



DARE

DIGITAL LIFELONG PREVENTION

CODE NO. PNC0000002

Spoke 1 Deliverable

S1.D3.1

Analysis of the Technological Framework and Interoperability Requirements

This research is co-funded by the Ministry of University and Research within the Complementary National Plan PNC-I.1 "Research initiatives for innovative technologies and pathways in the health and welfare sector"
D.D. 931 of 06/06/2022, PNC0000002 DARE - Digital Lifelong Prevention



S1.D3.1 Analysis of the Technological Framework and Interoperability Requirements - Update

Deliverable information	
Spoke number and title	Spoke 1 - Enabling Factors and Technologies for Digital Prevention
WP number and title	WP 3 - Interoperability Governance
Related task(s)	Task 3.1 - Multi-dimensional governance and coordination
Lead beneficiary	INFN
Contributing beneficiaries	UNIBO, EXP
Dissemination level	Public
Due date	14/12/2024
Actual date of delivery	14/4/2025
Author(s)	Barbara Martelli (INFN)
Contributors	Andrea Chierici (INFN), Sabato Mellone (UNIBO)
Quality Assurance	Sabato Mellone (UNIBO), Giuseppe Parrinello (EXP)

Document history

Version	Date	Author(s) /Reviewer(s) (Beneficiary)	Description
0.1	01/11/2023	Barbara Martelli Andrea Chierici	First draft
0.2	28/11/2023	Giuseppe Parrinello	Revision
0.3	01/12/2023	Sabato Mellone	Draft of Section 2.2 and document revision
0.4	15/12/2023	Barbara Martelli	Final revision
1.0	18/12/2023	Barbara Martelli	Final document
1.1	09/12/2024	Sabato Mellone	Draft of Chapter 4, document update and revision
2.0	14/04/2025	Barbara Martelli	Final revision of the update and final update of the document

Disclaimer

This publication reflects only the author's views and the Funding Agency is not liable for any use that may be made of the information contained therein.

Table of contents

1. Publishable summary	5
2. Introduction	6
3. Technological framework	6
3.1. INFN DataCloud	6
3.2. AlmaHealthDB	7
3.3. Integration between INFN DataCloud and AlmaHealthDB	8
4. Interoperability requirements	9
4.1. Metrics for interoperability and portability evaluation	11
4.2. Evaluation process	12
4.3. Classification of DARE applications	13
5. National Development and Integration Strategy	14
5.1. Technical Activities in DARE	14
5.2. Use Cases	14
5.3. First Phase and short-term goals	16
5.4. Second phase and long-term goals	18
5.5. Update on implementation activities described in the short-term plan (section 5.3)	19
6. DARE-ICSC Agreement	20
7. Bibliography	21

1. Publishable summary

This document provides a comprehensive update on the technological framework, interoperability requirements, and integration strategies for digital health data management in Italy in the context of the DARE project. It discusses two primary technological pillars: the INFN DataCloud and the AlmaHealthDB, which are designed to ensure secure data management and compliance with legal and regulatory standards.

The document defines cloud interoperability as the ability of systems to interact and exchange information, emphasizing the importance of avoiding vendor lock-in and ensuring data portability.

It proposes metrics to evaluate interoperability and portability, including transport interoperability, policy compliance, and data retrieval capabilities, which will aid in assessing the cost-effectiveness of porting applications to the DARE platform.

Moreover, the document proposes a National Integration Strategy to create a unified approach for clinical data processing across various projects, promoting the use and reuse of clinical data for research while adhering to legal constraints. The strategy starts from the definition of short-term objectives, including the secure deployment of the Salus Ratio application on certified cloud services and the establishment of a centralized authentication system using Keycloak. Then, a long-term vision is proposed as a second phase of the strategy, focusing on making Salus Ratio as a Service available on the ICSC cloud, allowing for the management of sensitive data in compliance with GDPR.

2. Introduction

According to the Italian Department for Digital Transformation and National Security Agency, [1] migration to the Cloud allows to exploit secure, efficient, and reliable technological infrastructures, in line with the principles of privacy protection and the recommendations of European and national institutions, while maintaining the necessary guarantees for the country's strategic autonomy, security and national control over data. They recommend a “Cloud First” strategy to simplify and optimize the management of IT resources and facilitate the adoption of new digital technologies.

With reference to the Italian Cloud Strategy, DARE data can be classified as *Critical*, that is data and services the impairment of which could be detrimental to the maintenance of functions that are important to society, health, safety, and the economic and social well-being of the country. This kind of data need to be managed in a *qualified cloud service with management of encryption keys in Italy*.

In order to meet these regulatory and governance recommendations, the DARE project decided to base its computing infrastructure on the combination of two main technological pillars:

- the **INFN DataCloud**, based on secure cloud technologies developed by INFN in the last two decades to cope with the big data challenges faced by physics and life science scientists;
- the **AlmaHealthDB** software stack developed by UNIBO, IOR, IRCS AOU BO, and IRCCS ISNB to solve the issue of secure clinical data collection and analysis in compliance with legal, ethical, organisational, and regulatory requirements.

This document constitutes an official updated version of Deliverable D3.1, submitted at M12.

3. Technological framework

3.1. INFN DataCloud

The INFN DataCloud platform is described in D4.1. It is based on the following key components:

- INDIGO IAM: enabling the federation of identities through standard protocols like OpenID Connect and OAuth 2.0;
- INDIGO PaaS Orchestrator: enabling service orchestration and dynamic allocation of resources in a pool of geographically distributed data centers;
- INDIGO PaaS Dashboard: enabling service composition and deployment using the *Infrastructure as a Code* paradigm.

3.2. AlmaHealthDB

AlmaHealthDB The AHDB infrastructure consists of a series of virtual machines and a private network (VPN) hosted within the Regional Health Service network managed by Lepida S.c.p.A, the in-house company of the Emilia Romagna region. All virtual machines in the "trust" area can see and communicate with each other and can connect to the virtual machine(s) with a public IP exposing the data ingestion services.

In the AHDB model, Lepida S.c.p.A is the IaaS provider, certified according to ISO 9001 quality standards, ISO/IEC 27001 information security standard, ISO/IEC 27017 security controls for cloud services standard, and ISO/IEC 27018 on protecting personal data on public cloud systems.

AlmaHealthDB user interfaces for data transfer include:

- REDCap - Clinical trial manager that allows the input of structured data and small files (a few MB).
- SFTP/HTTPS Client - Application for secure file transfer, even for large files. For standard formats, it is possible to include rules for anonymization/pseudonymization before the transfer.

Ingested structured and unstructured data can become an input to a data standardization pipeline. In order to achieve a declarative, modular and maintainable transformation pipeline, AlmaHealthDB chose templating as the conversion strategy. Specifically, it makes use of the HL7 FHIR Mapping Language [2], a mapping language specification used to convert data from the custom formats of the input data to the FHIR model.

AlmaHealthDB selected Matchbox [3], as a FHIR templating engine implementation, due to its versatile features, including offline transformation, standalone microservice deployment, seamless integration as a Java library, and extensive documentation.

Experts in HL7 FHIR R5 annotate input variables into corresponding FHIR resources. This phase also requires domain experts' contributions for resolving discrepancies in the mapping process, leading to the establishment of an agreed resource mapping. Once all variables had been successfully mapped, in accordance with HL7 recommendations, AHDB administrators proceed in identifying an appropriate coding system.

Next steps in the development of the AlmaHealthDB standardization pipeline include the addition of a FHIR to OMOP (Observational Medical Outcomes Partnership) CDM (Common Data Model) converter and the continuous extension and validation of the clinical study templates (Figure 1).

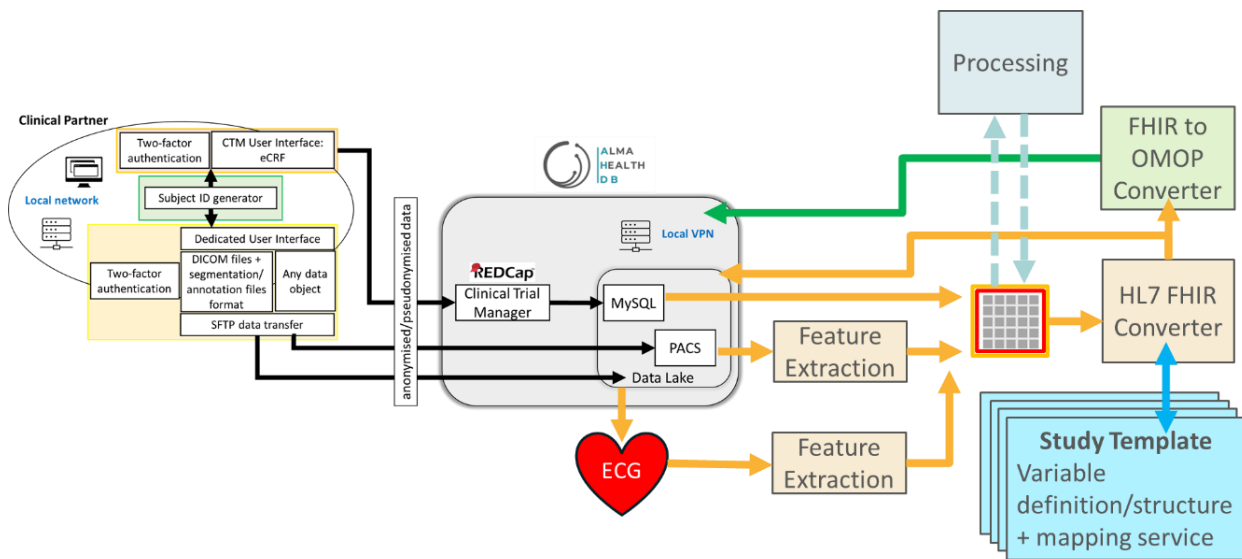


Figure 1 - The AlmaHealthDB standardization pipeline. In the example structured data (e.g. data from the clinical trial manager) is a direct input of the HL7 FHIR converter, while unstructured data, such as medical images and biomedical signals, are an input of dedicated feature extraction functional blocks whose output are structured datasets that can be inputted into the HL7 FHIR converter. Similarly for the output of other data and signals processing functions. An integration with a FHIR to OMOP CDM converted is expected downstream of the FHIR converter.

3.3. Integration between INFN DataCloud and AlmaHealthDB

The AlmaHealthDB infrastructure allows to apply machine/statistical learning algorithms that are not particularly resource intensive. Whenever data and signal processing require large computational capacity, the AlmaHealthDB strategy is to exploit computational resources and services provided by INFN and, more in general, of the National Research Centre for High Performance Computing, Big Data and Quantum Computing. The ability to leverage VMs equipped with multi-CPU and GPUs for deep learning applications, or otherwise resource-intensive tasks, is an activity that has recently been initiated in collaboration with INFN and CINECA (Figure 2). An integration plan is under development.

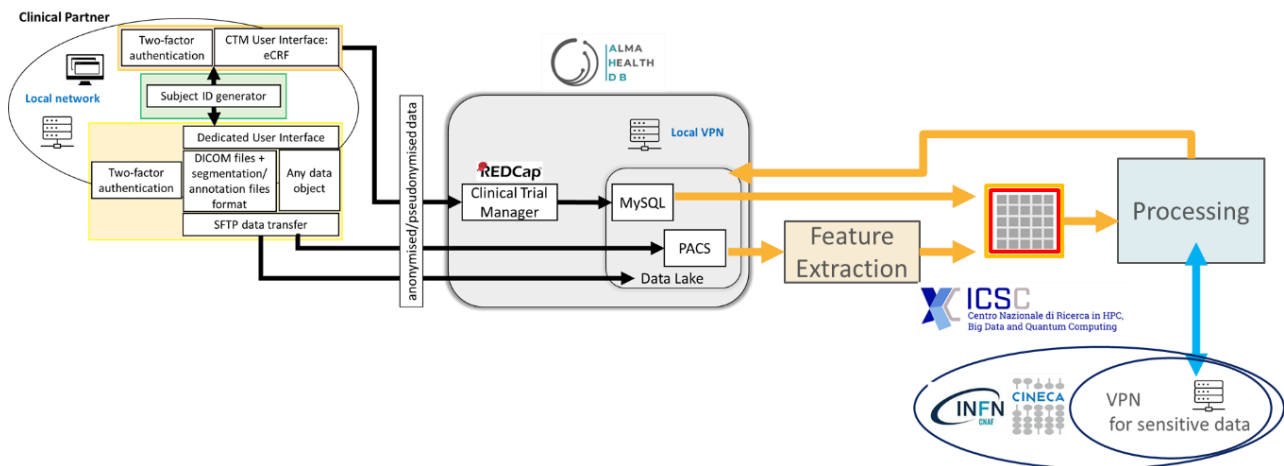


Figure 2 - Integration of the AlmaHealthDB infrastructure with the cloud services provided by INFN, and, more in general, by the ICSC.

4. Interoperability requirements

As defined in ISO/IEC 22123 [4], cloud interoperability is the ability of a cloud service customer's system to interact with a cloud service, or the ability for one cloud service to interact with other cloud services, by exchanging information according to a prescribed method to obtain predictable results. It is based on the concept of cloud portability: the ability of a Cloud Service Customer (CSC) to move their data or their applications between two different cloud services at a low cost and with minimal disruption. Interoperability and portability are significant in cloud computing because CSCs are interested in avoiding lock-in when they choose to use cloud services. In fact, vendor lock-in is one of the most relevant risks in the adoption of cloud services according to the ENISA Cloud Computing Risk Assessment [5].

On the other hand, the same report [5] lists a number of advantages that the adoption of cloud computing brings to users. Some examples are the benefits of scale in term of information confidentiality, integrity and availability; standardized interfaces for managed security services; rapid and smart scaling of resources; more timely, effective and efficient updates and defaults. That's why many DARE applications choose to adopt cloud services and in the context of the project, invest on enhancing the cloud readiness of their storage and computing tools.

DARE applications are proposed by Spoke2 and Spoke3 in the form of pilot projects or twin projects. After a review process performed by Spoke1, the applications wanting to exploit the DARE platform are evaluated by Spoke1-WP3 to identify gaps and improvement actions needed to enhance their interoperability and standardization with

the final goal of building an ecosystem of tools for preventive medicine usable by researchers in an user friendly and composable way.

Interoperability and portability in a cloud computing environment is not a binary concept. Almost all applications running in a cloud service can be made interoperable with others or can be ported to another cloud service offering equivalent capabilities if enough resources are invested. The critical considerations are:

- the interoperability and/or porting time and cost,
- the risks associated with interconnection of applications or their porting.

In this deliverable we aim to define metrics, evaluation processes and strategies to control the costs and risks compared to the expected benefits of porting applications on the DARE computing platform.

Facets of Interoperability and portability

Interoperability and portability have different facets [6]. The most relevant in the context of DARE are:

- **Transport interoperability:** the transfer mechanism between various cloud computing components, either between Cloud Service Customer (CDC) components and Cloud Service Provider (CSP) components or between CSP components related to different cloud services.
- **Policy interoperability:** the ability of two or more systems to interoperate while complying with the legal, organizational and policy frameworks applicable to the participating systems. The DARE applications are located in the Italian territory, therefore applicable rules and regulations are the European and Italian ones. Policy interoperability includes policies which relate to specific capabilities of cloud services.
- **Data portability - retrieval of cloud service customer data:** possibility to extract data from source and store it at destination. One key aspect is the amount of data and the bandwidth available to data movement. Is the hospital connected through a high bandwidth network? Is it connected through a GARR link? (in this case temporary dedicated links can be arranged to allow the data transfer in a reasonable amount of time). If data is too large and the transfer has a relevant impact on the available bandwidth between systems, data might be moved by transfer of physical storage media.
- **Data portability - syntax of the data:** the ability of two or more systems or services to understand the structure of exchanged information.

- Data portability - semantics of the data: the ability for the systems exchanging information to understand the meaning of the data model within the context of a subject area.
- Application syntactic portability: the received application should be in understood format. Need to agree to common package format.
- Application instruction: the received application should be executed in a functional equivalent manner. Supported runtime environments are C++, Java, C#, BPEL
- Application metadata: the received application and the cloud environment should have a mutual understanding of environmental dependencies.
- Application behavior: the application should produce expected results.
- Integration with the INDIGO PaaS Orchestrator: as described in D4.1 [7], the cloud orchestration in the DARE cloud platform is the INDIGO PaaS Orchestrator. Applications can be integrated using TOSCA [8] templates.
- Identity and Access Management federation: when porting the applications to a target cloud service, there is the choice of either switching to use an IdAM system supplied by the cloud service or else continuing to use a private IdAM system. As described in D4.1 [7], DARE DataCloud platform IdAM is the INDIGO IAM which supports identity federation through the standard protocols OAuth 2.0 and OpenID Connect. INDIGO IAM can be exploited as a general IdP Proxy for the entire infrastructure or as an “AAI-as-a-Service” solution that can be self-instantiated through the INFN DataCloud Service Catalog. In both cases, the aim is to have a single location for the control of the identities of DARE users. This improves security since only one IdAM system needs updating for significant events such as the removal of access for a user.

4.1. Metrics for interoperability and portability evaluation

- Transport Interoperability metric: the following protocols are supported by the DARE platform HTTP/S, SOAP, Advanced Message Queuing Protocol (AMQP) and Message Queuing Telemetry Transport (MQTT)
- Policy interoperability metric: in case different cloud providers are involved, DARE requires that all are certified ISO 27001 27017 27018 (information security and privacy).
- Data portability - retrieval of cloud service customer data - metric: amount of data compared to bandwidth available without impact on the hospital normal activities.
- Data portability - syntax of the data - metric: recommended formats are JSON [9] and DICOM for images [10]. If the syntax is different, evaluate the cost of translating to a JSON format.

- Data portability - semantics of the data - metric: recommended ontologies are Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) [11], HL7 FHIR [12]. The federated data storage will have to comply with the ISO 13606 “EHR interoperability”. In case of inconsistencies, evaluate the cost of translating to this standard reference, following a process like the one described in [13].
 - Application syntactic portability metric: the accepted package formats are Zip, tar, jar;
 - Application instruction portability: supported runtime environments are C++, Java, C#, BPEL;
 - Application metadata portability metric: supported metadata models are XML, JSON, YAML, TOSCA;
 - Application behavior portability metric: supported application tests suites are: Apache JMeter, SoapUI, Robot Framework;
 - Integration with the INDIGO PaaS Orchestrator metric: effort to create the TOSCA template in order to integrate the application in the DARE DataCloud dashboard and enable its on-demand deployment;
 - Identity and Access Management federation metric: effort to integrate the hospital IdP with INDIGO IAM or to migrate the user data to the DARE INDIGO IAM service. Supported identity management protocols are OAuth 2.0 and OpenID Connect.

4.2. Evaluation process

The evaluation process will consider usage of common programming languages, standards, tools, frameworks, models, run times and APIs, based on metrics defined in the previous paragraph. Those metrics will be composed in order to compute the following cumulative metrics:

- [M1] time required for porting both applications and data to the DARE federated cloud platform.
- [M2] effort required for porting both applications and data to the DARE federated cloud platform.

To evaluate M1 and M2 metrics, the applications will be classified based on their cloud readiness. Legacy, monolithic applications will need to be reengineered to be run on cloud, a development plan will be therefore required. Cloud ready application we'll only need a deployment plan to be presented to Spoke1, in order to evaluate the effort for their activation on the DARE computing platform.

4.3. Classification of DARE applications

The applications in the DARE context can be legacy applications or cloud native applications.

- Legacy applications: require a development strategy to be converted into cloud-ready applications, before being deployed on cloud. The development is the responsibility of the application provider. The development strategy needs to consider the following aspects:
 - o in addition to application logic, it may be necessary to port or reconfigure the cloud application or the components upon which it depends, e.g., libraries, databases and web servers.
 - o The sequence of virtual machine or component start-up may also be important. Portability of complex applications may also require INFN DataCloud to share application metadata. Examples of this metadata are details regarding the relationships and dependencies between various application components, requirements such as the acceptable range of component versions, start-up sequence, network and firewall configuration, processing capacity, co-location rules and load balancing requirements.
 - o Data architecture should provide separation between processing and data, loosely coupling the applications.
 - o The application should be horizontal scalable (not multi-tier scalable), enabling for performance scaling in cloud.
 - o Depending on how the application is structured to use CPU and memory resources, the virtualized versions of these components provided by the cloud environment must provide an equivalent level of support for the application to function properly. The evaluation of resource allocation must be included in the development plan.
- Cloud native applications: ready to be deployed on the cloud infrastructure, only a **deployment plan** is needed. The deployment plan must consider the following aspects:
 - o **the ability** of a target cloud service **to replicate the environment** that the source cloud service has for the application or at least create an environment that similarly satisfies the dependencies of the application.
 - o Different cloud services rarely provide identical **capabilities to support** all the activities for **all sub-roles**. The effort necessary to adjust for these differences and the potential benefits need to be considered. For example, a cloud application implemented on an Infrastructure as a Service moved to a different cloud service of the same type might provide identical capabilities to support the *user sub-role* deploying and operating the application, but very different capabilities *administrator sub-role* managing the use of the cloud service.
 - o **interfaces needed by the application** in the source environment **must also be available** in the target environment. In particular, the application provider must list the needs for service discovery and communication protocols, for the provision of access to the environment capabilities or for the management interfaces of the underlying resources.

5. National Development and Integration Strategy

Starting from the scenario reported in section 2.3 – Integration between INFN DataCloud and AlmaHealthDB – the consortium is developing a nationwide integration strategy by leveraging synergies among the ICSC, AHDB, and DARE projects. The overall objective of the strategy is to create a common approach for clinical data processing, in terms of standards, formats and exposed services.

On one hand, the use of ICSC infrastructures requires understanding when personal data is involved and how to select the most appropriate resources to comply both with legal and computational capacity requirements, on the other hand, DARE requires developing digital tools for primary, secondary, and tertiary prevention on large amount of data, e.g. for public health monitoring, data management models for particular populations, and continuous monitoring of citizens and patients for population-based interventions.

Therefore, the national development and integration strategy aims to set up certified and appropriately configured tools to enable the use and reuse of clinical data for biomedical research purposes within the limits dictated by current laws.

5.1. Technical Activities in DARE

As part of the Solution Frameworks (**refer to the SP1 Scientific report of Reporting Period 4, RP4**) that DARE is developing, *Solution Framework #2 – Salus Ratio* – that will ultimately implement a GDPR-compliant AlmaHealthDB as a Service, to be made available on certified local and cloud services. The work must adhere to the processes defined in the Information Security Management System (ISMS) of ICSC, including risk analysis, secure software development, and change management. In ICSC, resources from both INFN and CINECA will be exploited:

- ISMS of EPIC Cloud for INFN resources;
- ISMS of CINECA Cloud for CINECA resources.

5.2. Use Cases

Digital Twin on HPC resources Use Case

- Ansys Structures and PrepPost Suite (Mechanical APDL and ICEM CFD); the workflow has been tested on virtually all versions of Ansys from 2019R2 (v192) onwards
- Python (preferably Anaconda/Miniconda):
 - numpy
 - scipy
 - pandas
 - numba

- o itk

License server for Ansys that needs to be connected to start and maintain simulations. Ansys runs only on Intel or AMD processors, not on PowerPC architectures. It is available for Windows and some Linux distributions: RHEL (we have tested CentOS and Rocky Linux clones), SUSE (not OpenSUSE), and Ubuntu (from the penultimate version of Ansys or so, to compensate for the end of CentOS).

The simulation of each virtual patient requires about 50 core-hours in total, divided into 28 independent jobs each using about 30 GB of RAM and 4-6 cores (20-30 minutes) with MPI parallelism (IntelMPI or OpenMPI) and produces about 5-10 GB of Ansys binary files. Almost all of these files are deleted at the end of each simulation, and the results saved for each patient are a total of 20 MB of text files.

The workflow implementing the Digital Twin application will run both on HPC resources with SLURM scheduler in CINECA and on the INFN cloud platform with Kubernetes and Nextflow.

Radiomics/Deep Learning Use Case

The current implementation hypothesis for exploiting GPU-based resources is to develop deep learning applications, such as radiomics pipelines, based on ClearML Server, a client-server tool used to manage experiments. Pytorch is required for running the ClearML client. The following software packages are required to use GPUs:

- Cuda Toolkit 12.1
- cuDNN 8902
- Python 3.11.9
- clearml-agent 1.8.0
- torch 2.3.0
- clearml 1.15.1
- nvidia-cublas-cu12 12.1.3.1
- nvidia-cuda-cupti-cu12 12.1.105
- nvidia-cuda-nvcc-cu12 12.3.107
- nvidia-cuda-nvrtc-cu12 12.1.105
- nvidia-cuda-runtime-cu12 12.1.105
- nvidia-cudnn-cu12 8.9.2.26
- nvidia-cufft-cu12 11.0.2.54
- nvidia-curand-cu12 10.3.2.106
- nvidia-cusolver-cu12 11.4.5.107
- nvidia-cuspars-cu12 12.1.0.106
- nvidia-nccl-cu12 2.20.5

- nvidia-nvjitlink-cu12 12.3.101
- nvidia-nvtx-cu12 12.1.105
- virtualenv 20.26.1
- monai 1.3.0

The worker nodes should have direct internet access or enable a connection to external license servers.

5.3. First Phase and short-term goals

The first phase of the strategy is the secure deployment of Salus Ratio. based on the AlmaHealthDB application layer, on an EPIC tenant

- Defining and configuring the tenant on EPIC Cloud, including necessary resources;
- Defining user groups to be configured on Keycloak;
- Deploying Keycloak with secure configuration, multitenancy support, and enabling two factor authentication;
- Deploying REDCap in a secure and highly reliable configuration. Determine necessary security configurations. The MySQL database connected to REDCap should be installed on a separate machine within the protected network, so that REDCap resides on a machine with a public IP, and the database is on a non-exposed machine;
- Configuring single sign-on for VPN, REDCap, and any other installed applications.

Applications currently installed on AlmaHealthDB

Functional Applications

- REDCap
- MySQL
- minIO
- sftpGO
- PostgreSQL
- Python (installed by default on all Linux systems)
- Django (Python library)
- XNAT
- R Studio

Service Applications

- Apache Server in front of REDCap and SFTP for HTTPS
- SSHGuard
- Postfix (mail server, a mail server is necessary for REDCap to send emails)

In this first phase of the strategy, will result in a prototype to be tested by the end of the project. As reported in Figure 3, the data is managed by data processors and controllers (on the left side), within a certified and authorized infrastructure such as AHDB, while the INFN and/or CINECA cloud services within ICSC only receive/transfer anonymous derived data

- All the services are exposed by AHDB/Salus Ratio (data ingestion, processing, transformation, and anonymization);
- A transfer mode (including authentication modalities) is defined, allowing the data to be shared externally (ICSC).

The expected result in DARE is to replace AHDB with a first version of Salus Ratio, deployed in any certified IaaS, integrated with INFN cloud resources. Requirements for Salus Ratio are:

- To integrate a workflow manager;
- To integrate authentication and exchange modalities (formats, standards) agreed with ICSC

There are two possible modalities for the workflow management:

1. *Dynamic management*: ICSC receives anonymized data and decides how to sort data to be processed by the different entities within ICSC
2. *Static management*: All the management is in charge of AHDB/Salus Ratio, which decides to which entities to send data to.

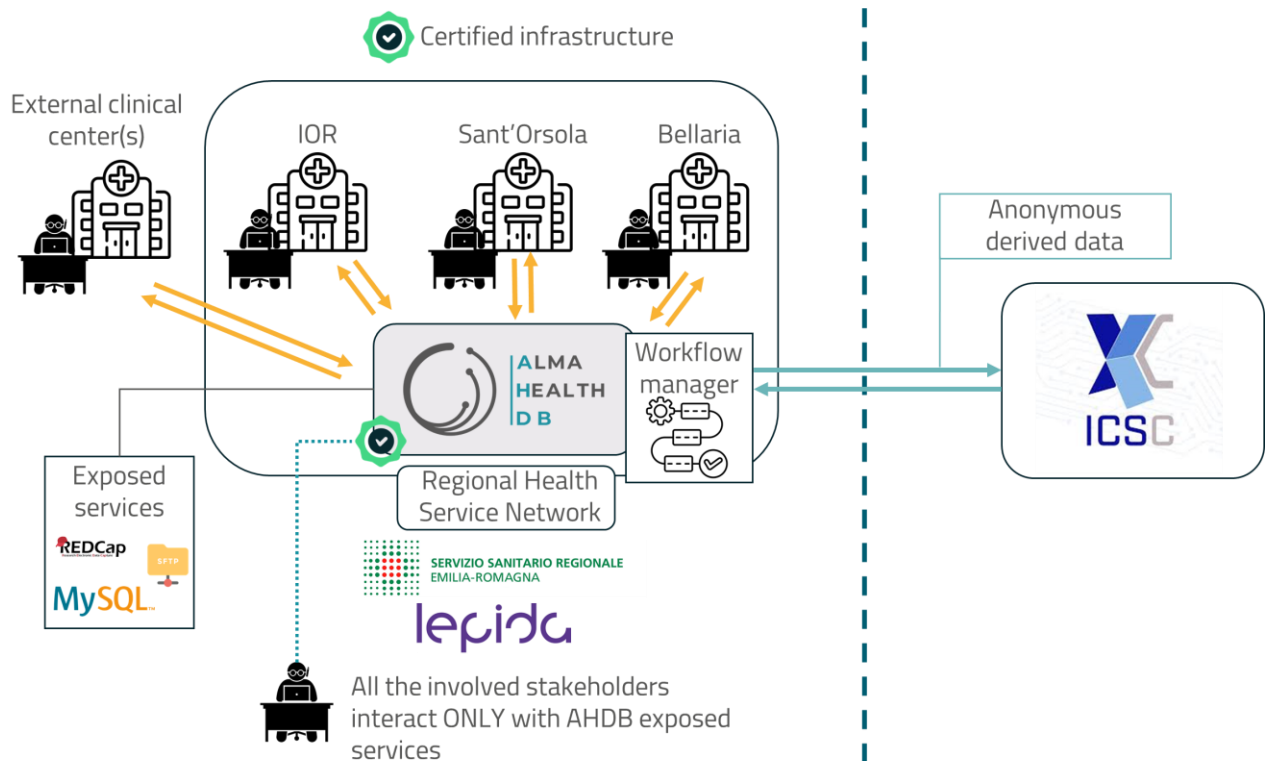


Figure 3 - Scenario and data flow for the first phase of the national development and implementation strategy.

5.4. Second phase and long-term goals

The second phase of the strategy, which will go beyond the duration of the DARE project, will require Salus Ratio as a Service to be made available on the ICSC cloud. The start of the second phase depends on the completion of two activities by ICSC:

1. Availability of an orchestrator in a GDPR-compliant version;
2. IAM (Identity and Access Management) in a GDPR-compliant version.

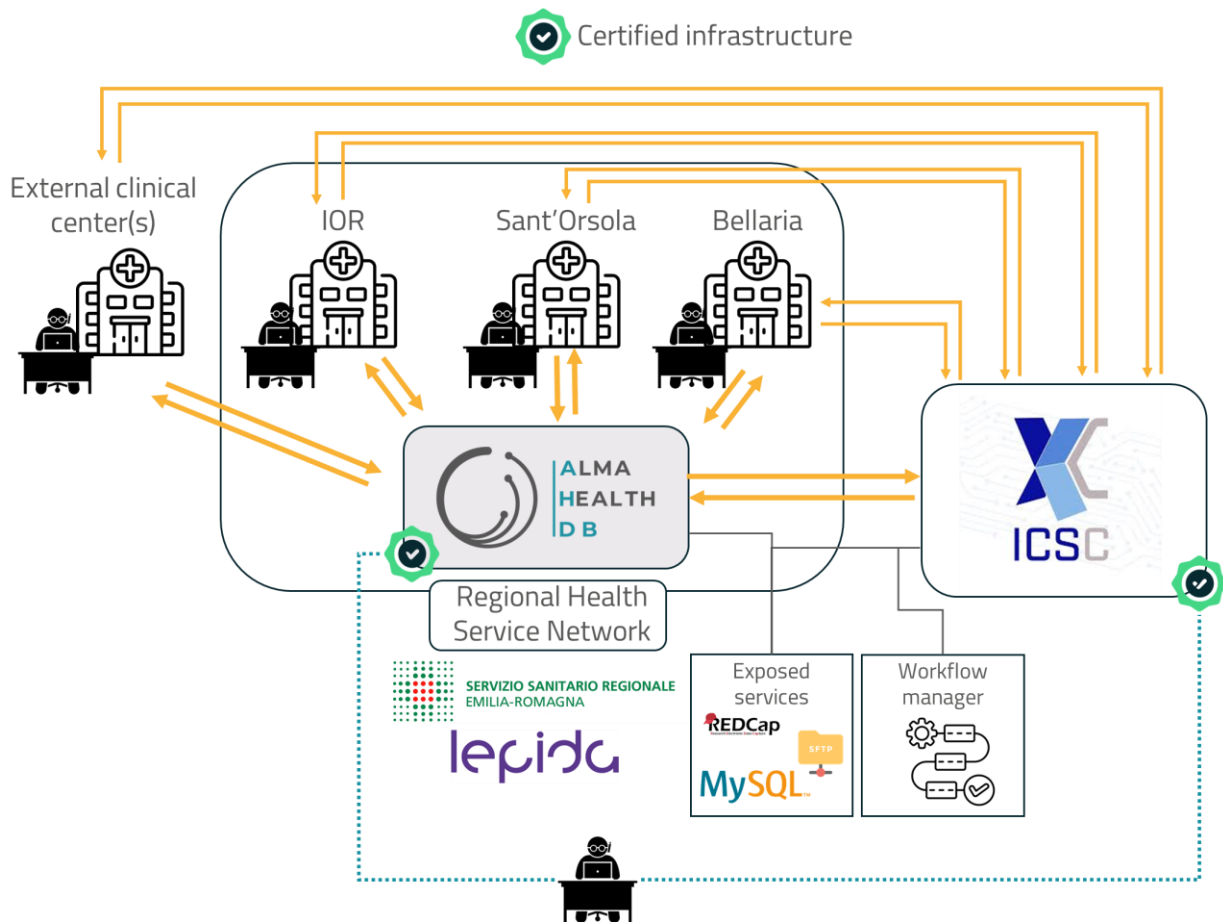


Figure 4 - Scenario and data flow for the second phase of the national development and implementation strategy.

In the second phase, as reported in Figure 4, the ICSC infrastructure, or at least part of it, becomes a certified infrastructure so that sensitive data can be sent to and be managed by ICSC. Hence, no more rigid separation between AHDB/Salus Ratio and ICSC since ICSC entities can be data processors too. Services can be exposed both by AHDB/Salus Ratio and ICSC. In this scenario, the first implementation of the workflow manager is expected to be static, but the long-term objective is to evolve towards a more dynamic management,

5.5. Update on implementation activities described in the short-term plan (section 5.3)

Over the reference period, we have planned the activities required for the integration of AlmaHealthDB/Salus Ratio services with those provided by INFN and CINECA, ensuring seamless interoperability between the two platforms. The plan details the high-level plan described in section 5.3. Furthermore, we have initiated the testing phase of the pipeline for data ingestion and standardization, aligning with the project's outlined objectives. To

this end, we agreed on a detailed action plan that involves implementing a centralized authentication system using Keycloak to ensure secure and unified access to the organization's software and hardware resources.

Keycloak has been deployed on an INFN DataCloud tenant, presently we are in the process of identifying all applications that need authentication and authorization, as well as cataloguing resource providers, including both software and hardware components.

In the next months, a gap analysis will be conducted to determine if these resources require additional integration for compatibility with Keycloak.

Subsequently, Keycloak will be installed and securely configured to establish a functional test environment. This setup includes defining external Identity Providers (IdPs) to enable federation within Keycloak. The configuration will be validated with the INFN security team to ensure compliance with EPIC security policies.

Following validation, any necessary integrations identified during the gap analysis will be developed. Finally, Single Sign-On (SSO) will be configured for the resource providers, utilizing the validated Keycloak setup to streamline and secure user access across the organization's resources. This structured approach ensures a robust and compliant authentication system, enhancing security and user experience across the organization's infrastructure.

These activities collectively form the foundation of our implementation strategy, facilitating robust authentication, authorization, and data standardization infrastructure. The collaboration across institutions such as UNIBO, INFN and CINECA is instrumental in this phase, with all stakeholders actively contributing to the defined acceptance criteria and timelines.

6. DARE-ICSC Agreement

DARE and ICSC are working on a joint research agreement to use their resources and skills to promote joint research also in the industrial field in computational sciences, high-performance computing, big data and quantum computing oriented to biomedical sciences, with a focus on the promotion of public health and disease prevention; collaborate in the design and delivery of training activities; develop, also with collaborative projects and with the promotion of synergies between activities already in place, an overall framework that promotes a new data culture (digital and health literacy); organize workshops and interdisciplinary and interinstitutional continuing education programs on digital and medicine; collaborate on the definition of infrastructural, organizational, ethical and techno-legal aspects for the processing of sensitive data; promote support for the clinical use of bioinformatics solutions and in silico medicine; organize joint events for the dissemination of the results of the collaboration.

7. Bibliography

- [1] <https://assets.innovazione.gov.it/1634299767-strategiaclouden.pdf>
- [2] <https://build.fhir.org/mapping-language.html>
- [3] <https://www.matchbox.health/>
- [4] <https://www.iso.org/standard/82758.html>
- [5] <https://www.enisa.europa.eu/publications/cloud-computing-risk-assessment>
- [6] ISO/IEC 19941
- [7] Deliverable 4.1
- [8] OASIS TOSCA Templates <https://docs.oasis-open.org/tosca/TOSCA/v2.0/TOSCA-v2.0.html>
- [9] JSON <https://www.json.org/json-en.html>
- [10] DICOM <https://www.dicomstandard.org/about-home>
- [11] OHDSI OMOP <https://ohdsi.github.io/CommonDataModel/>
- [12] HL7 FHIR <https://www.hl7.org/fhir/>
- [13] Frid S, Pastor Duran X, Bracons Cucó G, Pedrera-Jiménez M, Serrano-Balazote P, Muñoz Carrero A, Lozano-Rubí R. An Ontology-Based Approach for Consolidating Patient Data Standardized With European Norm/International Organization for Standardization 13606 (EN/ISO 13606) Into Joint Observational Medical Outcomes Partnership (OMOP) Repositories: Description of a Methodology. *JMIR Med Inform.* 2023 Mar 8;11:e44547. doi: 10.2196/44547. PMID: 36884279; PMCID: PMC10034609.