Improving cancer diagnosis and prediction with AI and big data

A Multimodal AI-based Toolbox and an Interoperable Health Imaging Repository for the Empowerment of Imaging Analysis related to the Diagnosis, Prediction and Follow-up of Cancer

INCISIVE Node

Procurement & Installation Guidelines

17/01/2023

Revision 1.1

Status: Final Version

Grant Agreement n 952179

European Commission

DOCUMENT CONTROL	
Project reference	Grant Agreement number: 952179
Document name	INCISIVE Node Procurement & Installation Guidelines
Work Package	WP 6
Work Package Title	INCISIVE pan-European repository of health images
Dissemination level	CO
Revision	-
Status	Final Version
Reviewers	-
Beneficiary(ies)	

Dissemination level:
PU = Public, for wide dissemination (public deliverables shall be of a professional standard in a form suitable for print or electronic publication) or CO = Confidential, limited to project participants and European Commission.

AUTHORS		
	Name	**Organisation**
Document leader	Paris Laras	MAG
	Gianna Tsakou	MAG
	Sotiris Athanasopoulos	MAG
	Alberto Gutierrez Torre	BSC

REVISION HISTORY				
Revision	**Date**	**Author**	**Organisation**	**Description**
0.1	22/07/2022	Paris Laras	MAG	First draft - Inserted Node requirements and core sections
0.2	5/08/2022	Paris Laras	MAG	Second Draft - Tweaked core sections, restructured guidelines
0.3	25/08/2022	Paris Laras	MAG	Third Draft - Added Security and Backup sections
0.4	29/09/2022	Paris Laras	MAG	Final Draft - Re-structured entire document using INCISIVE document format
1.0	8/12/2022	Paris Laras	MAG	Base Version - Add Updated Legal Requirements
1.1	17/1/2023	Paris Laras	MAG	Final Version – Document Validated

Table of Contents

1. Federated Node in INCISIVE

This document aims to provide INCISIVE users with guidelines for acquiring and installing dedicated hardware to operate as an INCISIVE Node for your organization.

Each Data Provider that participates in the INCISIVE federated data repository is called to provide infrastructure that will be able to store the data they will be contributing to the federated repository of INCISIVE.

This infrastructure (a powerful computer, server, or virtual machine) is connected to the INCISIVE platform and acts as a Federated Node in the system's architecture. This node is used to facilitate federated data sharing, train AI models and contribute to the INCISIVE platform with its data processing.

Data providers that are not able to obtain or host a Node within their organization's infrastructure are also able to use the Central Node of INCISIVE. This Node is the GDPR-compliant Central Data storage of INCISIVE, created and managed by the project to make the data within it centrally available for the platform and cloud services.

2. Procurement of INCISIVE Data Provider Node

2.1. Description

Procurement of the Federated Node has been agreed on a project level to be carried out on an individual-organization basis. Each organization is called to obtain the infrastructure needed to host a Federated Node in which their retrospective and/or prospective data will be stored and shared through data federation as part of INCISIVE's core innovation activity.

2.2. Federated Node Hardware Requirements

Below you will find a table outlining the Hardware requirements for a single Node as they have been defined so far in the context of the project. The requirements below are suggested to guarantee that the majority of INCISIVE use cases will be supported in terms of Node performance, and that no critical performance bottlenecks will occur during platform operations.

Hardware	Option 1	Option 2	Notes
CPU	Intel i9 12XXX or better (socket 1700) Minimum	AMD Ryzen 5900X or better (socket AM4) Minimum	• If a GPU is not present, a server-grade, high core-count CPU is necessary for the Second Prototype • If not comparable by cores, ideal thread count is 24+

	Cores: 16 >=1.8GHZ	Cores: 12 >=3.0Ghz	
RAM	64GB DDR5 for Intel	64GB DDR4 for either Intel or AMD	• DDR5 is ideal • ECC memory is highly recommended for stability • 128GB will be needed down the line, so 64GB should be in 2x32GB configuration
Motherboard	Socket 1700 for Intel 4+ Ram Slots	Socket AM4 for AMD 4+ Ram slots	• Make sure to double check compatibility of selected CPUs with the Chipset of the motherboard • In the case of DDR5, double check motherboard compatibility with DDR5
Storage	1TB SSD Drive for Operating System (Either NVMe M.2 PCI Gen4 or SATA III) 1TB++ SATA III Drive (SSD or HDD) for local storage of medical data		• M.2 NVMe Gen4 Drives are suggested for the OS • For data storage size, DPs are expected to plan their purchase depending on the size of the Data they will provide. 1TB is a minimum, with some DPs already planning for 2 TB + datasets • For data storage, SSD are preferred for speed but are not mandatory
Graphics card	NVIDIA Quadro	NVIDIA RTX 3XXX	• 12GB RAM+ is preferred • Maximizing the amount of Tensor Cores is a priority, most recent GPUs will generally have higher Tensor Core counts • Ampere and Volta architectures are preferred
Operating System	Linux		• Latest version of any mainstream Linux distribution is acceptable Ubuntu, Alpine or other • Windows is NOT acceptable, unless absolutely impossible for a DP to setup a Linux environment

Power Supply	-	Each DP must make calculations depending on the hardware setup that will be selected to make sure that needed Wattage is covered and ideally exceeded to prepare of any future upgrades to the machine
Internet	100mbps (baseline)	Each DP must make best efforts to provide the best possible connection to their Node. Network performance will directly affect node stability and can invalidate AI training or prevent successful demonstrations of the INCISVE platform

2.3. Virtual Machine Case

In the case of your organization chooses to deploy a Virtual machine internally to host the Node, through a server or third-party service, the above requirements continue to apply, and the appropriate Cores, RAM and GPU processing must be allocated to the virtual machine acting as the node.

2.4. Existing Hardware Case

In the case of existing hardware within your organization that might be used for hosting an INCISIVE Node that might not exactly fit the Suggested Hardware Requirements above, feel free to contact Maggioli to confirm if said hardware can meet the project's needs.

3. Federated Node Installation

3.1. Hardware Installation

Hardware for the Node in most cases will arrive already assembled into a ready-to-boot machine, either a desktop computer or in a server unit format.

If the hardware arrives as separate parts, Maggioli will support as much as possible remotely but eventually it is up to each organization to allocate time and resources to build the separate parts into the final machine.

The Node must be installed physically close to a network connection and connect to the network through an appropriate cable and not through a wireless connection. Its connecting USB and power cables should not hang over tables or lay in the way common footpaths or near machinery. Liquids, cleaning substances, food and other materials that might damage the Node's hardware should always remain away from the Node. The Node should be

installed in a secure, not publicly accessible location and away from any plumbing or electricity installations. Ideally, the node is installed in a controlled-access environment through usage of security passes, security cameras and other monitoring capabilities.

3.2. Operating System Installation

In most cases, the machine will arrive with a selected Operating system already installed. If this is not the case, Maggioli will support as much as possible remotely but eventually it is up to each organization to install the Operating System so the node can be configured and integrated.

While setting up the Operating System, please keep in mind that a user will have to be created for Maggioli to connect to the Node via SSH Remote Access.
Therefore, it is suggested that this designated user has a strong and secure password, using at least 12 if not more characters, as well as capitalized letters, numbers and any valid symbols.

Example Ubuntu installation

We currently suggest the installation of Ubuntu 20.04 (link) distribution of Linux in all Nodes. To install, download the appropriate image from the link and proceed with installation according to this guide.

Of the two image options (Desktop or Server image) select the desktop version if the machine will be used independently through a monitor screen within the organization, or in the case that it is deployed in a server configuration, feel free to install the Server version of 20.04, but please note that this will initially limit the node to only be used through a Command line interface.

4. Federated Node Configuration

4.1. Network Configuration

In order to operate, every Node must be able to connect the public internet.
Furthermore, unless your organization makes use of a VPN (see below) the Node **must** have a public & static IP address, which has been allocated by the organization in coordination with whichever entity provides the organization with Internet and/or networking.

4.2. Security Configurations

Finally, as it relates to Nodes, it is important for security that every organization takes steps to secure their Nodes and other infrastructure at any opportunity they can find. Your organization is the most suitable entity to decide which security measures will be used. As usual best practices applied to similar health data storage structures, we currently suggest that any DP connecting to INCISIVE should make use of Firewalls and VPN software, if possible.

4.3. Firewall Configuration

If your organization makes use of a Firewall, configurations will have to be made to allow traffic to ports 22 TCP and 53 UDP so the node is able to operate and connect to the system.

4.4. VPN (Virtual Private Network) Configuration

It is strongly advised for any Data Provider joining the INCISIVE platform and
If your organization makes use of a VPN to protect access to internal infrastructure, credentials to connect to the VPN will have to be provided to Maggioli.

These credentials can either be username/password or certificate files generated by your organization.

4.5. GPU Driver Configuration

Once a node has been configured and integrated, the Node's GPU needs to have driver software installed and enabled to allow for the GPU to be utilized in a Linux environment. To achieve this, Maggioli will be reaching out to technical partners with credentials. These credentials will be used locally when re-booting the Node to enable the GPU drivers through Secure Boot.

5. Node Integration

5.1. Install OpenSSH

For Maggioli to integrate your organization's node with the INCISIVE platform, remote access is required. This remote access is provided by installing **open-ssh.**

This can be achieved by the following series of commands:

```
sudo apt update
```

```
sudo apt install openssh-server
```

It is recommended to not execute these commands without some basic technical knowledge of Linux systems. If a person within your organization exists that is qualified in technical matters, we recommend that this installation is executed by them or with their supervision.

Once SSH has been installed and the service is running, the machine will now be able to be accessed by anyone that has access to the IP address of the machine & credentials for any user on the machine.

5.2. Provide SSH Access for Integration

Following this, you are requested to provide the IP address and user & password credentials to Maggioli. Maggioli will also remain available to assist remotely in these configurations if needed, but the extent of assistance we will be able to provide depends on the type of

hardware configuration that has been selected as well as the technical knowledge available on-site.

Maggioli will then continue with installing *Docker Engine*, the first step of Integration. Following that, multiple components will be downloaded and installed to the Node and configured, connecting it to the INCISIVE platform.

Once the Integration process has been completed and the green light has been given by Maggioli, the Node will be ready for use. This means that the node will be ready for data to be transferred and stored in it, as well as to receive training and/or inference jobs from the platform.

6. Data Redundancy

Organizations integrating a Node for INCISIVE will also proceed with storing the contributed data within the Node. To minimize the chance of data loss and to mitigate the impact of data being deleted or corrupted during the project and beyond, two suggestions are made:

6.1. Backup Policy

Firstly, the creation of a backup policy within your organization, allowing you to store an exact replica of the Node's data every 2 weeks (for example) in a physical (external) drive or a network drive.

6.2. Data Redundancy

We suggest the use of a RAID (Redundant Array of Independent Disks) Technology or similar for the Node's data storage drives. This technology creates a software-level replica of the Data by using twice (or more) times the storage drives and allowing your organization to have hardware drives fail or become corrupted, but still being able replace some of the drives in case of failure, without losing any data.

Please keep in mind that RAID configurations cannot be applied to populated drives. This means that RAID configurations must be done on empty storage volumes before they are filled in with data.

7. Legal Requirements

When setting up the federated nodes, the Data Providers need to follow their legal and organizational requirements for storing medical data. Any hardware and infrastructure used for processing of personal data must be checked for IT security and compliance with any other local rules. Data Provider should check with their legal department and/or Data Protection Officer (DPO) for further guidance. Data Providers are responsible for ensuring technical and organizational measures of privacy protection (including, but not limited to,

data access procedures, security breach reporting) of any data that they host in federated nodes.

Moreover, if the Data Provider plans to outsource the storage (hosting) of contributed patient personal data and images, or otherwise to use service providers that process data on their behalf, such Data Provider needs to comply with their own data protection requirements. In particular, Data Providers need to ensure:

i. appropriate contracts with the infrastructure/cloud providers and

ii. the security of the hosted data, in line with the recommendations of their IT, legal and DPO/security departments.

Any documentation regarding concluded contracts and security level should be kept on file and made available for Incisive, if required for project's privacy assessment purposes or the project's legal compliance.

For any EU based Data Providers, particular attention should be paid when making a decision to outsource the storage of the pseudonymized patient data to a cloud server (provided by a third-party provider), as it will have GDPR implications. In particular Data Providers should be aware that:

i. hiring a processor means that such processor should be verified and the agreement with them should include data processing annex and should be compliant with GDPR. Please check with your legal teams for any templates or conditions which may be applicable. Any templates from the processor should also be verified for legal compliance.

ii. hosting the medical data outside of EU would be considered a transfer of data to a third country.

In theory, it is possible to use providers which host data in the US (or other countries outside of EEA/EU), but this raises additional legal considerations. In particular, such hosting (or even access to data for technical reasons by the provider located outside of EU) would be considered a transfer of data to a third country, which requires additional safeguards, defined by the GDPR. For the US, those measures would mean putting in place a special agreement, so-called standard contractual clauses. Moreover, transfer impact assessment (TIA) would be required to determine if further security measures are needed. This exercise involves examining the legal systems in the US, with input from the provider, which will likely prolong the process. Once completed, the additional security measures most likely will include encrypting the data and keeping the keys in the EU as the condition for transfer. It would thus need to be verified if encrypted data may still be used in the project.

All in all, due to privacy considerations, it is recommended to choose a provider which hosts the data in the EU and does not access it from outside the EU (even for technical reasons). Even more, from a legal perspective, choosing an EU cloud provider, if one is available, is a better option.

For non-EU based Data Providers, there may be local legal requirements in place, thus please check with your legal department that your chosen solution is compliant with them.

8. Final Notes

8.1. Contact

Each Data Provider is recommended to designate a contact-person within their organization to manage their infrastructure locally when there is a need for it. Ideally this person can act on a technical level and if not, Maggioli will assist by providing guidelines and support for these matters.

If you have any questions whatsoever, we encourage you to contact us at Maggioli to discuss.

Gianna Tsakou

Gianna.tsakou@maggioli.gr

Sotiris Athanasopoulos
sotiris.athanasopoulos@maggioli.gr

Paris Laras
paris.laras@maggioli.gr