AFRICAI-RI: A Pan-African Research Infrastructure for Collaborative Biomedical Imaging and Artificial Intelligence

Deliverable D2.5:

Study Initiation Package: Infrastructure Mapping

Reference	D2.5_ AFRICAI-RI_EMC_v1
Lead Beneficiary	EMC
Author(s)	Mahlet Birhanu
Dissemination level	Public
Type	Report
Official Delivery Date	3 December 2025
Date of validation of the WP leader	27 November 2025
Date of validation by the Project Coordinator	2 December 2025
Project Coordinator Signature	

AFRICAI-RI is funded by the European Union's Horizon Europe Framework
Under Grant Agreement No 101190731

Version log

Issue Date	Version	Involved	Comments
25 Nov 25	V0.1	Mahlet Birhanu (EMC)	First version.
26 Nov 25	V0.2	Aisha Goedhart (EMC)	Version reviewed.
2 Dec 2025	V1	Cecilia Kessler, Lidia Garrucho Moras, Karim Lekadir (UB)	Revised and corrected final version.

Table of Contents

Acronyms

Name	Abbreviation
Sub Saharan Africa	SSA
Artificial Intelligence	AI
Findable Accessible Interoperable Reusable	FAIR
Point of Care Ultrasound	POCUS
Digital Imaging and Communications in Medicine Format	DICOM
Graphic Processing Unit	GPU
Local Area Network	LAN
Wide Area Network	WAN
Uninterruptible Power Supply	UPS
Health Information System	HIS
Picture Archiving and Communication Systems	PACS
Radiology Information System	RIS
Electronic Medical Records	EMR

Executive Summary

AFRICAI-RI is an initiative aimed at establishing the very first federated, multi-institution imaging data infrastructure for Africa, addressing a critical gap in the continent's biomedical resources using Artificial Intelligence (AI). This study initiation package marks the first of many steps in a multi-year effort to build a sustainable, interoperable, and FAIR-aligned imaging data ecosystem across partner Sub Saharan African (SSA) institutions. By laying the groundwork for coordinated infrastructure and workflow assessments, this package ensures that the project begins with a clear, accurate, and shared understanding of existing capacities and challenges.

Medical imaging data plays a pivotal role in clinical decision-making and in the development of AI-driven diagnostic tools. Yet across many multi-centre research settings, imaging systems and data workflows remain fragmented, inconsistently documented, or constrained by limited and/or varying infrastructure. Although the degree to which these issues are encountered can vary with participating institutions, it is safe to assume that they can hinder progress if realistic expectations and strategic planning are not set ahead of time.

Additionally, understanding current workflows and infrastructure is foundational for designing federated architectures that are technically feasible, ethically compliant, and tailored to local realities. By revealing bottlenecks, interoperability challenges, and areas where automation or standardization could dramatically improve data utility, this phase supports the creation of a long-term roadmap. It aims to ensure that subsequent phases (ranging from infrastructure deployment to federated analytics and AI integration) are informed by real-world evidence, guided by partner needs, and built upon a solid foundation of shared understanding.

To this end, the study initiation phase launches a systematic plan to map technical infrastructure (including PACS, servers, storage, networks, cybersecurity, and data governance structures) and operational data workflows (from image acquisition to annotation, archiving, analysis, and sharing). There will be 10 participating centres from different SSA countries. The methodology section explains each requirement in detail and reports the strategy to gather data from each participating SSA centre. Additional standardized tools planned to be used during surveys and data mapping are also mentioned in later sections. This document also covers the expected gaps, a strategy on how to inventories metadata standards, and references a hardware recommendation document that is currently under review by the project's infrastructure working group.

As the starting point of AFRICAI-RI's multi-year vision, this document should be considered as one part of a bigger study initiation package that includes stakeholder and community engagement (D1.4) as well as imaging and clinical data inventory.

1 Objective

This study initiation package aims to establish a strong and consistent foundation for the AFRICAI-RI infrastructure and data workflow mapping activities. Its specific objectives are to:

1. **Ensure Consistent Data Collection Across All Partner Institutions**
   Develop and distribute standardized templates, tools, and guidelines so that every participating institution collects infrastructure and workflow information in a uniform manner. This consistency is essential for generating comparable insights and enabling cross-site analysis.

2. **Provide a Clear, Comparable, and Auditable Methodology**
   Define a transparent and reproducible assessment methodology that outlines how data should be gathered, validated, and reported. This ensures that the mapping process is robust, traceable, and aligned with internationally recognized best practices.

3. **Establish a Common Understanding of Terms, Concepts, and Definitions**
   Introduce shared terminology for infrastructure components, data workflows, metadata standards, and governance processes. This harmonization prevents misunderstandings and supports coherent communication across multidisciplinary and multi-country teams.

4. **Align Partner Institutions Around Common Expectations**
   Clarify the scope, roles, responsibilities, deliverables, and timelines associated with the mapping exercise. This alignment ensures that all partners understand what is expected at each stage and promotes coordinated progress throughout the project.

2 Methodology

This section is structured around two core components: Infrastructure Assessment and Data Workflow Assessment. Together, these sections provide a comprehensive approach to understanding the technical, operational, and organizational foundations required to establish a federated imaging data infrastructure across partner institutions within AFRICAI-RI.

Each component includes clear definitions of key terms, explanations of required inputs, and standardized criteria to ensure a shared understanding across all participating sites by explicitly outlining concepts.

2.1 Infrastructure Assessment

In the context of AFRICAI-RI, infrastructure refers to the complete set of physical, digital, and organizational systems that enable the acquisition, storage, management, and use of medical imaging data. This includes the full spectrum of imaging equipment and modalities such as stationary and portable/Point-of-Care (POC) ultrasound machines, X-ray systems, and any associated acquisition hardware. These components form the foundation of clinical imaging services and determine the types, quality, and volume of data that can be generated at each site. Assessing these modalities not only reveals the technological capabilities of partner institutions but also highlights variability in equipment age, maintenance status, vendor support, and imaging protocols, all of which influence the reliability and standardization of datasets.

Infrastructure also encompasses the broader IT and network ecosystem required to support imaging workflows and enable secure, efficient data management. This includes Picture Archiving and Communication Systems (PACS), reporting systems, servers, storage devices, backup solutions, network connectivity, cybersecurity measures, and policies governing data access and governance. These elements determine how images move from acquisition to archiving, how they are shared within and across institutions, and whether they can support federated storage for AI-based analysis. Together, these combined layers of hardware, software, and connectivity define the institutional capacity to participate in a durable, multi-institution imaging data ecosystem.

2.1.1 List of Requirements

Here is the list of requirements that are covered by the infrastructure assessment. The information pertaining to each of the concepts defined below will be gathered from each centre by using the standardized tools explained in section 4. It is probable that some institutions might not have all the following information, but this document should serve as a baseline mentioning all relevant information to be gathered.

Imaging Equipment and Modalities

- **X-ray Systems (number, type, age, maintenance)**
  This refers to how many X-ray units were used to image patients, whether they are digital or analogue, their age, and maintenance status to assess imaging capacity and data quality reliability.

- **Ultrasound Systems (number, type, usage contexts)**

Documenting the quantity, portability, and clinical use cases of ultrasound devices reveal workflow diversity and the potential volume of imaging data generated.

- **Connectivity and DICOM Capability**
  Evaluating whether devices can export images digitally—especially in Digital Imaging and Communications in Medicine Format (DICOM)—determines their interoperability with PACS and data integration systems.

- **Preventive Maintenance and Calibration Schedule**
  Regular maintenance information indicates equipment reliability, uptime, and compliance with safe clinical operation standards.

- **Service Contracts and Technical Support Arrangements**
  The presence of vendor or third-party support contracts shows how quickly technical issues can be resolved and how well equipment can be sustained long-term.

IT and Network Infrastructure

- **PACS/RIS Systems (vendor, version, functionality)**
  This refers to the PACS and RIS (Radiology Information System) used for image archiving and reporting.

- **Server and Storage Capacity (local, centralized, or cloud)**
  This refers to the availability of a central server or the use of a cloud virtual machine for storage of medical images.

- **Network Capacity (LAN/WAN bandwidth, uptime, latency)**
  Network performance influences image transfer speeds, remote access, and the feasibility of participating in federated data infrastructures.

- **Backup, Redundancy, and Disaster Recovery Systems**
  These systems ensure data protection, continuity of operations, and resilience against hardware failures or outages.

- **Hardware Resources for AI or Data Processing (GPUs, workstations)**
  Understanding available compute resources shows the institution's readiness for AI workloads, image processing, and advanced analytics.

- **Power Supply and UPS ((Uninterruptible Power Supply) Systems**
  Reliable power and backup units are essential to prevent equipment damage, data loss, and workflow interruptions in environments with unstable electricity.

2.1.2 Data Collection

The data collection process for this assessment is designed to be structured, and consistent across all partner institutions. To achieve this, a standardized Excel-based data collection template will be provided, outlining all requirements defined above. The template includes predefined categories, definitions, and instructions to reduce ambiguity and ensure that each site reports information in a uniform manner. Sites will be asked to complete the Excel sheet

using a combination of local records, system logs, and consultations with technical and clinical staff. The infrastructure group has already set up this template (see Annex Section A) and collected input from many sites although some of the input will still need to be verified.

To ensure accuracy and contextual understanding, all submissions will undergo a validation phase consisting of follow-up interviews with IT personnel at each institution. These interviews will clarify any ambiguous entries, verify critical details, and capture nuances that are not easily represented in a spreadsheet. This two-step process combines the efficiency of structured data collection with the depth of qualitative validation, ensuring that the final dataset reflects both quantitative metrics and local realities. The result will be a robust and reliable baseline assessment that can confidently inform planning, capacity building, and future infrastructure development within AFRICAI-RI.

2.1.3 Analysis of Gaps and Recommendation

After the two step data gathering and verification step is complete, the infrastructure working group will take a close look to analyse the gap. It is expected that different institutions will have different hardware types and systems. Based on this, a hardware recommendation document will be drafted by the working group highlighting the necessary equipment to be purchased by each SSA centre with the project budget allocated. This recommendation document (see Annex Section A), which is already in progress, will then be shared with responsible stakeholders to proceed to the next step. Hence, the outcome of this assessment will be a hardware recommendation document from the working group followed by a purchase status/report from the partner SSA centres.

2.2 Data Workflow Assessment

Data workflow refers to the end-to-end pathway that medical imaging data follows from the moment it is generated to its eventual storage and its use in research. This begins with data acquisition, which involves how images are produced on X-ray or ultrasound, or other modalities, including the protocols used, image formats generated, and whether devices automatically transmit data to a PACS or rely on manual transfer methods. The acquisition stage determines not only the immediate clinical utility of the images but also the consistency and completeness of metadata, both of which are essential for interoperability and future use in research or AI development.

Following acquisition, data storage, protection, and anonymization become central components of the workflow. This includes how images are archived in local servers or PACS, the capacity and structure of storage systems, the reliability of backups, and the security measures that ensure compliance with ethical and legal requirements. Anonymization or de-identification processes—whether automated or manual—play a critical role in preparing data for research, safeguarding patient privacy while enabling safe reuse. The broader workflow may also involve data retrieval, image annotation, quality control, and data export for federated analytics or cross-site studies.

2.2.1 List of Requirements

Here is the list of requirements that are covered by the data workflow assessment. The information pertaining to each of the concepts defined below will be gathered from each centre by using the standardized tools explained in section 4.

Medical Imaging Data Acquisition and Transfer

- **Integration Between Imaging Devices and PACS/RIS**
  Assesses if the scanners can properly connect to the PACS/RIS for automated data flow.

- **Manual vs. Automated Data Entry Steps**
  Identifies where staff must manually input information versus where data is automatically captured or transferred.

- **Frequency and Speed of Data Transfer**
  Measures how often images are sent to storage systems and how quickly transfers occur, affecting workflow efficiency.

- **Quality Assurance at Acquisition**
  Refers to checks performed during imaging to ensure data accuracy, completeness, and adherence to protocol.

- **Storage Backup Routines and Retention Policy**
  Covers how often data is backed up and how long images are kept according to clinical, legal, or research requirements.

- **Archiving Protocols for Research**
  Describes specific procedures for storing and preparing imaging data intended for research reuse.

- **Data Retrieval Processes and Access Frequency**
  Explains how images are accessed for clinical or research purposes and how often/if retrieval occurs.

Data standards, interoperability and governance

- **Use of DICOM Standards and Conformance Level**
  Assesses how consistently devices and systems follow the DICOM standard for image formatting and communication.

- **Integration With Hospital Information Systems (HIS/EMR)**
  Evaluates how imaging systems connect with electronic medical records to exchange patients and study information.

- **Use of Metadata Models (MIABIS, DCAT-AP, FHIR)**
  Determines whether advanced metadata standards are used to describe, catalogue, and exchange imaging data.

- **Existing APIs or Interfaces for Data Exchange**
  Identifies whether systems support secure, programmable interfaces that enable interoperability and automated workflows.

- **Data Ownership, Consent, and Sharing Policies**
  Reviews institutional rules governing who controls imaging data and under what

conditions it may be shared.

- **Compliance With GDPR and National Data Protection Laws**
  Ensures that data handling practices meet legal requirements for privacy, security, and patient rights.

2.2.2 Data Collection

For data workflow assessment, the data collection will be done in a multistep approach. The first step will be designing a standard excel based template or questionnaire on google form to be filled in by the data managers in participating SSA clinical sites. This template will include all the requirements listed above with instructions to avoid ambiguity. Next, a first round of interviews will be conducted with principal investigators (PIs) and radiologists to confirm the inputs and to cover potential gaps missed. The output of the interview will be combined with the previous assessment interviews (Infrastructure assessment with IT managers) for further clarification. Another round of interviews can also be planned with other stakeholders if there is a need for it.

2.2.3 Analysis of Gaps and Recommendation

After completing the round of interviews with responsible parties, the infrastructure working group will analyse the standard excel inputs for potential gaps. Based on the gap analysis, a data management plan will be drafted with our recommendations. The data management plan will outline the concrete tasks that will follow to implement federated infrastructure setup in the AFRICAI-RI nodes. Additionally, a software management plan (SMP) will be drafted that should contain necessary instructions on recommendation of software usage.

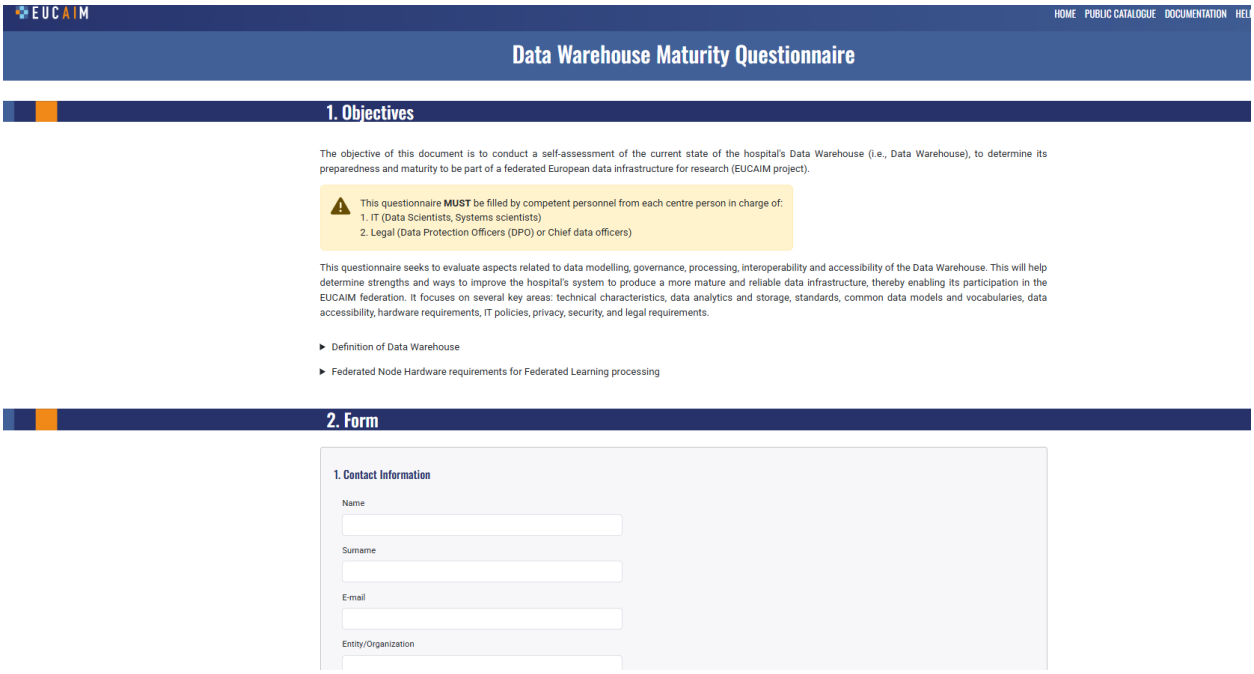
3 Standardized Assessment Tools

In many EU funded multi-year projects, many standardized assessment tools are used to ensure consistent and comparable data collection and reporting. These tools typically include structured questionnaires, interview guides, scoring frameworks, and standardized checklists.

As mentioned in previous sections, we plan to use standardized excel sheets and semi-structured interview guides as we find them more fitting to the type and scope of the work. For data workflow analysis, we plan to design structured questionnaires on Google forms by using EUCAIM's data maturity questionnaire as a reference standard.

Example of the EUCAIM's questionnaire:

https://dashboard.eucaim.cancerimage.eu/data-warehouse-maturity-questionnaire



4 Timeline

The following table summarizes the list of tasks mentioned in this study initiation report and indicates their estimated month of completion. It should be noted that there is currently a budget approval delay affecting employment of man power in all the clinical sites. This means that many if not most of the listed tasks might be significantly delayed if the issue is not solved in a timely manner.

Table 1: Timeline for the completion of list of study initiation package tasks during the first year (M12)

	1	2	3	4	5	6	7	8	9	10	11	12
Data collection for hardware assessment												
First round of interviews with IT personnel												
Data collection for workflow assessment												
Additional round of interviews												
First draft of Data management Plan (DMP)												
Architectural design and Software Management Plan (SMP)												

5 GDPR and Data Protection

This deliverable does not involve the processing of personal data. All activities described herein relate exclusively to the assessment of institutional infrastructure and technical workflows at participating centres. No clinical data, patient identifiers, or operational logs containing personal information were collected or accessed as part of this study initiation package.

The AFRICAI-RI project follows the principles of the EU General Data Protection Regulation (GDPR), including purpose limitation, data minimisation, integrity, confidentiality, and accountability. Any future activities involving data flows, metadata inventories, or federated learning configurations will ensure that identifiable information remains entirely within each participating clinical site. Only pseudonymised or anonymised metadata may be shared across sites when strictly necessary and in accordance with approved data governance policies.

A full and detailed Data Management Plan (DMP), together with a Software Management Plan (SMP), will be developed as part of subsequent project phases. These plans will describe applicable legal bases, roles and responsibilities of data controllers and processors, retention periods, data security measures, documentation requirements, and local compliance with national data protection frameworks. No cross-border transfer of identifiable data will occur at any stage of the project.

6 Conclusion

This study initiation package establishes the methodological and operational foundation necessary to conduct a harmonised infrastructure and workflow mapping exercise across ten AFRICAI-RI partner institutions. By defining standardised assessment tools, common terminology, and clear data collection procedures, this deliverable ensures that all participating sites will follow a consistent and comparable approach. The outputs of this initial phase, including validated infrastructure inventories, workflow assessments, and subsequent gap analyses, will directly support the design of the federated architecture and inform both hardware and software procurement decisions.

Through this work, AFRICAI-RI strengthens its capacity to develop the first pan-African, federated imaging ecosystem aligned with FAIR principles and local realities. The insights generated during this initiation phase will guide the implementation of secure, interoperable, and sustainable data infrastructures in the coming project stages, laying the groundwork for long-term research collaboration and AI-enabled innovation across SSA.

7 Annexes

Section A: Standardized excel sheets for hardware assessment

Site	Country	IT Manager	Installation Responsible	email	Equipment Budget	Type of connection	Internet Connectivity	Possibility of improvement	Server	Dedicated room
MRCG	Gambia	Badou Gaye		bgaye@mrc.gm	€37,000.00	Optical Fiber	>200 Mbps 1Gbps	☐	Rack	Yes
CSIM	Mozambique	Joao Gemo Novele		joao.novele@manhica.net	€13,500.00	Optical Fiber	50-100 Mbps	☑	Rack	Yes
KUHES	Malawi	Vincent Kantunga Phiri		vincentkatungaphiri@gmail.com	€20,000.00		50-100 Mbps	☐		
IDRC	Uganda	Alan Asiimwe		aasiimwe@idrc-uganda.org	€20,000.00	Optical Fiber	Stable 20 Mbps	☑	Rack	Yes
JU	Ethiopia	Aboma Balcha		balchaaboma74@gmail.com	€20,000.00	Optical Fiber	Stable 1Gbps	☐	Rack	
CERMEL	Gabon	Dr Bertrant Lell; Anthony Mintsa Menie; Dani Nzinga	Anthony Mintsa Menie	bertrand.lell@cermel.org; anthony.mintsa@cermel.org; dani.nzinga@cermel.org	€20,000.00	Optical Fiber	Stable 1Gbps >200 Mbps	☑	Rack	Yes
UNIKIN	Congo	Primo Kumbulu		kimbuluprimo@gmail.com	€20,000.00	Satellite	>200 Mbps	☑	PC/Workstation	Yes
UIDT	Senegal	Mamadou Bousso		mbousso@univ-thies.sn	€20,000.00	Optical Fiber	10 Mbps	☑	Rack	Yes
NIMR	Nigeria	David Taiwo	David Taiwo	davtaiwo@gmail.com	€20,000.00	Optical Fiber	10 Mbps	☑	Rack	Yes
DIG	Ghana	Oswald Osei		oosei@delft.care	€0.00	Optical Fiber	50-100 Mbps	☑	Rack	Yes
UB	Spain				€5,000.00			☐		
FORTH	Greece				€5,000.00			☐		
EMC	Netherlands				€2,500.00			☐		
ISGLOBAL	Spain				€1,000.00			☐		

	IDRC - Uganda
Specifities of the local data storage system	**Main server:** Dell PowerEdge R730 Windows Server 2016 Standard Processor: 12 Cores, 24 Logical Processors RAM: 160 GB RAM HDD: 11 TB Backup server: Dell PowerEdge T440 Windows Server 2016 Standard Processor: 16 Cores, 32 Logical Processors RAM: 32 GB RAM HDD: 8 TB
Purchase of your last computer	N/A
Internet connectivity	Stable internet connectio with symmetric speeds of 20 Mbps for both download and upload
Computational resources	We only have the servers listed above, but we can allocate a portion of their resources for machine learning training tasks.
Procedures for data access	Our data is stored on our local servers and is accessible via a password protected data management website. A data sharing agreement wold have to be agreed upon before accessing the data.
Computer with Linux?If not, pecify if you have on with Window	Planning to set up Linux VM
IT manager	Alan Asiimwe aasiimwe@idrc-uganda.org

Section B: Infrastructure recommendation document

Central Hub

Component	Minimum / Recommended Specification	Upgrade / Advanced Option	Approx. Cost (€)	Justification / Purpose
CPU	AMD EPYC 9354 (32 cores) or Dual Intel Xeon Gold 6330	Dual EPYC 9454	6 000	High-core CPUs handle concurrent FL coordination and data processing.
RAM	256 GB ECC DDR4/DDR5	512 GB	1 000 – 2 000	Large ECC RAM ensures smooth aggregation of multiple models and containerized services.
System Storage	2 × 2 TB NVMe SSD (RAID 1)	4 TB NVMe scratch RAID 0	300	Redundant NVMe ensures reliability and fast access.
Enterprise NAS (Storage)	12-bay NAS (e.g. Synology RS3621RPxs) + 12 × 8 TB HDDs (≈ 70 TB usable)	12 × 12 TB HDDs (≈ 100 TB usable)	4 000 – 5 000	Enterprise NAS provides secure and redundant central data storage in line with FAIR principles.
Network	10 GbE switch + fibre uplink	Redundant 10 GbE links	1 000	High-speed network ensures fast communication with nodes.
Rack UPS (SAI)	6 kVA Online Smart-UPS	10 kVA UPS with extended batteries	4 000	Rack UPS protects hub from power interruptions ensuring data integrity.
Cooling / Rack Accessories	42U rack + PDU + AC duct integration	-	1 000	Physical organization and airflow in data-centre environment.
Contingency (10%)	**Installation, shipping, customs buffer**	**-**	**3 000**	**Contingency covers installation, logistics, and customs costs.**
Total Estimated Cost		**-**	**33 000 – 36 000**	**Fits within 37 k€ central hub budget limit.**

Local Node Rack

Component	Minimum	Upgrade / Advanced Option	Approx. Cost (€)	Justification / Purpose
Form Factor	2U rack server chassis	4U multi-GPU server	-	Rack servers are suited for data-centres with cooling and monitoring systems.
GPU	1 × RTX A5000 (24 GB)	1 × RTX 6000 Ada (48 GB) or 2 × A5000	2 500 – 6 000	GPU enables high-performance AI computation; 24–48 GB VRAM supports large medical models.
CPU	Intel Xeon Silver 4314 / AMD EPYC 7313	Dual Xeon Gold / EPYC 9354	1 500 – 3 000	Server-grade CPUs ensure stability for 24/7 operation.
RAM	64 GB ECC	128–256 GB ECC	300 – 800	ECC RAM prevents memory errors critical for long training sessions.
Storage	1 TB NVMe SSD + attached NAS (24 TB)	2 × NVMe in RAID1 + larger NAS	1 500 – 2 000	High-speed NVMe and NAS offer redundancy and speed.
Network	1 GbE	10 GbE	200	Fast network supports model synchronization across multiple FL nodes.
Rack UPS (SAI)	3 kVA Smart-UPS Rack	6 kVA Online UPS	1 500 – 3 000	Rack UPS protects equipment and allows safe shutdowns.
Total per Node		-	8 000 – 12 000	Remains within 20 k€ per-node limit.

Local Node PC

Component	Minimum (Recommended)	Upgrade / Advanced Option	Approx. Cost (€)	Justification / Purpose
Primary Storage	1 TB NVMe SSD	2 TB NVMe SSD	80 – 160	Fast SSD enables rapid data access and OS stability during training.
Data Storage (NAS)	Synology DS923+ + 4 × 8 TB HDD (RAID5, ≈ 24 TB usable)	Larger NAS (6 × 8 TB ≈ 40 TB usable)	1 400 – 2 000	Local NAS provides safe redundant storage (RAID5) for 10 000 studies and ensures FAIR data management.
Operating System	Ubuntu 22.04 LTS	-	-	Ubuntu ensures open-source compatibility and long-term support.
Network	1 GbE Ethernet	10 GbE optional	100 – 200	Reliable connection for federated learning model exchange; 10 GbE optional for high throughput.
UPS (SAI)	APC Smart-UPS 1500 VA (line-interactive)	APC Online 3000 VA	400 – 800	UPS ensures stability in sites with frequent power cuts, preventing hardware damage.
Total per Node		**-**	**6 500 – 8 000**	**Overall node cost within 20k € limit.**