, Yoda, Dataverse, etc. in order to perform its assessment activities, and deposit data from the performed assessments and other types of actions there. In addition, the RA Tool will be interacting with the Graphical User Interfaces/Application Portal to ascertain user-specific access and utilization.

4.3.3.4 ERMES Modelling tool

ERMES (Electric Regularized Maxwell Equations with Singularities) is an open-source tool provided by the NextGEM partner CIMNE and is primarily used to visualize EMF-related simulation outputs. The novelty of this computational tool rests on its formulation, in that it provides a simplified version of the weighted regularized Maxwell equation method that as a result can provide the capacity of solving problems in the low (quasi-static) and high frequency regimens in EMF-related simulation scenarios.

ERMES is available for both Windows and Linux operating systems, featuring a user-friendly interface integrated into the pre- and post-processor GiD, which can handle geometrical modelling, data input, meshing, and result visualization.

In the NIKH platform, the ERMES modelling tool will provide the results of different electromagnetic simulations and scenarios as an add-on functionality. The ERMES tool through its capability for visualizing numerical simulations can complement the understanding of the real impacts of EMF technologies, and thus foster transparency, accessibility and collaboration. A detailed description of the internal processes and functionalities of the ERMES tool can be found in Deliverable D3.2, "Development of modelling approaches to assess internal and external exposure – Initial version”.

4.4 External platform’s components

Linking to the external sources’ distinction presented in Section 2 of this report, we distinguish between local premises, external EMF-related initiatives and repositories and present them from an architectural, component-level standpoint in order to calibrate integration with the NIKH platform.

4.4.1 Local premises

Local premises are defined as an interconnected network of local facilities, with capabilities for persistent, local data storage capabilities. Local premises will, thus, comprise a server infrastructure along with the physical resources (e.g., data storage), running on it and owned by a local premise operator/owner.

4.4.2 Zenodo

Zenodo is a general-purpose open repository developed under the European OpenAIRE program, developed by the European Organization for Nuclear Research (CERN), but is open to researchers from outside the EU. It allows researchers from all disciplines to deposit research papers, data sets, research software, reports, and any other research related digital artefacts and store them in the CERN Data Center, which provides long-term preservation and the ability for users and stakeholders to access it at a moment’s notice.

Free to upload and free to access, Zenodo makes scientific outputs of all kinds citable, shareable, and discoverable for the long term. Zenodo does not impose any requirements on format, access restrictions or licence. Finally, Zenodo also abides by the FAIR principles and is compliant with the data management requirements of Horizon Europe, the ERC and other EU research and innovation funding programmes.

Zenodo’s core functionalities include:

- Zenodo accepts large files (up to 50GB) without format restriction. The data is assigned a **Digital Object Identifier** (DOI), which means that publishing in a journal is not required in order to obtain a unique and permanent DOI, thus makes the publication more easily citable according to international standards.
- Datasets, documents, and other research materials can be located via the Zenodo **search engine**.
- Other features are very helpful for researchers: flexible licensing and accessibility, long-term storage, and automatic integration in reporting for European Commission-funded projects via OpenAIRE.
- Zenodo offers the possibility to house closed and restricted content, so that artefacts can be captured and stored safely whilst the research is ongoing, such that nothing is missing when they are openly shared later in the research workflow.
- The metadata of each record are indexed and searchable directly in Zenodo's search engine immediately after upload.
- Metadata are publicly accessible and licensed under public domain. No authorization is ever necessary to retrieve it.
- Data and metadata will be retained for the lifetime of the repository. This is currently the lifetime of the host laboratory CERN, which has an experimental programme defined for the next 20 years at least.
- The web interface is supplemented by a rich API which allows third -party tools and services to use Zenodo as a backend in their workflow. The Zenodo REST API currently supports:
    - Deposit — upload and publishing of research outputs (identical to functionality available in the user interface).
    - Records — search published records.
    - Files — download/upload of files.
- Zenodo uses JSON Schema as internal representation of metadata and offers export to other popular formats.

The Metadata Manager service of the Controller provides an endpoint that accepts the ID of a public record in Zenodo in string format. The service then uses the API of Zenodo, described previously, to retrieve the record metadata in a JSON schema. If the API call is successful, the response is the JSON schema itself, otherwise an error message is returned.

4.4.3 EMF-Portal

The EMF-Portal [8] is a project of the working group of the Institute for Occupational, Social and Environmental Medicine of the Uniklinik RWTH Aachen University , which systematically summarizes scientific research data on the effects of EMF. All information is made available in both English and German. The core of the EMF-Portal is an extensive literature database with an inventory of more than **38,000** publications and over **7,000** summaries of individual scientific studies on the effects of electromagnetic fields.

The main scope of the EMF-Portal is to collect scientific information on the effects of electromagnetic fields and present the data in an easily understandable form and to make it available to all interested parties. The publications collected in the EMF-Portal must have been written by scientists (such as, biologists, physicists, engineers, epidemiologists, physicians, etc.) and must have been published in a scientific journal with a publisher and/or editorial board (with peer-review process editor/editorial board).

4.4.4 Dataverse

As stated in Section 2.2.2.4, the Harvard Dataverse Repository is a free data repository open to all researchers where you can share, archive, cite, access, and explore research data. All requests for data retrieval towards the Dataverse repository will be handled by the Controller API. The Data Services component of the NIKH platform might be used to temporarily store retrieved data that might need to go into processing operations.

4.4.5 Yoda

As stated in Section 2.2.2.5, Yoda is a research data management service that enables researchers from Utrecht University as well as their partners to deposit, share, publish and preserve large amounts of research data during all stages of a research project. Yoda is currently used as a repository for partners in the ETAIN consortia. Integration

[8] https://www.emf-portal.org

with Yoda is currently under investigation and will be reported and implemented over the upcoming deployment actions, as described in Task 6.4 "Full system integration, release planning and dry-run testing."

4.4.6 European Health Data Space (EHDS)

The European Health Data Space (EHDS) is an ambitious project of the European Union, with the scope of transforming the EU healthcare (European Health Union). The EHDS aims to be an ecosystem combining rules, common standards, practices, and infrastructures, under a common governance framework.

It will rely on two different pillars:

- MyHealth@EU, which is focused on health data exchange between patients and health professionals across member states. The aim is to give European citizens, travelling or living abroad, access to the same healthcare as they would have in their home country. Some of its services are already operational, such as e-Prescriptions and Patient Summaries, and the rest will be implemented progressively across member states until the end of 2025 (clinical reports, lab tests, discharge reports etc).
- HealthData@EU, which will be focused on what experts term the secondary use of data. Researchers, policy-makers and companies will be able to use and study patients' medical records if they receive a permit from a health data access body that will be set up in each member state.

In terms of governance, the EHDS rules on the re-use of health data built on the framework introduced by the Data Governance Act. A **new European Health Data Space Board** will be created, composed of the representatives of digital health authorities and new health data access bodies from all the Member States, the European Commission, and observers. It will contribute to a consistent application of the rules throughout the EU, including by advising the European Commission, while cooperating with other EU bodies and stakeholders such as patient organizations. In terms of data privacy and security, the EHDS builds upon the GDPR, proposed Governance Act, draft Data Act and NIS Directive.

The EHDS project was first presented in March 2022 and it will take a number of years before all of afore-mentioned functionalities are put in place.

5 Updated flow component architecture

This section attempts to further delineate the processes and interactions between the NIKH components and aggregate all key functionalities of each component and how each component is associated with the rest of the components. In what follows, we provide tables per each component of the NIKH platform, detailing their functionality, their relevance and applicability to the afore-presented Usage Scenarios and the corresponding Requirements they pertain to.

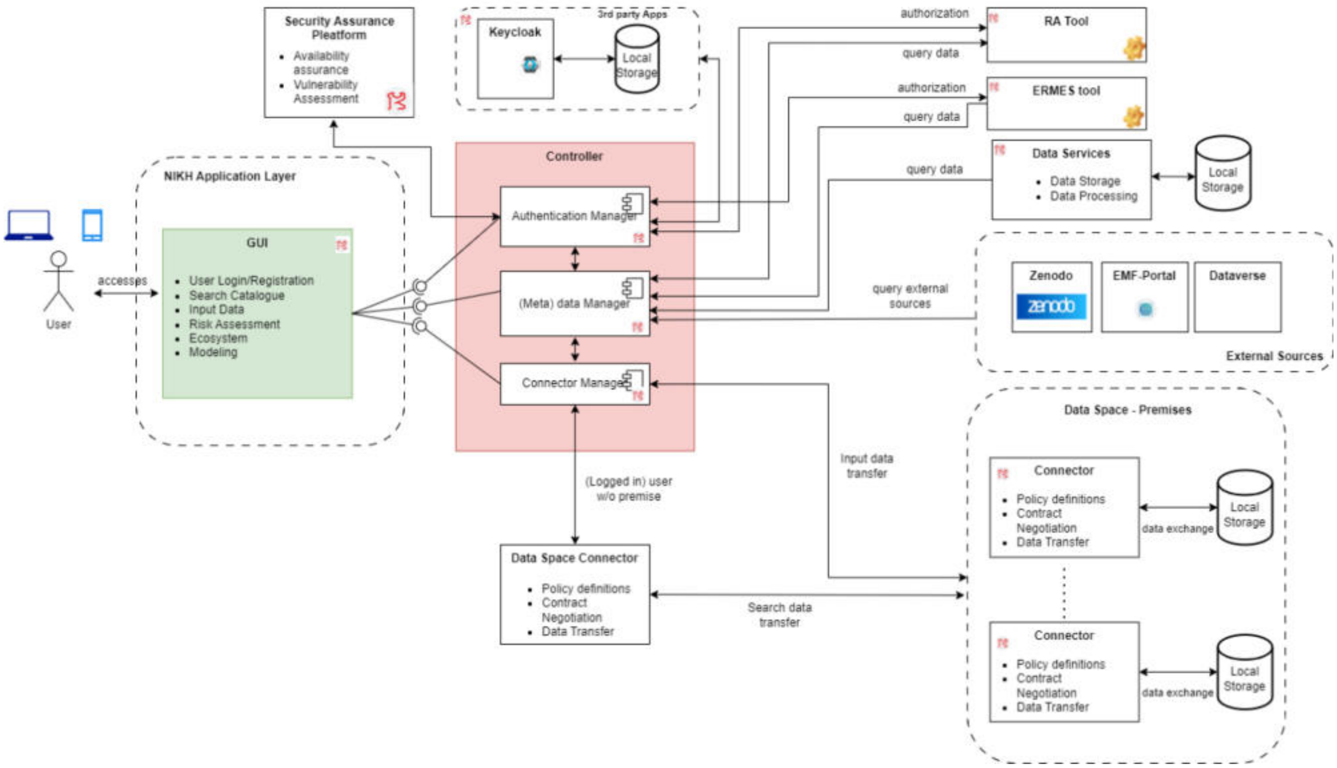


Figure 8: NIKH Architectural Flow diagram

We begin our platform components' descriptions starting from left-to-right as depicted in Figure 8.

Table 17: The Graphical User Interface (GUI)

Component	Graphical User Interface (GUI)
Architectural Layer	Application Layer
Description	The Application Portal and GUI serve as the primary components of the NIKH platform, focusing on core functionalities such as user authentication, role-based access control, and intuitive data interaction. This layer ensures seamless user engagement with EMF exposure data through an efficient and user-friendly interface.
Provided functionalities	• **Login:** Allow registered users to log in securely using their unique usernames and passwords. • **Registration:** Enable new users to create accounts by entering required information such as username, email, password, organization details, title, role, and premises.

	• **Search Catalog:** Implements a GUI that allows users to efficiently search, filter, and retrieve specific data from the catalogue. • **Input Data:** Facilitate data contributors in selecting the data source, uploading real-time data, and providing necessary metadata. • **Risk Assessment Tool:** Provide a GUI for conducting risk assessments based on EMF exposure scenarios. • **ERMES Modelling:** Provide a GUI for ERMES modelling, allowing scientists and researchers to simulate and analyze electromagnetic scenarios. • **Ecosystem/Data Space services:** Offers a centralized space for information on consortium partners, fostering collaboration and knowledge sharing.
Relation to other components	• **Controller:** The GUI interacts with the Controller to send requests and receive appropriate responses. • **Access Management Component:** As the GUI involves user authentication and role-based access control, it interacts with this component for user-related functionalities. • **Data Processing Component:** The GUI may interact with this component to trigger specific data processing operations or visualize processed data. • **Risk Assessment Tool:** The GUI interacts with this tool to initiate risk assessments, visualize assessment results, or provide inputs for risk governance activities.
Related Usage Scenarios	• All Usage Scenarios.
Related Requirements	R1.1, R2.4, R3.3, R4.1, R4.2, R10.1, R10.2, R10.3, R10.4, R11.1, R11.2, R12.1, R13.1, S1, S2, S3, S4

Table 18: The Controller

Component	Controller
Architectural Layer	Service Layer
Description	The Controller acts as the mediator to facilitate the flow and relay of requests among the rest of the NIKH components. It is comprised of the web services that provide the functionality for the end user to register, search, retrieve or upload data. Additionally, it provides access control and authentication for securing the sensitivity of the data.
Provided functionalities	• **Authentication Manager**: Provides user registration, sign-in and user management APIs for the end user. It has an interface with Keycloak, which acts as a third-party identity and access management system and stores information about the platform users.

	• **Metadata Manager**: Provides APIs that enable an end user to upload, remove, update, and retrieve resource (meta-) data. Additionally, it acts as wrapper service that enables the implicit usage of a Data Connector by the end user, for the purpose of registering resources to the data space, through the Connector M. • **Connector Manager:** handles the connectors instantiated in 3rd party premises and the Dataspace connector as part of the NIKH platform, and any other external 3rd party repositories or sources.
Relation to other components	• GUI • Identity and Access Management (Keycloak) • Data Space connectors • External repositories
Related Usage Scenarios	• Usage Scenario 1 – NextGEM Members • Usage Scenario 2 – NextGEM Local Premises • Usage Scenario 3 – NextGEM Limited Local Premises • Usage Scenario 4 – EMF and Health Cluster (CLUE-H) • Usage Scenario 10 – Scientific Community
Related Requirements	R1.1, R4.3, R5.3, R6.3, R8.1, S1, S2, S3, S4

Table 19: Third Party App for Authentication and authorization

Component	Authentication and Authorization
Architectural Layer	Service Layer
Description	Identity and Access Management in the NIKH platform is implemented by the third-party service Keycloak
Provided functionalities	• Token-based authentication • User management
Relation to other components	• GUI • Controller
Related Usage Scenarios	• Usage Scenario 1 – NextGEM Members • Usage Scenario 2 – NextGEM Local Premises • Usage Scenario 3 – NextGEM Limited Local Premises • Usage Scenario 4 – EMF and Health Cluster (CLUE-H) • Usage Scenario 10 – Scientific Community
Related Requirements	R1.1, R2.3, R10.1, S1, S2

Table 20: Security Assurance Platform

Component	SAP Platform
Architectural Layer	Service Layer
Description	The Security Assurance Platform is the component that is used to provide insights on the security posture for the entire NIKH platform and its constituent components.
Provided functionalities	• Vulnerability analysis • Continuous monitoring • GUI for accessing the SPA Platform
Relation to other components	• Log Shipper: The SPA platform aassesses event logs gathered by event shippers from the application layer.
Related Usage Scenarios	• All Usage Scenarios
Related Requirements	S3, S4

Table 21: Log shipper

Component	Log shipper
Architectural Layer	Application layer
Description	Instances of Elastic Beats which gather event and log information from the machines on which they are deployed
Provided functionalities	Gather logs and events to be sent to the SA platform for assessment
Relation to other components	Communicates with the SPA platform. Will be deployed on other components from the application layer so that they can be monitored. Such components are: • Controller • Data Services Component • Data Space Connector • Risk Assessment Tool • GUI
Related Usage Scenarios	• Usage Scenario 1 - NextGEM Members • Usage Scenario 2 - NextGEM Local Premises • Usage Scenario 3 - NextGEM Limited Local Premises • Usage Scenario 4 – EMF and Health Cluster (CLUE-H) • Usage Scenario 10 - Scientific Community • Usage Scenario 11 - Public Authorities and Regulators • Usage Scenario 12 - Standardization Bodies and International Organizations • Usage Scenario 13 - Citizen Awareness and Risk Communication
Related Requirements	S3, S4

Table 22: Data Space Connector

Component	Data Space Connector
Architectural Layer	Service Layer
Description	The Data Connector is the component that is used to securely manage and exchange data assets between NextGEM members.
Provided functionalities	• Data exchange • Policy creation/handling • Contract Negotiation
Relation to other components	• Controller. The access information of the Connectors is stored by the Controller. Additionally, the metadata of the available data assets are handled by the NIKH Controller and can be queried and retrieved by the users in the form of a metadata catalog.
Related Usage Scenarios	• Usage Scenario 1 – NextGEM Members • Usage Scenario 2 – NextGEM Local Premises • Usage Scenario 3 – NextGEM Limited Local Premises • Usage Scenario 4 – EMF and Health Cluster (CLUE-H) • Usage Scenario 10 – Scientific Community
Related Requirements	R1.1, R2.1, R2.3, R2.4, R3.1, R3.3, S3, S4, S5, S6

Table 23: The Risk Assessment tool.

Component	Risk Assessment Tool
Architectural Layer	Application Layer
Description	The RA Tool and its components is an application for risk management activities performed by NextGEM members, members of CLUE-H, and other relevant stakeholders.
Provided functionalities	The Tool will perform qualitative and quantitative risk assessment and provide advice for risk mitigation and risk communication.
Relation to other components	• Application Programmable Interfaces (APIs) at the Service Layer. • The Graphical User Interfaces at the Application Layer.
Related Usage Scenarios	• All Usage Scenarios.
Related Requirements	S1, S2, S3, S4, S5, S6.

Table 24: The ERMES Modelling tool

Component	ERMES tool
Architectural Layer	Application Layer
Description	The ERMES modelling tool is an open-source tool used to visualize EMF-related simulation outputs.
Provided functionalities	The ERMES tool can visualize numerical simulations based on a predefined electromagnetic scenario.
Relation to other components	• Application Programmable Interfaces (APIs) at the Service Layer. • The Graphical User Interfaces at the Application Layer.
Related Usage Scenarios	• All Usage Scenarios.
Related Requirements	R3.3, R10.2, R11.2, R12.1, R13.1

Table 25: Data Services

Component	Data Services
Architectural Layer	Service Layer
Description	The Data Services component facilitates the needs of the NIKH platform for storage of data and metadata pertaining to user information and/or data assets or offerings that are gathered from various premises, either from local premises or other external repositories, such as the Zenodo, EMF-portal, etc.
Provided functionalities	• Data Storage • Data Processing.
Relation to other components	• Controller
Related Usage Scenarios	• All Usage Scenarios.
Related Requirements	S1, S2, S3, S4, S5, S6

Table 26: External Sources

Component	External Sources
Architectural Layer	Infrastructure Layer
Description	External sources are considered all sources of information and data that a 3rd party can provide to the NIKH platform. At the

	time of writing this deliverable, the identified 3rd party sources include the Zenodo, the EMF-portal and Dataverse. We are also exploring integration routes for Yoda repository and PubMed. However, for clarity reasons, Figure 7 just depicts the first 3 external sources.
Provided functionalities	• (Meta-) data search and retrieval
Relation to other components	• Controller
Related Usage Scenarios	• All Usage Scenarios.
Related Requirements	S1, S2, S3, S4, S5, S6.

6 Communication diagrams

This section describes the communication diagrams that schematically depict the flow of interactions, taking place among individual components of the NIKH platform, in order to perform a certain operation.

6.1 User initialization workflow

This section describes the user initialization workflow, which is comprised of the sequence of actions taking place to admit a user into the NIKH platform (aka grant access).

This applies to internal stakeholders, which are users within the NextGEM consortium and CLUE-H members. These so-called "internal" members would need to register themselves in the NIKH platform and log in with their credentials in order to interact with NIKH and be able to upload metadata records, make appropriate edits to records and initiate data sharing procedures.

The detailed user initialization workflow is described in the figure below.

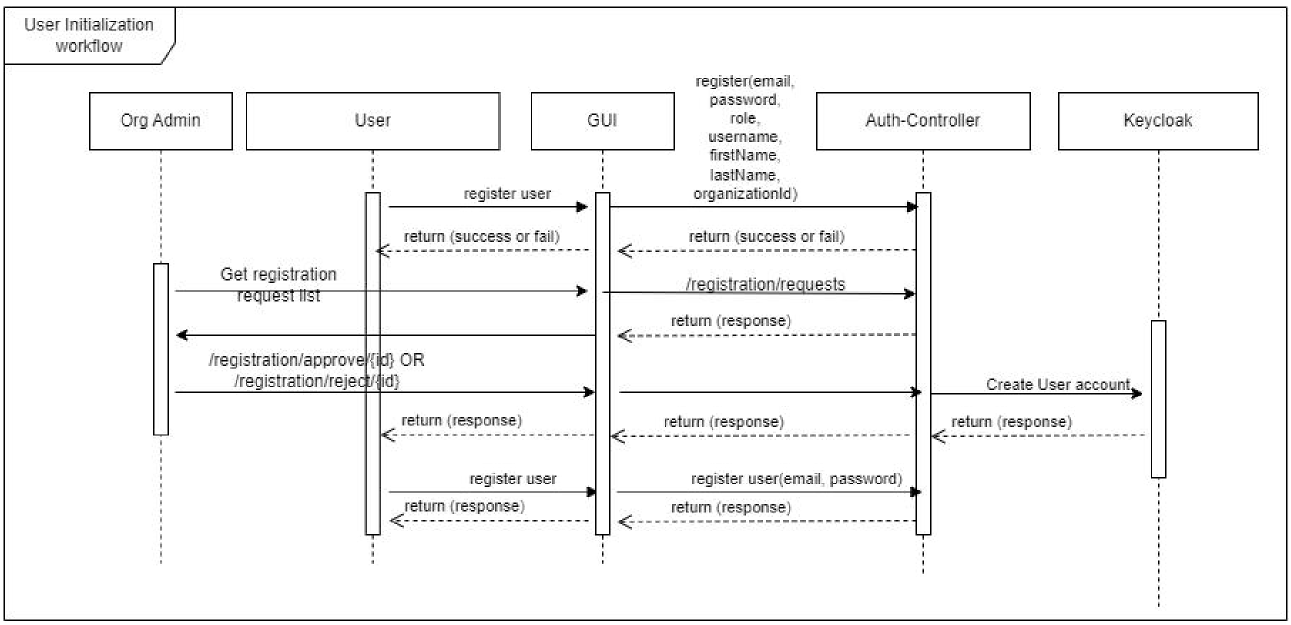


Figure 9: User Initialization workflow.

- A user registers to the NIKH platform via NIKH's GUI by filling out registration fields, comprised of an *email, password, a role, a username, the firstName* and *lastName* and an organizationID. The GUI relays this request to the Authorization part of the Controller, which handles the request and returns an appropriate response.
- Once a user registers for the first time, their request needs to go by their Organization Administrator, who will approve or reject the registration request.
- To do so, the Organization Administrator might need to access the list of all registration requests to check whether the request in question has been previously handled in the past. This functionality is handled by the Controller, Authentication Manager.
- Then, the Organization Administrator approves or rejects the request specified by an {id}, which triggers an analogous process for the creation of the User profile in Keycloak.
- After the successful registration, the user logs in to the NIKH platform, via the GUI, using their email and password.

6.2 Metadata workflows

This section describes the communication diagrams pertaining to operations on the metadata. Based on the usage scenarios identified earlier, the operations that can be performed on metadata are the following:

- The creation of metadata records (or upload)
- The search of metadata record(s)
- The view of metadata record(s)
- The editing of metadata record(s)
- The deletion of metadata record(s).

In the context of a role-based access policy, we have distinguished between two types of users, each of which has different rights related to operations on (meta)-data:

- A simple user, who is able to only search and view metadata record(s)
- An Administrator-type of user, who has full access rights to perform all five operations.

Table 27: User roles and admissions.

	Create/Upload	Search	View	Edit	Delete
Administrator	✓	✓	✓	✓	✓
Simple User		✓	✓		

6.2.1 Upload metadata

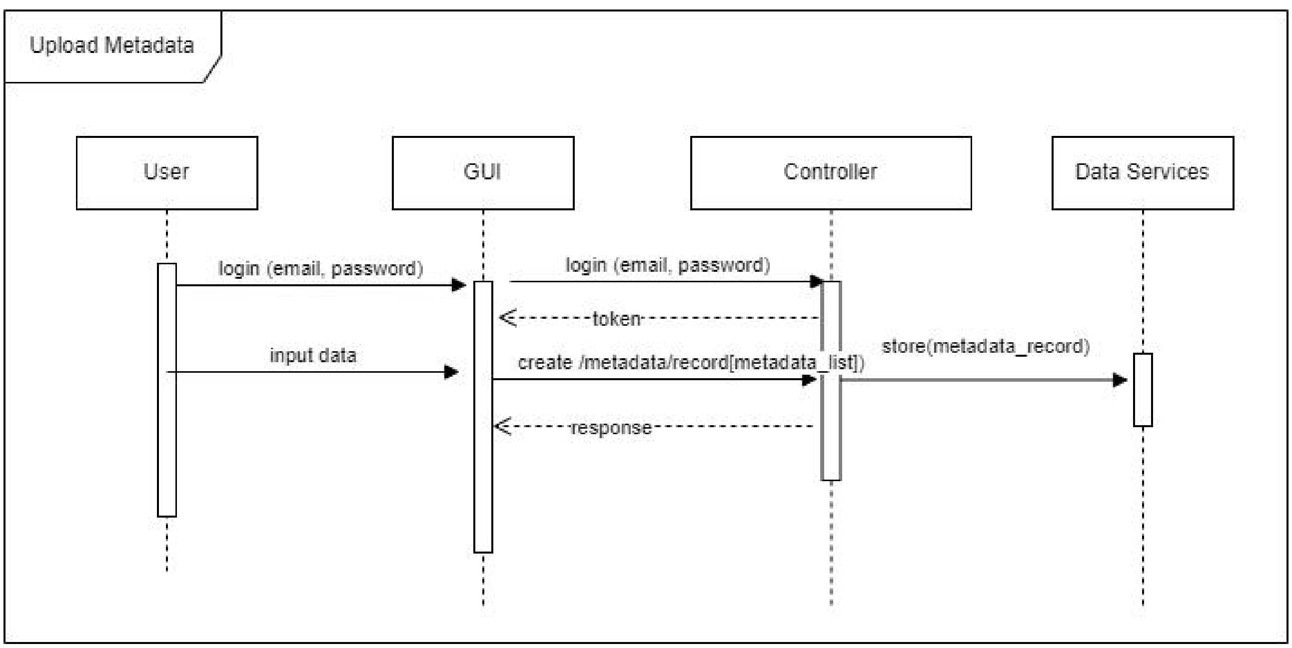


Figure 10: Communication Diagram for uploading metadata records.

In order to be able to upload data, a registered user with Admin rights logs into the NIKH platform via the GUI, with his/her email and password credentials. Via the GUI, the user selects the appropriate button that triggers a REST-API endpoint for the creation of the metadata record that is handled by the Controller. The metadata records are subsequently forwarded for storage to the Data Services component. The Controller returns a response to the GUI pertaining either to a successful creation, or unsuccessful operation and/or denied operation.

6.2.2 Search metadata

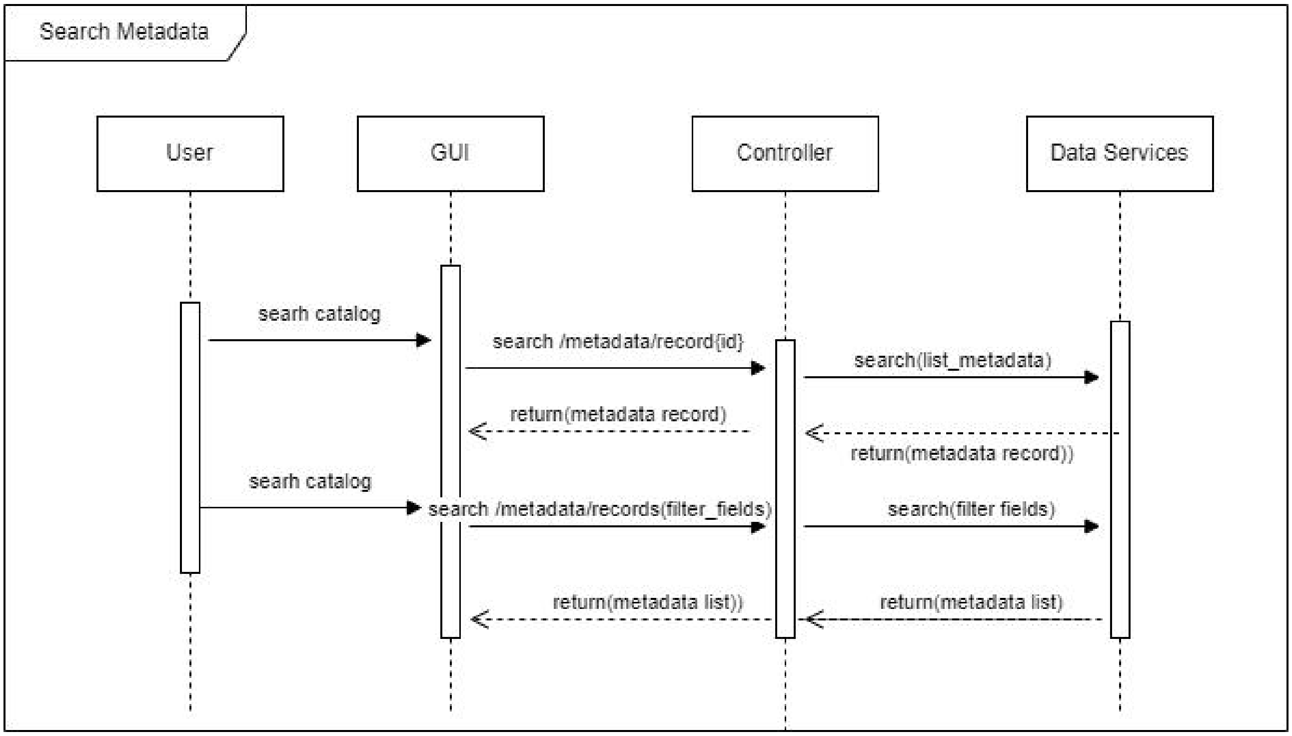


Figure 11: Communication diagram for searching metadata records.

Any user, whether they are registered in the NIKH platform or not, is able to search for metadata records, via the GUI. The user has then the option to either list all the available metadata records and also to view a specific one. Based on the request, appropriate calls are made to the appropriate Controller endpoints, which handle the requests and access the Data Services for retrieval.

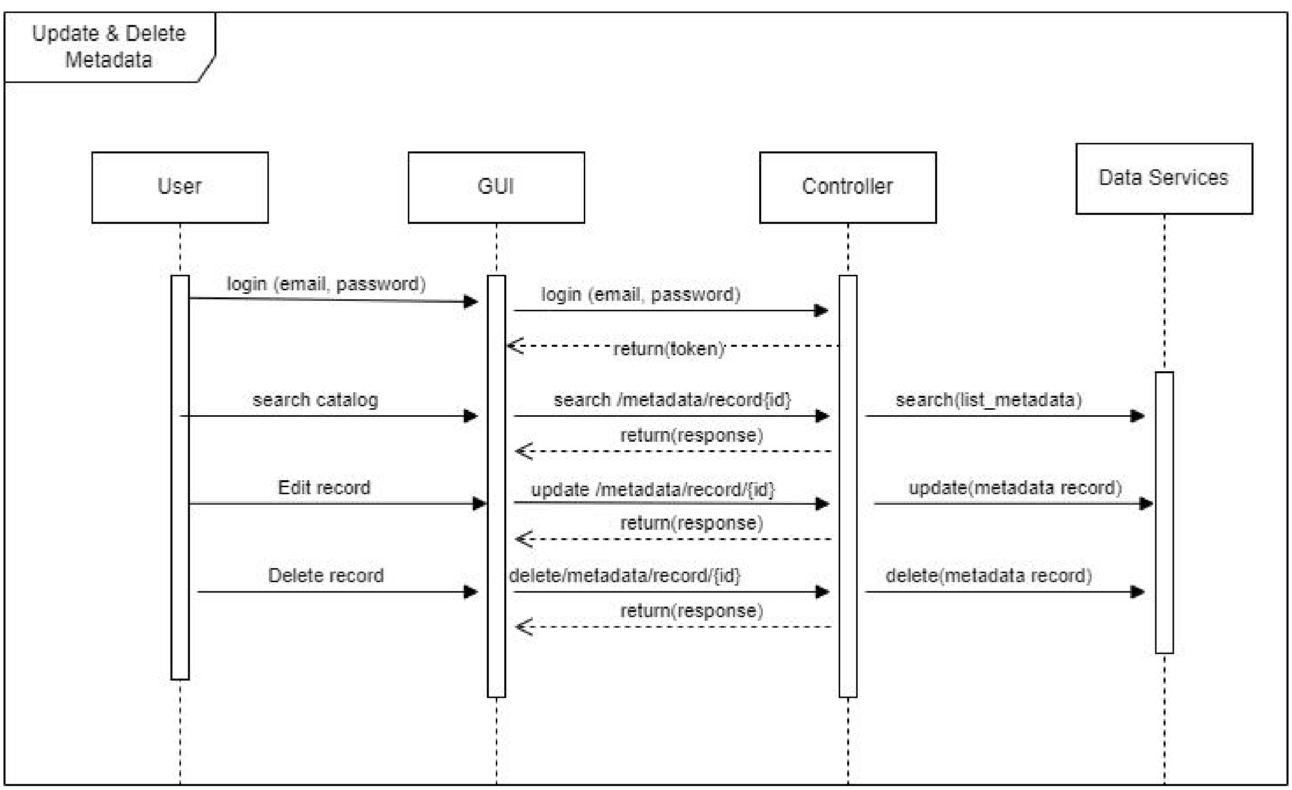


Figure 12: Communication diagram for updating or deleting metadata records.

Operations pertaining to editing and deleting a metadata record can only be accessed by registered users. Therefore, registered users need to log in first to the NIKH platform via the GUI, by inserting their email and password. Then, after searching for the desired metadata record they can perform edit and deletion operations, by opting for the right functionality in the GUI, which will relay the request to the Controller and handle the operation. The Controller connects to the Data Services component to retrieve the metadata record and then save the edited record or delete it altogether, accordingly.

6.3 Data transfer workflows

In the event the user wants to initiate a data transfer (e.g. download actual data), they should first search the catalogue to identify the metadata record they are interested in, and then initiate a data transfer flow. Depending on the type of access policy corresponding to the data the user wishes to download, we have identified the following scenarios:

1. **A user wants to download a document under the "Public" access policy**. In this case, the user does not have to be registered or logged in to the NIKH platform. He/she will be able to download the desired file after identifying it in the Search Catalog. If it is "Public" and the data rests in the local premises, then Data Connectors get involved in the process of data transfer and hence, a login process (to the NIKH platform) is required. If data under "Public" labelling lay on external, 3rd party sources, such as Zenodo, EMF-portal, etc., then transfer/download process can be directly handled by the NIKH platform itself, without any need for logging in.
2. **A user wants to download a document under the "Public after embargo" access policy.** In this case, the document is not at first accessible to the users of the NIKH platform, regardless of them being logged in or not. After the embargo period, namely after the date that the data owner has defined, the data are public, thus the same procedures as to what has been defined for the "Public" access policy apply as well.
3. **A user wants to download a document under the "Restricted" access policy.** In this case, in order to be able to request the specific document, the user has to be registered and logged in to the NIKH platform with their credentials. After searching the metadata catalogue and identifying the desired data, the user can request access to the data and they are then provided with contact information related to the owner of the respective file(s), along with information about the owner's Data Connector and the contract that is associated with the data. If the user requests to access the data, after accepting the terms of the accompanying contract, the NIKH platform initiates the data transfer process, by engaging the Data Space Connectors (as presented in the previous sections) that carry out the complete transfer cycle, ensuring transparency in a trustworthy and sovereign data exchange scheme.
   If the user who requests the data, participates in the NIKH dataspace with their own local Data Connector, then the complete data exchange is performed between their Connector and the data owner's Connector. Alternatively, if the user requesting access to the data does not use a Data Connector of their own, then the data exchange process involves the NIKH Connector as a data consumer's Connector and the data owner's Connector as the data provider.
   At the time of writing this deliverable, a number of options are being considered for the implementation of the Data Space Connectors. It is our intention to present the data transfer workflows in the upcoming deliverable D6.2 "Network provision and links with EU health data space - Initial report", after finalizing our research and implementation efforts.
4. **A user wants to download a document under the "Closed" access policy.** In this case users, either logged in or not, cannot access and download the data.

7 Integration strategy

The micro-services architecture will be used for the design and development of the NIKH platform to enable independent development, maintenance, and deployment of each individual component of its core architecture. The micro-services architecture allows to structure an application as a collection of autonomous but loosely coupled services which are able to work together. This section describes the integration approach that will be used for the development and deployment of the NIKH technical platform.

7.1 Components integration

7.1.1 Containerization

Containerization is a form of operating system virtualization that allows to package applications and their dependencies in containers (i.e., loosely isolated environments) to support their secure and isolated execution on a same shared operating system and hardware. It allows applications to run quickly and reliably in different computing environments, independently of the underlying hardware and operating system. Applications are encapsulated in Docker [9] container images, executable packages of software that include everything required to run the application, such as code, runtime, libraries, and settings. Container images become containers at runtime.

Each NIKH component will be encapsulated in a container image to be deployed independently in the form of a Docker container.

The consortium partners that own NIKH components will provide the configuration files required to build the Docker images for the deployment of their components, such as Dockerfiles or Docker Compose files for more complex components.

7.1.2 Container orchestration

It is typical in the microservice architecture that all services operate within their own containers, including third-party components commonly used for integration and storage, like message queues and databases. A microservice platform can thus be composed of several interconnected containers, which proves a challenge to managing them individually. There are several container orchestration solutions available, the most widely known being Kubernetes.

Kubernetes [10] is popular for hosting microservice projects due to its ability to manage a large number of containers, not only on a single host, but across multiple hosts and clusters, often located in different geographical coordinates. Besides handling the lifecycle of each container, Kubernetes also takes care of resource allocation (e.g., CPU, storage, RAM), dynamic service scalability, networking, load balancing and failure amendments. Additionally, it offers a wide variety of administration tools including cluster management, resource monitoring and role-based access control.

It is important to distinguish that Kubernetes only manages the orchestration of the containers and not their runtime, as this is handled by the virtualization engines mentioned previously.

The smallest logical unit that Kubernetes can manage is called a Pod (13). A Pod may contain one or more containers, which correspond to different processes that would typically run on a physical server. In the majority of microservices, however, each Pod is assigned only one container, essentially operating one service per Pod. This method eases the complexity of the architecture as it is easier to manage, configure, debug and monitor the standalone services. Then, then containers placed into Pods run on *Nodes*. A node may be a virtual or physical machine, depending on the cluster. Each node is managed by a control plane and contains the services necessary to run Pods.

Lastly, besides some initial configuration, if any, the developer and/or maintainer of the services rarely needs to step in, as management of resources per service is also taken care of by Kubernetes.

[9] Docker, https://www.docker.com/

[10] Kubernetes, https://kubernetes.io/

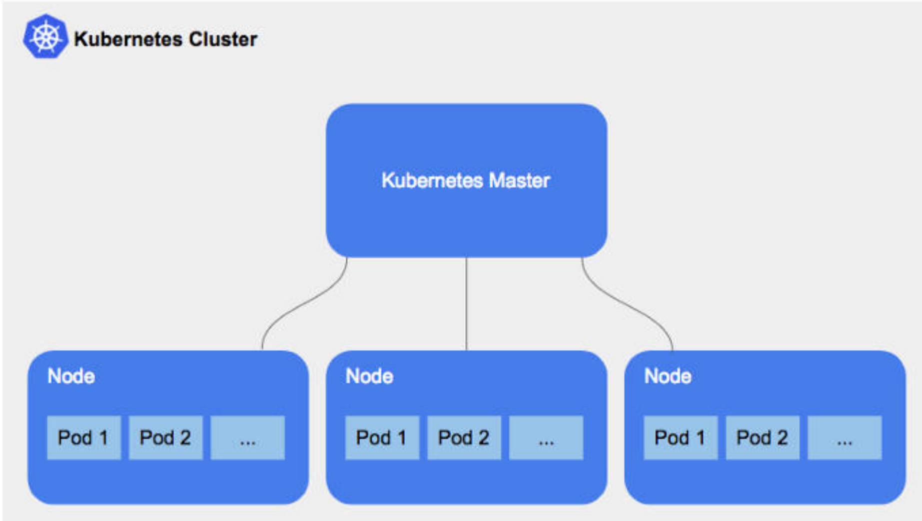


Figure 13: Architecture of a Kubernetes Cluster.

7.2 Continuous deployment and integration

An important concept in software development is deployment automation which is employed frequently in monolithic applications and eventually became an indispensable part of microservice-based development. Deployment usually refers to the stages following the initial development, where the different source code segments are merged, built, and tested in a testing environment, and subsequently deployed to the production environment. It quickly becomes apparent that this process, especially when it involves complex software that requires resources and dependencies outside the base framework used for development, is unmanageable if conducted manually. To rectify this, software teams use automation scripts to handle the stages mentioned, with minimal human intervention. This method is referred to as Continuous Integration/Continuous Delivery (CI/CD) as depicted in Figure 14. By using this automation, new features and extensions can be added to the software quickly and efficiently and be ready for production in a relatively small amount of time.

Continuous Integration (CI) scripts are applied after development has concluded to merge the different software pieces, build them, and then test their functionality. These scripts often contain system parameters, configuration of the execution of the code and instructions for the build environment. The latter may especially apply when dealing with independent services that may run on different operating systems or may have been developed using different programming languages. CI then requires the instructions to be written only once and all future iterations of the system are built and tested without the overhead of manual configuration.

Continuous Delivery (CD) follows the CI phase and requires the successfully tested software components to be deployed on a local or remote server infrastructure, ready for public use. This is referred to as the production environment and all software running on it is presumably fault-free, as it has already been tested on a development environment that replicates the conditions (configuration) of the production one. The deployment script is usually set to be triggered after the CI script has completed successfully. The result of a successful CD can either be a service, e.g.. a website, or a software package that can be downloaded, installed, and executed.

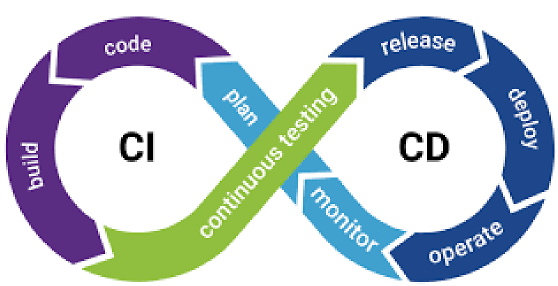


Figure 14: The CI/CD Pipeline

7.3 GitLab code repository

GitLab [11] is a cloud-based Git repository hosting service for open-source projects. It will be used to host and make available the source code of the NIKH platform and also instructions for the deployment of the NIKH platform components, which are part of activities in WP6 of the project.

A collaborative GitLab repository has been set up on ICOM's premises.

A snapshot of the GitLab repo is presented below.

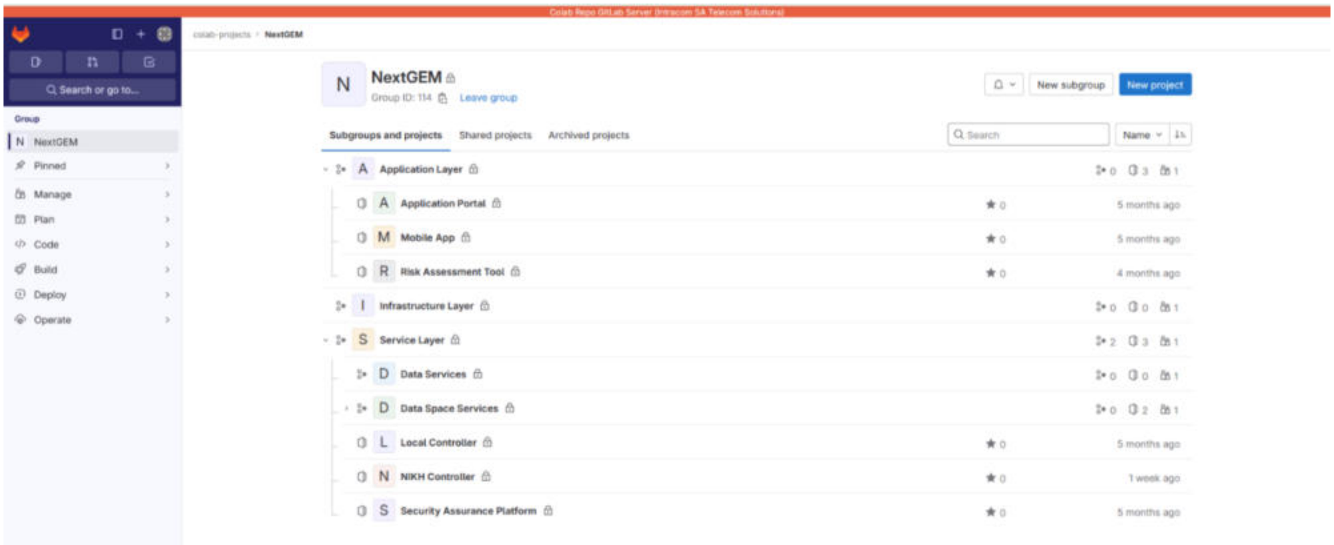


Figure 15: Snapshot of GitLab.

[11] GitLab, https://about.gitlab.com/

8 Conclusion

This deliverable presented the final version of the NIKH architectural framework, showcased the usage scenarios, highlighted the aims and needs for the NIKH platform, and provided a detailed overview of the core components comprising the platform.

Section 2 provided an updated overview of the usage scenarios and the functional and non-functional specifications to be considered for the deployment of the NIKH platform.

Section 3 provided a description of the data that will be managed within NIKH and laid the foundations for further identifying storage and interoperability needs, as well as privacy and access rights.

Section 4 provided a description of the core components of the NIKH platform, which will provide a basis for the deployment activities in WP6 "Development of NextGEM Innovation and Knowledge Hub".

Section 5 detailed the architectural flow diagram of the NIKH platform, along with summaries of the functionalities, requirements, related US for each of the NIKH components.

Section 6 provided an initial view of the communication between individual components of the NIKH platform that allow for the enactment of functionalities provided by NIKH platform to interested stakeholders.

Section 7 laid the foundations for the micro-services' type deployment of the core components and provided an insight to the integration strategy for the deployment of the integrated NIKH prototype.

As a conclusion, NIKH is considered to be a game changer and can be developed as a tool which is versatile, flexible, adaptable and ready to employ AI/ML for securing the overall goal of NextGEM which is to provide a safe environment for use of 5G NR and coming generations of mobile communication.