Ref. Ares(2025)1611740 - 28/02/2025

NOSTRADAMUS

D7.6 DATA MANAGEMENT PLAN (M6)

29/01/2025

Funded by
the European Union

PARTNERS

NOSTRADAMUS

Grant Agreement No.: 101134888
Call: HORIZON-CL6-2023-GOVERNANCE-01
Topic: HORIZON-CL6-2023-GOVERNANCE-01-13
Type of action: HORIZON Research and Innovation Actions

D7.6 DATA MANAGEMENT PLAN (M6)

Data management

Work package	WP 7
Task(s)	Task 7.4: Data management
Due date	28/02/2025
Submission date	28/02/2025
Deliverable lead	ERATOSTHENES Centre of Excellence (Acronym: ECoE)
Version	0.3
Authors	Thomaida Polydorou (ECoE - BE1) Ioannis Varvaris (ECoE - BE1) Diofantos Hadjimitsis (ECoE - BE1)
Reviewers	Tomaz Bokan (ITC) Nils Borchard (DLG eV) Ioannis Varvaris (ECoE - BE1) Diofantos Hadjimitsis (ECoE - BE1) Thomaida Polydorou (ECoE - BE1)
Abstract	This deliverable establishes a structured and secure framework for handling data throughout the Nostradamus project lifecycle. It ensures compliance with Horizon Europe Open Science mandates, FAIR Data Principles (Findable, Accessible, Interoperable, and Reusable), GDPR regulations, and Intellectual Property Rights (IPR) policies. The plan outlines governance structures, data collection methodologies, processing strategies, security frameworks, and long-term storage approaches. The DMP also specifies guidelines for data storage and backup policies, ensuring datasets remain accessible for at least five years post-project.
Keywords	Data Management; FAIR Data; Open Science; GDPR Compliance; Data Governance; Intellectual Property Rights (IPR)

NOSTRADAMUS

Document Revision History

Version	Date	Description of change	List of contributor(s)
0.1	29/01/2025	Creation, First Draft	Thomaida Polydorou (ECoE), Ioannis Varvaris (ECoE), Dimitrios Koumoulidis (ECoE), Diofantos Hadjimitsis (ECoE)
0.2	22/02/2025	Revised Draft	Thomaida Polydorou (ECoE), Tomaz Bokan (ITC), Nils Borchard (DLG eV) Ioannis Varvaris (ECoE - BE1) Diofantos Hadjimitsis (ECoE - BE1)
0.3	26/02/2025	Minor editing, formatting	Thomaida Polydorou (ECoE)

DISCLAIMER

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Health and Digital Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

COPYRIGHT NOTICE

The Nostradamus Consortium is the following:

Participant number	Participant organization name	Short name	Country
1	ERATOSTHENES Centre of Excellence	ECoE	CY
2	DEUTSCHES ZENTRUM FUR LUFT - UND RAUMFAHRT EV	DLR	DE
3	AG FUTURA TECHNOLOGII DOOEL SKOPJE	AGFT	MK
4	CROPT DOO NOVI SAD	CROPT	RS
5	ITC - INOVACIJSKO TEHNOLOSKI GROZD MURSKA SOBOTA	ITC	SI
6	DLG EV	DLG e.V.	DE
7	CESKA ZEMEDELSKA UNIVERZITA V PRAZE	CZU	CZ
8	BIOSENSE INSTITUTE - RESEARCH AND DEVELOPMENT INSTITUTE FOR INFORMATION TECHNOLOGIES IN BIOSYSTEMS	BIOS	RS
9	TERRADUE SRL	TERRADUE	IT
10	KMETIJSKO GOZDARSKA ZBORNICA SLOVENIJE KMETIJSKO GOZDARSKI ZAVOD MURSKA SOBOTA	KGZS	SI
11	F6S NETWORK IRELAND LIMITED	F6S IE	IE
12	YPOURGEIO GEORGIAS, AGROTIKIS ANAPTYXIS KAI PERIVALLONTOS	MARDE	CY
13	ARISTOTELIO PANEPISTIMIO THESSALONIKIS	AUTH	EL
14	UNIVERSITE DE GENEVE	UNIGE	CH
15	UNIVERSITAET ST. GALLEN	HSG	CH
16	EIDGENOESSISCHES DEPARTEMENT FUER WIRTSCHAFT, BILDUNG UND FORSCHUNG	WBF	CH

EXECUTIVE SUMMARY

This document, Deliverable 7.6 "Data Management Plan", or the Nostradamus Data Management Plan (DMP) provides the framework for managing the collection, processing, storage, access, and sharing of multi-source datasets used in the project. In alignment with Horizon Europe Open Science mandates, GDPR regulations, and best practices in data governance, this document ensures that all research data remains Findable, Accessible, Interoperable, and Reusable (FAIR) while maintaining the highest standards of security and ethical compliance.

The Nostradamus project integrates Earth Observation (EO) satellite imagery, IoT sensor networks, AI-driven analytics, and socioeconomic data to support sustainable agricultural decision-making. To ensure efficient data handling, the project employs a hybrid cloud-HPC infrastructure that enables real-time data ingestion, scalable storage, and high-performance computational processing. Additionally, Data Cube architecture and machine-readable APIs enhance interoperability, making datasets easily accessible to researchers, policymakers, and industry stakeholders.

The data security and privacy framework include AES-256 encryption, TLS 1.3 secure data transmission, Role-Based Access Control (RBAC), and Multi-Factor Authentication (MFA) to prevent unauthorized access. Ethical compliance is a core principle, with strict GDPR-aligned policies, anonymization protocols for personal data, and ongoing ethics reviews ensuring transparency in data use.

To guarantee long-term accessibility and data preservation, Nostradamus employs redundant backup strategies, immutable storage snapshots, and archiving through institutional and open-access repositories such as Zenodo, OpenAIRE, and the EU Open Data Portal. Regular FAIR Data Audits and compliance monitoring ensure that datasets remain structured, machine-readable, and openly available.

This DMP is a Deliverable of Nostradamus project Task 7.4: Data management and will be continuously updated throughout the 48-month duration of the project to reflect technological advancements, regulatory changes, and evolving project needs. Future updates will focus on enhancing metadata completeness, expanding AI-driven data cataloging, and improving interoperability across platforms.

By implementing structured data governance, robust security mechanisms, and sustainable Open Science policies, Nostradamus ensures that its research outputs remain accessible, reusable, and impactful, supporting data-driven agricultural innovation and informed policy-making in Europe.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

AI	Artificial Intelligence
API	Application Programming Interface
CA	Consortium Agreement
CAP	Common Agricultural Policy
CoE	Centre of Excellence
DCs	Data Cubes
DMP	Data Management Plan
ECoE	ERATOSTHENES Centre of Excellence
EC	European Commission
EO	Earth Observation
EU	European Union
FAIR	Findable, Accessible, Interoperable, and Reusable
F2F	Farm to Fork Strategy
GA	Grant Agreement
GDPR	General Data Protection Regulation
IoT	Internet of Things
IP	Intellectual Property
IPR	Intellectual Property Rights
PII	Personally Identified Information
PMQP	Project Management and Quality Plan
QA	Quality Assurance
RIA	Research and Innovation Action
RRI	Responsible Research and Innovation
WP	Work Package
WPL	Work Package Leader

1 INTRODUCTION

The Nostradamus Data Management Plan (DMP) document, or Deliverable 7.6 defines how research data will be collected, processed, stored, protected, and shared throughout the 48-month duration of the Nostradamus project, funded through the Horizon Europe Programme. The plan follows the guidelines set by Horizon Europe, the Grant Agreement (GA No. 101134888), and the Consortium Agreement (CA), ensuring full compliance with FAIR Data Principles (Findable, Accessible, Interoperable, and Reusable), GDPR regulations, and Open Science policies. Furthermore, the DMP will be continuously updated throughout the project duration to reflect evolving data governance policies, updates to data collection methodologies, adjustments in security, access, and sharing procedures, as well as improvements in FAIR data management.

By implementing a robust data management strategy, Nostradamus aims to maximize the impact of its research outputs, enhance data transparency, and facilitate secure data exchange among consortium members and external researchers.

1.1 PROJECT OVERVIEW

The Nostradamus Project is dedicated to enhancing food security and reducing reliance on agrochemical inputs across the European Union (EU). In alignment with EU policies such as the Farm to Fork Strategy (F2F), the Common Agricultural Policy (CAP), and the European Strategy for Data, the project employs a multi-actor collaboration approach to develop a scalable and resilient digital platform for near-real-time agricultural data collection and analysis.

By integrating multi-source datasets, including Copernicus Earth Observation (EO) data, Internet of Things (IoT) technologies, and open-access databases, Nostradamus delivers data-driven insights that support sustainable agricultural practices, enhance sector resilience and sustainability and improve competitiveness. A key component of this initiative is the development of Data Cubes (DCs)-centralized digital hubs that provide stakeholders with continuous access to relevant, actionable information for informed decision-making and policy development. Leveraging state-of-the-art digital solutions, Nostradamus develops an inclusive, data-driven agricultural ecosystem that addresses pressing challenges, unlocks new opportunities, and strengthens the sustainability and transparency of farming operations. The project pioneers edge-, cloud-, and hybrid computing technologies to boost agricultural productivity while reducing input costs and minimizing environmental impacts, particularly in areas with limited connectivity infrastructure.

Beyond technological advancements, Nostradamus is committed to strengthening administrative and monitoring capacities to enhance the efficiency, accountability and transparency of agricultural practices. Upholding ethical standards, gender inclusivity, and cultural diversity, the project ensures that its processes and outputs remain accessible to diverse user groups, paving the way for a more sustainable, equitable, and resilient agricultural future.

1.2 OBJECTIVES OF THE NOSTRADAMUS DMP

The primary objective of this Deliverable is to establish a comprehensive and structured approach for managing all data generated, collected, and processed throughout the project lifecycle. The Nostradamus project Data Management Plan (DMP) serves the key purposes, described as follows:

a) Ensure Compliance with Horizon Europe's Open Science & FAIR Data Policies-

Adopting FAIR principles to enhance data reuse and interoperability and ensuring compliance with Horizon Europe's Open Access mandates for publicly funded research.

b) Define Data Governance & Security Standards-

Implementing GDPR-compliant data protection measures for sensitive datasets, as well as guaranteeing data integrity through encryption, backup strategies, and controlled access.

c) Facilitate Collaboration & Data Sharing within the Consortium

Enabling secure, role-based access to project datasets for partners, and standardizing data formats and metadata schemas to improve interoperability.

d) Maximize Research Impact through Open Science & Public Data Access

Publishing high-value datasets on OpenAIRE, Zenodo, and the EU Open Data Portal and ensuring long-term accessibility for post-project data reuse.

e) Provide a Risk Management Framework for Data Handling

Identifying potential data security risks and mitigation strategies and implementing disaster recovery plans for critical datasets.

By fulfilling these objectives, the Nostradamus Project DMP ensures that all project data are managed in a secure, ethical, and transparent manner, promoting open-access science while protecting intellectual property and sensitive information.

2 DATA GOVERNANCE & COMPLIANCE FRAMEWORK

The Nostradamus Data Management Plan (DMP) is built upon a strong governance structure that ensures legal compliance, data integrity, and ethical handling of information in alignment with the Horizon Europe Grant Agreement (GA No. 101134888) and the Consortium Agreement (CA). This section outlines the regulatory framework, Intellectual Property (IP) management, access rights, and ethical considerations that govern data collection, processing, and data sharing throughout the project lifecycle.

The primary objectives of the data governance framework are to ensure full compliance with European and international data protection laws (GDPR, Horizon Europe mandates, and Open Science policies), define clear data ownership and access rights for consortium members, guarantee transparency, accountability, and interoperability in data handling, and protect sensitive and confidential data through robust security measures.

2.1 LEGAL COMPLIANCE

Ensuring compliance with European and international data protection laws, research ethics, and open science mandates is fundamental to the Nostradamus project. This section outlines the legal frameworks governing data collection, processing, security, and accessibility, ensuring that all project activities align with Horizon Europe regulations, GDPR (Regulation (EU) 2016/679), and FAIR Data Principles while maintaining transparency, accountability, and ethical integrity.

2.1.1 COMPLIANCE WITH GENERAL DATA PROTECTION REGULATION (GDPR – REGULATION (EU) 2016/679)

The Nostradamus project strictly adheres to the General Data Protection Regulation (GDPR), ensuring that all data collection, processing, and sharing activities comply with European Union (EU) legal standards. The project follows key GDPR principles, such as: Lawfulness, fairness, and transparency: for example, Data subjects will be informed of how their data is processed, and explicit consent will be obtained where necessary, Purpose limitation, e.g. Data will only be used for the specific research objectives outlined in the project proposal, Data minimization, meaning Only necessary data will be collected, avoiding excessive or redundant information, Storage limitation: for example, Personal data will be retained only for the required period and securely deleted afterward, Integrity and confidentiality: e.g. All data will be protected against unauthorized access, breaches, or leaks.

2.1.2 HORIZON EUROPE OPEN SCIENCE AND FAIR DATA REQUIREMENTS

The Horizon Europe Programme mandates that all publicly funded research projects comply with Open Science principles, ensuring that research data is accessible, reusable, and transparent. Nostradamus implements the FAIR (Findable, Accessible, Interoperable, and Reusable) Data Principles, which ensure Findability: All datasets will be assigned a Digital Object Identifier (DOI) and published with rich metadata, Accessibility: Data will be made available through open-access repositories such as Zenodo, OpenAIRE, and institutional platforms, Interoperability: Standardized metadata formats and vocabularies (e.g., OGC, Dublin Core, AGROVOC) will be used, and Reusability: Clear licensing (Creative Commons CC-BY or CC0) and detailed metadata documentation will be provided.

2.1.3 COMPLIANCE WITH ETHICAL AI AND RESPONSIBLE RESEARCH PRACTICES

The project follows the EU AI Act (European Parliament) ethical guidelines for Artificial Intelligence (AI) and Machine Learning (ML) to ensure:

a) Transparency, e.g. all AI-generated data will be accompanied by explainability statements,

b) Bias mitigation, meaning all AI models will undergo periodic audits to detect potential biases,

c) Accountability, thus a Data Ethics Advisor is responsible for overseeing compliance with AI ethics and privacy regulations.

2.2 INTELLECTUAL PROPERTY & ACCESS RIGHTS

The Nostradamus project generates and utilizes multiple data sources, including Copernicus Earth Observation (EO) data, Internet of Things (IoT) sensor datasets, stakeholder-generated content, and AI-processed insights. Proper management of Intellectual Property Rights (IPR) and access permissions ensures that partners can collaborate effectively while protecting proprietary data.

2.2.1 FOREGREOUND VS BACKGROUND IP RIGHTS

The Foreground IP (Newly Generated Data) of Nostradamus includes any data, software, or knowledge created during the project that will belong to the partner(s) responsible for its generation. Foreground IP may be jointly owned if developed collaboratively. Data access rights for other partners will be granted under the terms of the Consortium Agreement (CA).

Now regarding Background IP (Pre-Existing Knowledge) utilized within the Nostradamus project, any pre-existing datasets, software, or models used in the project remain the property of the original owner. Background IP will be accessible to consortium members on a need-to-use basis.

2.2.2 DATA ACCESS & LICENSING MODELS

The Nostradamus project adheres to clear access control mechanisms to ensure that project data is shared responsibly, securely, and in compliance with Horizon Europe Open Science mandates, GDPR, and intellectual property rights (IPR) agreements. This section defines the different levels of data accessibility, licensing options, and storage locations for project-generated datasets.

The data access policy is designed to balance open-access research and innovation with the need to protect sensitive, proprietary, and confidential information. The approach follows FAIR Data Principles, ensuring that data is Findable, Accessible, Interoperable, and Reusable while safeguarding the rights of data owners and contributors. Data generated in the Nostradamus project falls into three broad categories:

- Public Research Data – Openly available for unrestricted use.
- Restricted Project Data – Limited to consortium members for internal research purposes.
- Confidential & Sensitive Data – Strictly controlled, accessible only to designated project personnel and regulatory bodies.

Data access tiers applied in Nostradamus are described in Table 1. The table details the data classification, access permissions, licensing models, and storage locations for each type of dataset.

The definition of terms used in the table and some examples for clarification are provided in the following sections.

TABLE 1. NOSTRADAMUS PROJECT DATA ACCESS TIERS

Data category	Access level	Licensing	Storage Location
Public Research Data	Open Access	Creative Commons (CC-BY, CC0, CC-BY-NC-SA)	Zenodo, OpenAIRE, EU Open Data Portal
Restricted Project Data	Limited to Nostradamus Consortium	Custom Access Agreements under Consortium Agreement	Nostradamus SharePoint, Project Data Cube
Confidential & Sensitive Data	Restricted (Only for EC, GA, and Ethics Board)	GDPR-Compliant Internal Use Only	Secure Encrypted Cloud Storage, Private On-Premise Servers (ECoE, UNIGE)

2.2.2.1 PUBLIC RESEARCH DATA

Public Research Data (Open Access) are defined as Datasets that have no access restrictions and are made freely available for use by researchers, policymakers, industry stakeholders, and the general public. Licensing models include:

- CC-BY (Attribution) – Users must credit the original creators.
- CC0 (Public Domain Dedication) – No restrictions on use, modification, or redistribution.
- CC-BY-NC-SA (Non-Commercial, Share-Alike) – Data can be used non-commercially and must retain the same license.

The data storage platforms used by the Nostradamus project include the following:

- Zenodo (CERN) – European Open Science Repository
- OpenAIRE – Horizon Europe's Open Science Infrastructure
- EU Open Data Portal – Centralized data hub for publicly funded research

Examples of data to be managed accordingly are Climate models derived from Copernicus Earth Observation (EO) data, AI-driven agricultural risk assessments, and Crop yield forecasts for European farmlands.

2.2.2.2 RESTRICTED PROJECT DATA

Restricted Project Data (Limited to Consortium) are defined as Data that is intended for internal use within the consortium but is not immediately made public due to ongoing research, intellectual property rights, or commercial considerations. Their accessibility is restricted as follows;

- Access is restricted under the Consortium Agreement (CA).
- Individual partners can grant controlled access based on bilateral agreements.

Storage platforms used for storing Restricted Project Data include the Nostradamus project SharePoint, the Internal Collaboration & Document Management folders, and the dedicated Project Data Cube, a Private Repository for Partners. Examples of such Data are Pre-publication AI models and raw training data, Internal reports on stakeholder engagement findings, and Unpublished sensor data collected from pilot agricultural test sites.

2.2.2.3 CONFIDENTIAL & SENSITIVE DATA (RESTRICTED USE)

Confidential & Sensitive Data (Restricted Use) are Data that contain personally identifiable information (PII), proprietary research, or sensitive stakeholder information that cannot be publicly disclosed due to legal, ethical, or security concerns.

Their Licensing involves Strict GDPR Compliance, e.g. Access limited to authorized personnel only. Such data are for Internal Use, thus not available for external research or commercial purposes.

Storage Platform to be used for Confidential & Sensitive Data include Secure Encrypted Cloud Storage (AES-256 encryption at rest and TLS 1.3 encryption in transit), and Private On-Premise Servers (ECoE, UNIGE). Examples of Confidential & Sensitive Data are Personally identifiable information (PII) collected via farmer surveys, Proprietary AI algorithm training datasets used for machine learning research, Confidential regulatory reports related to EU agricultural monitoring policies.

2.2.3 DATA LICENSING COMPLIANCE & MONITORING

To ensure responsible sharing, data security, and regulatory compliance, the Nostradamus project consortium will implement a continuous monitoring and auditing framework, including Automated License Attribution & DOI Assignment, meaning all datasets will be assigned a unique Digital Object Identifier (DOI) and metadata-based licensing information; Regular Data Audits, e.g. a Data Governance & Ethics Advisor will conduct quarterly compliance checks to ensure data is shared according to access policies; and Secure Data Transfer Protocols, for example sensitive data will be transmitted via encrypted channels Transport Layer Security (TLS) 1.3 with Role-Based Access Control (RBAC).

Transport Layer Security (TLS) 1.3 is the latest encryption protocol designed to provide secure communication over the internet. It ensures that data transmitted between servers, applications, or users remains private and secure from unauthorized access. TLS 1.3 is the successor to TLS 1.2 and offers Stronger security by removing outdated cryptographic algorithms that were vulnerable to attacks; Faster performance, since it reduces the number of handshake steps, making encrypted connections more efficient; Perfect forward secrecy, because even if encryption keys are compromised, past communications remain secure.

Transport Layer Security (TLS) 1.3 can therefore protect data transmissions between project partners, databases, and storage platforms, ensure compliance with GDPR by securing sensitive datasets against unauthorized interception. It can further be used for encrypting API calls between Nostradamus Data Cubes and cloud platforms.

For example, when researchers download satellite imagery or IoT sensor data from a cloud-based repository, TLS 1.3 encrypts the data to prevent interception.

On the other hand, Role-Based Access Control (RBAC) is a security model that restricts access to data, systems, and services based on a user's role within an organization. In that way, it minimizes security risks by ensuring users only have access to data they need; enhances compliance with GDPR & data protection laws by restricting access to sensitive information and streamline management of permissions for different levels of users (e.g., Researchers, Administrators, Data Managers). RBAC is useful in the case of Nostradamus since it ensures only authorized project partners can modify or access specific datasets; prevents accidental data leaks by restricting access based on job roles; supports traceability by logging who accessed or modified data within the Data Cube infrastructure.

An example use case for RBAC in Nostradamus is when a Work Package Leader (WPL) may have "write" access to update datasets in the repository, while a public stakeholder may have read-only access to open datasets.

Both TLS 1.3 and RBAC are necessary so that Nostradamus will ensure that the project data remains secure, accessible only to authorized users, and compliant with the EU regulations.

GDPR requires secure data transmission for personal or sensitive information. TLS 1.3 is currently the best standard for encrypted communications. Furthermore, RBAC aligns with Horizon Europe's data protection requirements by ensuring only authorized individuals can access sensitive research datasets.

In addition, TLS 1.3 is the industry standard for encrypting API calls, cloud storage, and database connections in research and enterprise settings, while RBAC is widely used in cloud platforms (AWS, Azure, Kubernetes) to ensure fine-grained access control. By integrating TLS 1.3 and RBAC, Nostradamus ensures that data remains secure, accessible only to authorized users, and compliant with EU regulations.

2.3 DATA GOVERNANCE ROLES & RESPONSIBILITIES

Effective data management requires clear role assignments within the consortium. In Nostradamus, the Project Coordinator, ERATOSTHENES CoE is responsible for ensuring data governance compliance and supervise updates to the DMP. This document is the first DMP for the project and will be updated throughout the project duration.

The project Data Protection Officer oversees GDPR compliance and personal data security. The project coordinator (ERATOSTHENES CoE) and WP 7 (Project Management) Leader implements data storage, sharing, and FAIR compliance strategies. The project coordinator is overall responsible for ensuring proper data management and governance, while also leading all administrative, financial, and legal compliance aspects of data governance as per the Nostradamus project GA.

Furthermore, Work Package Leaders are responsible for ensuring data quality, metadata documentation, and accessibility. A data protection team comprised of Data Protection Officers (DPOs) from consortium members is designated to oversee GDPR compliance and data security, ethical compliance, AI data governance and privacy protection. Regular internal reviews will be conducted to ensure all partners adhere to data governance protocols. All project data must be anonymized where necessary before being shared in open-access repositories to ensure data protection and prevent privacy breaches, as per Nostradamus GA.

The Nostradamus consortium is responsible for depositing research data in certified repositories (e.g., Zenodo, OpenAIRE), ensuring FAIR data principles are followed.

In addition, the project governance model ensures that all Work Package Leaders (WPLs) are accountable for data quality, curation, and compliance with FAIR principles, as described in the GA.

2.4 DATA MANAGEMENT MONITORING & AUDITS

The Nostradamus project implements a continuous monitoring and auditing framework to ensure data governance, quality, security, and compliance with the Grant Agreement (GA), Consortium Agreement (CA), GDPR, and Horizon Europe Open Science policies.

The goal of this monitoring system is to:

- Ensure adherence to data governance and FAIR principles.
- Detect non-compliance or security vulnerabilities in data handling.
- Maintain transparency and accountability in data access and use.
- Provide timely updates on data integrity and security measures.

The Data Governance & Compliance Monitoring Framework consists of Quarterly Data Audits, which are basically periodic reviews to ensure GDPR compliance, metadata quality, and data integrity; Annual FAIR Compliance Assessments, e.g. Evaluations to verify that research data remains open, reusable, and properly documented; Data Security Reviews, such as Cybersecurity assessments to identify potential vulnerabilities and enhance data protection mechanisms; Performance Monitoring of Data Repositories for ensuring availability, accessibility, and storage efficiency of the Nostradamus Data Cubes and project databases.

2.4.1 PERIODIC DATA AUDITS

Quarterly data audits will be conducted to evaluate data handling, access permissions, security policies, and compliance with ethical guidelines. The Data Protection Officer (DPO) and Data Manager will oversee this process. The periodic audit scope is summarized as follows:

GDPR Compliance Check – Ensuring that personally identifiable information (PII) is anonymized and that data processing aligns with legal obligations.

Data Integrity & Quality Check – Assessing accuracy, completeness, and metadata documentation for all datasets.

FAIR Data Compliance – Verifying that datasets are findable (assigned DOIs), accessible (stored in Zenodo/OpenAIRE), interoperable (standardized formats), and reusable (open licenses assigned).

Access Control & Security Check – Reviewing user access permissions under Role-Based Access Control (RBAC) and monitoring data access logs for unauthorized activities.

The Audit Process & Reporting procedure commences with the Data Protection Officer (DPO) & Data Manager initiating an internal audit. The second step is when partners submit reports on their data management activities. Later findings are reviewed, and non-compliance cases are flagged, leading to a corrective action plan implementation if needed.

Findings from the Audit Process are documented in a quarterly compliance report submitted to the Project Coordinator (ERATOSTHENES CoE) and Project Management Board (PMB).

2.4.2 ANNUAL FAIR COMPLIANCE ASSESSMENTS

An annual review will evaluate how well Nostradamus datasets comply with FAIR principles and Open Science mandates.

The assessment criteria include checks on Findability – Ensuring that all datasets are indexed in public repositories (Zenodo, OpenAIRE, EU Open Data Portal) with rich metadata descriptions; Accessibility – Verifying that datasets are freely available under open-access licenses and accessible via standardized APIs; Interoperability – Checking that data formats follow OGC, Dublin Core, and AGROVOC metadata standards; Reusability – Confirming that datasets are well-documented and follow Creative Commons (CC-BY, CC0) licensing standards.

The assessment process initiates with compiling the findings. Subsequently, a scorecard system will track the progress of FAIR implementation across datasets. Lastly, the results will be presented to the General Assembly (GA).

2.4.3 DATA SECURITY REVIEWS

Cybersecurity reviews will be performed to protect sensitive project data. These will focus on the following:

a) Encryption Standards Compliance – Ensuring that all data at rest (stored) and in transit (transferred) is encrypted using AES-256 and TLS 1.3 security protocols.
b) Penetration Testing & Cyber Risk Assessment – Conducting simulated attacks to identify vulnerabilities in cloud storage and database systems.
c) Incident Response Planning – Establishing a data breach response plan to address potential security incidents.
d) Access Control Monitoring – Reviewing audit logs of data access and modification requests to detect suspicious activity.

2.4.4 PERFORMANCE MONITORING OF DATA REPOSITORIES

The availability, performance, and efficiency of data storage repositories (e.g., Nostradamus Data Cubes, Zenodo, and institutional repositories) will be assessed to ensure uptime and accessibility for all datasets; evaluate storage efficiency and scalability for handling increasing data volumes; monitor data retrieval speed and API performance for third-party integrations; and identify and resolve system failures or downtime risks.

A bi-annual Data Infrastructure Performance Report will be generated, summarizing repository uptime, storage capacity utilization, and potential infrastructure upgrades.

2.4.5 CONTINUOUS IMPROVEMENT & CORRECTIVE ACTIONS

If any deficiencies are identified in audits, security reviews, or FAIR compliance checks, the Project Coordinator (ERATOSTHENES CoE) and Data Governance Team will initiate corrective actions, such as: Revising access control policies to prevent unauthorized data exposure; Improving metadata quality for better data discoverability; Upgrading encryption standards to enhance security; Expanding storage resources to accommodate increased dataset sizes. A Corrective Action Plan (CAP) will be documented and followed up in the next monitoring cycle.

Quarterly Data Audits ensure GDPR compliance, data integrity, and security. Furthermore, annual FAIR Compliance Assessments track adherence to Open Science policies. Bi-annual Security Reviews protect against cyber threats and unauthorized access while further Performance Monitoring ensures the efficiency of data repositories. Overall, corrective Action Plans guarantee continuous improvement in data governance.

The Nostradamus Data Management Monitoring System provides a structured, proactive approach to governance, compliance, and security, ensuring that all data remains FAIR, secure, and ethically managed throughout the project lifecycle.

3 FAIR DATA MANAGEMENT & SHARING

The Nostradamus project is committed to implementing FAIR Data Principles, ensuring that all research data is Findable, Accessible, Interoperable, and Reusable (FAIR). These principles align with the Horizon Europe Open Science policy, promoting transparent, reproducible, and sustainable data management practices.

By ensuring data sharing and interoperability, Nostradamus facilitates Maximized scientific impact and knowledge dissemination; Seamless collaboration across stakeholders and research disciplines; and Long-term data accessibility for innovation and policymaking. The FAIR Data Strategy is implemented through four key dimensions:

- Findability – Ensuring datasets are discoverable via unique identifiers and rich metadata.
- Accessibility – Making datasets openly available with standardized access procedures.
- Interoperability – Using open formats and vocabularies for seamless data integration.
- Reusability – Assigning clear usage licenses and detailed metadata to enable future reuse.

This section details how Nostradamus will achieve full FAIR compliance, ensuring that research data remains valuable and accessible beyond the project's lifetime.

3.1 FINDABILITY: ENSURING DISCOVERABLE DATA

The first principle of FAIR ensures that all datasets are easily searchable and discoverable. This is achieved through persistent identifiers, metadata indexing, and standardized dataset registration. Findability is a fundamental principle of FAIR data management, ensuring that datasets are easily searchable and discoverable by researchers, policymakers, industry stakeholders, and the public. In the Nostradamus project, findability is achieved using persistent identifiers (PIDs), metadata indexing, dataset registration, and repository integration.

By implementing robust data cataloging and metadata management strategies, Nostradamus ensures that its research output can be easily located, cited, and reused while complying with Horizon Europe Open Science mandates.

3.1.1 UNIQUE IDENTIFIERS FOR ALL DATASETS

To ensure that Nostradamus datasets remain permanently accessible and citable, all research outputs will be assigned a persistent identifier (PID) in the form of a Digital Object Identifier (DOI). The implementation strategy incorporates that DOIs will be assigned through repositories such as

Zenodo and OpenAIRE, ensuring long-term accessibility. Furthermore, each dataset will be given a unique, machine-readable identifier, allowing seamless referencing in research publications and reports. In addition, Version control will be implemented to track dataset updates, ensuring reproducibility. As an example, a dataset containing AI-generated agricultural risk assessments will be assigned a DOI through Zenodo, ensuring that researchers can accurately cite and reference it in future studies.

3.1.2 METADATA INDEXING AND STRUCTURED DESCRIPTIONS

Metadata plays a critical role in making datasets discoverable, understandable, and reusable. Nostradamus will adopt international metadata standards to ensure consistency and interoperability. Metadata Standards applied in this project are outlined as follows:

a) Dublin Core (DCAT-AP) – For general dataset descriptions (title, author, keywords, abstract).

b) INSPIRE Metadata Directive – For geospatial datasets (coordinate systems, geographic coverage).

c) OGC Standards (ISO 19115) – For Earth Observation (EO) and remote sensing datasets.

d) FAIR Data Point (FDP) Compliance – For structured metadata publication.

Each dataset will be accompanied by rich metadata containing Title & Description – A human-readable summary explaining the dataset's purpose; Dataset Creator & Institution – Identifying the contributors and affiliated organizations; Keywords & Thematic Tags – Assigning standardized terms (e.g., "climate change," "soil health"); Spatial & Temporal Coverage – Indicating geographic extent and timeframes covered; Licensing Information – Clearly specifying usage rights (e.g., CC-BY, CC0).

For example, a Copernicus EO-derived dataset on drought risk will be indexed with ISO 19115 metadata, allowing policymakers to search by geographic region and date range.

3.1.3 DATASET REGISTRATION IN OPEN REPOSITORIES

To maximize discoverability, Nostradamus datasets will be indexed in multiple open-access repositories. These repositories ensure long-term accessibility and compliance with Horizon Europe mandates. Repositories for Data Registration include Zenodo, a general-purpose research data repository maintained by CERN; OpenAIRE, the Horizon Europe Open Science platform for EU-funded projects; EU Open Data Portal, the European Commission's public research repository, and Institutional Data Repositories (ERATOSTHENES CoE, UNIGE), long-term storage in partner institutions.

Repository Integration & API support will be available since each dataset will be registered with a DOI and indexed in repository search engines. Furthermore, machine-readable metadata APIs will allow external platforms to automatically retrieve dataset descriptions.

For example, a dataset containing IoT-based soil moisture readings will be deposited in Zenodo with an assigned DOI, allowing it to be indexed in OpenAIRE and referenced in scientific publications.

3.1.4 SEARCHABILITY ENHANCEMENTS & SEMANTIC TAGGING

To improve dataset searchability, Nostradamus will implement machine-readable tagging and semantic annotations. Semantic Data Enrichment will be achieved via the use of:

- Controlled vocabulary (AGROVOC and GEMET) will be used to standardize thematic keywords.
- Ontology-based classification will enable automatic dataset grouping (e.g., linking soil quality data to climate risk models).
- Natural Language Processing (NLP) algorithms will be applied to suggest relevant datasets based on search queries.

Furthermore, Search Engine Optimization (SEO) for Open Data Portals will be applied since Metadata will be structured in compliance with Google Dataset Search standards, improving ranking in public data repositories; and Machine-readable JSON-LD descriptions will be generated for each dataset, enhancing integration with global Open Science platforms.

For example, a farmer looking for historical crop yield data can type "corn yield trends in Spain" into an OpenAIRE search engine, which will return indexed Nostradamus datasets with matching metadata tags.

AGROVOC was selected because it is the standard for agricultural metadata classification in EU-funded projects. Ontology-based tagging was integrated to enhance dataset searchability, AI integration, and geospatial data interoperability. Both were derived from Horizon Europe requirements, best practices, and metadata standards for research repositories.

These enhancements ensure Nostradamus datasets are highly structured, searchable, and fully aligned with EU Open Science mandates.

3.1.5 CONTINUOUS FINDABILITY MONITORING

To ensure that datasets remain discoverable over time, Nostradamus will implement a Findability Monitoring System that tracks Dataset access statistics (views, downloads, citations) to assess impact; Broken link detection in case repository storage locations change; Metadata completeness audits to ensure datasets remain fully documented.

In addition, an Annual Findability Report will be generated, summarizing dataset indexing performance, metadata quality improvements, and repository search rankings.

In summary, each dataset will be assigned a persistent Digital Object Identifier (DOI) for long-term discoverability; Metadata will follow international standards (Dublin Core, INSPIRE, ISO 19115) for structured descriptions; Datasets will be deposited in Zenodo, OpenAIRE, and EU Open Data Portal for maximum accessibility; Semantic enrichment (AGROVOC, ontology-based tagging) will enhance dataset searchability; and Automated monitoring will track findability metrics and metadata quality over time.

The Findability Strategy in Nostradamus ensures that all research output(s) remain permanently accessible, easily searchable, and optimally indexed in open-access repositories, promoting transparency, usability, and long-term impact.

3.2 ACCESSIBILITY: ENABLING OPEN AND CONTROLLED DATA ACCESS

The Nostradamus project is committed to ensuring that its research data is easily accessible while adhering to Horizon Europe Open Science policies, GDPR compliance, and intellectual property (IP) protection. Accessibility is critical for fostering collaboration among researchers, policymakers, industry stakeholders, and the broader scientific community.

To achieve this, the project follows a two-tiered accessibility model:

Open Access Data – Research datasets that are freely available to the public under open-access policies.

Controlled Access Data – Datasets that contain sensitive, proprietary, or confidential information and require authentication-based access controls.

The project ensures that all public datasets comply with FAIR principles (Findable, Accessible, Interoperable, and Reusable), while sensitive datasets remain securely stored and ethically managed.

3.2.1 OPEN-ACCESS DATA REPOSITORIES

To maximize the impact of research outputs, Nostradamus will deposit open-access datasets in trusted public repositories. These include a) Primary Open-Access Platforms such as Zenodo, a European Commission-supported open science repository (DOI-enabled); OpenAIRE, a Horizon Europe Open Science platform linking research datasets; EU Open Data Portal, the official European Commission research repository; and Institutional Repositories, e.g. University-based platforms (e.g., UNIGE, ERATOSTHENES CoE), and b) Open-Access Licensing Models, e.g. all publicly available datasets will be released under Creative Commons licenses, such as CC0 (Public Domain Dedication) – No restrictions; freely reusable without attribution; CC-BY (Attribution Required) – Users must credit the original source; and CC-BY-NC (Non-Commercial Use Only) – Data can be used for research but not for commercial purposes.

An example involves AI-generated crop health assessments from Copernicus satellite data. Such data will be published in Zenodo under a CC-BY license, enabling researchers and policymakers to access, cite, and reuse the data.

3.2.2 ACCESS CONTROL FOR SENSITIVE DATA

Certain datasets will contain sensitive, proprietary, or confidential information and cannot be openly shared. These will be categorized as restricted-access datasets and will require authentication and access control mechanisms. Types of restricted-access data include:

Personally Identifiable Information (PII) – Data collected from farmers, industry stakeholders, and socio-economic surveys (GDPR-regulated); Proprietary Research Data – AI model training datasets and commercial research outputs developed in collaboration with industry partners; Unpublished Experimental Results – Datasets under embargo until research papers are finalized.

Controlled Access Mechanisms utilized are outlined in the following paragraph:

a) Role-Based Access Control (RBAC) – Only authorized personnel can access restricted datasets.

b) Multi-Factor Authentication (MFA) – Secure login protocols for accessing confidential data.

c) Data Use Agreements (DUAs) – Signed contracts defining permitted use of restricted datasets.

d) Data Encryption (AES-256) – Ensuring data security in transit and storage.

For example, a farmer survey results containing PII will be stored in a secure SharePoint system with RBAC permissions. Only approved researchers with an ethics board-approved request will be granted access.

3.2.3 API-BASED DATA ACCESS FOR EXTERNAL INTEGRATION

To enable automated data retrieval and external system integration, Nostradamus will offer machine-readable API endpoints, such as:

- RESTful APIs, for allowing real-time access to select datasets.
- OAuth 2.0 Authentication, for ensuring secure API access.
- Rate Limiting & Logging, for preventing unauthorized bulk data downloads.

For example, a policymaker analyzing climate trends can use an API to extract temperature anomaly data from the Nostradamus Data Cube in real time.

3.2.4 DATA ACCESSIBILITY METRICS & MONITORING

To ensure that datasets remain accessible and useful over time, Nostradamus will track download and citation statistics, evaluating dataset impact; access request logs, ensuring compliance with data-use agreements; metadata completeness assessments, verifying that datasets remain properly documented.

An Annual Data Accessibility Report will summarize access patterns, usability issues, and necessary improvements.

As Nostradamus follows a dual access model employing Open-Access (Zenodo, OpenAIRE) and Restricted-Access (GDPR-compliant storage), public datasets are published under Creative Commons licenses (CC0, CC-BY). Restricted datasets use RBAC, MFA, and encryption for security, and APIs enable machine-readable access for real-time data retrieval. Data accessibility is continuously monitored to ensure compliance and usability. By ensuring broad accessibility while maintaining security for sensitive data, the project promotes open science, collaboration, and long-term research impact.

3.3 INTEROPERABILITY: ENSURING DATA COMPATIBILITY ACROSS SYSTEMS

Interoperability is essential for ensuring that Nostradamus datasets can be seamlessly integrated, exchanged, and reused across multiple research disciplines, platforms, and stakeholders. The project follows FAIR Data Principles, emphasizing standardized data formats, controlled vocabularies, and machine-readable metadata to enhance data compatibility and usability.

By implementing open standards, geospatial ontologies, and API-based integrations, Nostradamus ensures that researchers can easily integrate datasets into their workflows, policymakers can access structured data for decision-making, and AI models and decision-support systems can efficiently process multi-source data.

The three pillars of interoperability in Nostradamus are outlined as follows:

- Standardized data formats: Ensuring compatibility with geospatial, AI, and IoT systems.
- Controlled vocabularies & ontologies: Enabling structured, machine-readable metadata.
- API-based system integration: Facilitating seamless data exchange across platforms.

3.3.1 STANDARDIZED DATA FORMATS

To maximize interoperability, Nostradamus adopts widely recognized open data formats that are compatible with GIS platforms, machine learning models, and cloud-based infrastructures.

Primary Data Formats Used include Tabular & Structured Data: CSV, JSON, XML (for statistical and IoT data); Geospatial & Remote Sensing Data: GeoTIFF, NetCDF, SHP (for Earth Observation datasets); AI Model Outputs: HDF5, ONNX, TensorFlow SavedModel (for deep learning models); and Sensor Data Streams: MQTT, OGC SensorThings API (for real-time IoT data collection). As an example, crop health monitoring datasets from Copernicus Sentinel-2 imagery will be stored in GeoTIFF format, ensuring compatibility with QGIS, ArcGIS, and Open Data Cube platforms.

3.3.2 ADOPTION OF CONTROLLED VOCABULARIES & ONTOLOGIES

To ensure semantic interoperability, Nostradamus uses controlled vocabularies and domain-specific ontologies for standardized metadata tagging.

Metadata Standards are widely applied, such as the INSPIRE Directive (EU Geospatial Metadata Standard), which ensures geospatial datasets are interoperable across EU data infrastructures; ISO 19115 (Geospatial Metadata Standard), which standardizes descriptions of remote sensing and GIS datasets; and the Dublin Core (DCAT-AP), which provides structured metadata for open-access research data.

As far as Controlled Vocabularies & Semantic Tagging, AGROVOC (FAO Agricultural Ontology), OGC Observations & Measurements (O&M), and GEMET (General Multilingual Environmental Thesaurus) are utilized.

AGROVOC (FAO Agricultural Ontology) provides a Standardized agricultural dataset classification. OGC Observations & Measurements (O&M) ensures IoT sensor data consistency and GEMET (General Multilingual Environmental Thesaurus), standardizes environmental dataset descriptions.

For example, AI-driven pest detection datasets will use AGROVOC terms to tag affected crops, making them searchable and interoperable across agricultural research networks.

3.3.3 CROSS-PLATFORM INTEROPERABILITY AND LINKED DATA

Nostradamus datasets will be cross-compatible with external platforms, ensuring multi-disciplinary reuse. To enable Geospatial Data Compatibility, datasets will be formatted to work with Copernicus DIAS platforms, Google Earth Engine, and ESRI ArcGIS. To allow for Machine Learning Dataset Integration, AI training datasets will follow TFRecord, HDF5, and ONNX standards for use in deep learning frameworks. For IoT & Sensor Data Sharing, IoT-derived datasets will use OGC Sensor Things API, allowing integration with agricultural monitoring systems. An example is when a policymaker analyzing EU-wide crop yield trends can integrate Nostradamus datasets directly into the Copernicus Climate Data Store (CDS).

3.3.4 CONTINUOUS INTEROPERABILITY MONITORING & COMPLIANCE

To ensure ongoing interoperability compliance, Nostradamus will: Conduct quarterly metadata audits to ensure datasets remain machine-readable; Track API usage metrics to optimize data retrieval performance; Update metadata schemas based on evolving EU Open Data standards; Adopts standardized data formats (CSV, GeoTIFF, HDF5) for seamless integration; Uses controlled vocabularies (AGROVOC, GEMET) to improve semantic interoperability; Provides RESTful APIs and GeoServer endpoints for external data exchange; Ensures cross-platform compatibility with GIS, AI, and IoT infrastructures; Implements ongoing monitoring to maintain interoperability compliance.

By prioritizing open standards, linked data, and structured metadata, Nostradamus ensures that its research outputs are universally accessible, easily reusable, and fully interoperable across domains and disciplines.

3.4 REUSABILITY: ENABLING DATA SHARING AND LONG-TERM USABILITY

Reusability is a fundamental aspect of FAIR Data Principles, ensuring that research data can be efficiently reused by researchers, policymakers, and industry stakeholders. In the Nostradamus project, all datasets will be structured, documented, and licensed for future use, maximizing the impact and longevity of the project's outputs.

- Reusability in Nostradamus is achieved through: Clear licensing & intellectual property (IP) policies: Ensuring that datasets can be shared and cited appropriately.
- Detailed documentation & provenance tracking: Providing transparency on data collection, processing, and methodology.
- Long-term data preservation & sustainability: Ensuring accessibility beyond the project's duration.

This section outlines how the Nostradamus project will ensure data reusability, making datasets valuable resources for future research and innovation.

3.4.1 CLEAR LICENSING & USAGE RIGHTS

To guarantee legal clarity and ethical data sharing, the Nostradamus project datasets will be released under standardized licensing models. Specifically, the Licensing Strategy for Open-Access Datasets incorporates Creative Commons (CC) Licensing, meaning all public datasets will be shared under CC0 (Public Domain Dedication), which entails no restrictions on reuse, modification, or redistribution; or CC-BY (Attribution Required), where users must credit the dataset's original creator; and CC-BY-NC (Non-Commercial Use Only), which prevents commercial exploitation while ensuring research use. Regarding Machine-Readable Licensing, all metadata will include structured license information, ensuring automatic recognition by OpenAIRE, Zenodo, and EU Open Data repositories.

In terms of Attribution and Citation Requirements, datasets will have a preferred citation format to credit authorship and incorporate persistent Digital Object Identifiers (DOIs) to enable accurate referencing.

3.4.2 DETAILED DOCUMENTATION & PROVENANCE TRACKING

For datasets to be reusable, users must fully understand their content, origin, and limitations. Nostradamus ensures this through comprehensive metadata documentation and provenance tracking. Metadata Documentation Standards are summarized as follows:

a) ISO 19115 & INSPIRE Directives – Ensuring geospatial data descriptions follow EU standards.

b) Dublin Core (DCAT-AP) Schema – Providing structured metadata for OpenAIRE & Zenodo indexing.

c) OGC Observations & Measurements (O&M) – Standardizing IoT sensor metadata.

To ensure provenance Tracking & Data Lineage, all datasets will include detailed documentation on:

- Data collection methods (e.g., Copernicus EO, IoT sensor feeds).
- Processing workflows (e.g., AI model training, statistical cleaning).
- Data validation techniques (e.g., accuracy assessments, uncertainty quantification). Automated version tracking will ensure that dataset updates and modifications remain transparent.

3.4.3 LONG-TERM DATA PRESERVATION & SUSTAINABILITY

To ensure that research data remains accessible beyond the project's lifetime, Nostradamus follows a long-term preservation strategy. The Data Retention Policy implies that all project datasets will be preserved for a minimum of 5 years post-project (in accordance with Horizon Europe mandates), and archival storage will be maintained in Zenodo & OpenAIRE (for public datasets), Institutional Repositories (ERATOSTHENES CoE, UNIGE) (for internal long-term storage).

Permanent DOIs will be assigned for Dataset Tracking, and DOIs will be generated for all finalized datasets, ensuring persistent referencing. Furthermore, metadata records will be maintained in multiple repositories, preventing data loss.

The Nostradamus Digital Preservation Strategy incorporates periodic format migration, to ensure that datasets remain compatible with modern standards; Backup redundancy will be implemented to prevent accidental data loss; AI model weights and training datasets will be stored in ONNX, HDF5, and TensorFlow formats for future reuse in machine learning applications.

3.4.4 MONITORING DATASET REUSE & IMPACT

To assess how Nostradamus datasets are used and cited, a reuse monitoring system will track Download and citation statistics via Zenodo & OpenAIRE, usage patterns through API request logs, and user feedback surveys to improve dataset usability.

An Annual Data Reusability Report will summarize findings and suggest improvements. Furthermore, Nostradamus datasets will be published under open-access licenses (CC0, CC-BY).

In summary, comprehensive metadata and provenance tracking will ensure transparency and usability, long-term preservation will be secured through Zenodo, OpenAIRE, and institutional repositories, and monitoring dataset reuse will provide insights into impact and usability.

By ensuring structured documentation, legal clarity, and long-term preservation, Nostradamus datasets will remain highly reusable, supporting future research and policy development across multiple disciplines.

3.5 CONTINUOUS FAIR DATA COMPLIANCE MONITORING

Continuous monitoring and quality assurance mechanisms are required to ensure that datasets remain FAIR-compliant. The Nostradamus project implements a structured compliance framework to ensure completeness and adherence of metadata to FAIR principles, assess data usability and accessibility in open repositories, monitor repository performance and dataset availability, and regularly review compliance with Horizon Europe Open Science policies.

This continuous monitoring system ensures that datasets remain usable, traceable, and interoperable for future research and policy development.

3.5.1 FAIR DATA AUDITS & COMPLIANCE CHECKS

To maintain data quality and accessibility, quarterly FAIR Data Audits will be conducted by the Data Manager and Project Coordinator (ERATOSTHENES CoE), as per the Nostradamus project Grant Agreement. Audit Criteria include:

- Findability Check – Verifying that all datasets have DOIs and metadata descriptors.
- Accessibility Verification – Ensuring that datasets remain available in Zenodo, OpenAIRE, and EU Open Data Portal.
- Interoperability Review – Assessing compliance with OGC, ISO 19115, and AGROVOC ontologies.
- Reusability Assessment – Checking whether datasets have clear licensing (CC-BY, CC0) and provenance tracking.

In the case where a dataset fails a FAIR compliance check due to missing metadata fields, it will be flagged for correction, ensuring that users have complete dataset descriptions.

3.5.2 METADATA QUALITY ASSURANCE & VERSION CONTROL

To prevent metadata decay, Nostradamus will conduct automated metadata validation before dataset publication, as per the project GA. This process involves the following steps:

a) Automated Metadata Validation – AI-based tools will detect missing or inconsistent metadata fields.

b) Version Tracking & DOI Management – All dataset updates will be archived under version-controlled repositories.

c) Cross-Repository Synchronization – Metadata records will be synced across Zenodo, OpenAIRE, and institutional repositories.

Therefore, in this case, if a researcher updates soil health data with new geospatial layers, the system will automatically generate a new version and retain prior versions for reproducibility.

3.5.3 DATA ACCESSIBILITY PERFORMANCE MONITORING

To ensure datasets remain accessible and discoverable, Nostradamus will track Repository Uptime & Performance Metrics – Ensuring Zenodo and OpenAIRE maintain 99.9% availability; Download & API Access Logs – Monitoring how often datasets are retrieved and accessed by third parties; User Engagement Analytics – Collecting feedback from researchers and stakeholders on dataset usability. If a dataset receives high download rates but low citations, additional metadata enrichment efforts will be initiated to improve dataset discoverability.

3.5.4 OPEN SCIENCE COMPLIANCE & REPORTING

Annual FAIR Compliance Reports will be submitted to the Project Management Board (PMB), detailing Dataset publication rates in open repositories, Compliance with Open Science mandates, and FAIR adherence improvements over time. If a dataset is not compliant with OpenAIRE metadata standards, a corrective action plan will be initiated to update missing metadata fields.

Quarterly FAIR Data Audits ensure datasets remain usable and compliant. Furthermore, automated metadata validation prevents errors and enhances dataset quality. Repository performance monitoring ensures uninterrupted data access, and annual compliance reports track Open Science progress and dataset impact. By ensuring continuous FAIR compliance monitoring, Nostradamus guarantees that research datasets remain high-quality, well-documented, and easily reusable for scientific and policy-making purposes.

4 DATA SUMMARY

The Nostradamus project generates and processes large-scale, multi-source datasets from Earth Observation (EO) platforms, Internet of Things (IoT) sensors, AI-driven analytics, and open-access databases. This section provides a comprehensive overview of the data types, sources, formats, expected volumes, and processing methods to ensure efficient storage, management, and accessibility.

By categorizing datasets according to their source, format, and usage, this section ensures that all partners and stakeholders have a clear understanding of how project data is generated, structured, and stored.

4.1 TYPES OF DATA COLLECTED

The Nostradamus project integrates four primary types of data; namely, Earth Observation (EO) Data, IoT Sensor Data, AI Model Outputs, and Socioeconomic & Stakeholder Data. A description of each type, their sources, formats and expected volumes are summarized in Table 2. Data collection is based on Consortium, projects, interested parties and stakeholders needs, and aligned with EU strategies and initiatives as described in Figure 1.

TABLE 2. NOSTRADAMUS DATA TYPE DESCRIPTION SUMMARY

Data Category	Description	Source	Format(s)	Expected Volume
Earth Observation (EO) Data	Remote sensing images for agricultural monitoring	Copernicus Sentinel-2, Landsat	GeoTIFF, NetCDF, HDF5	~20 TB over 4 years
IoT Sensor Data	Real-time environmental data (temperature, soil moisture, weather)	Smart farm sensors, IoT networks	CSV, JSON, MQTT, XML	~5 TB per year
AI Model Outputs	Machine learning-generated insights on crop health, climate risks	AI pipelines trained on historical datasets	HDF5, ONNX, TFRecord	~3 TB over project duration
Socioeconomic & Stakeholder Data	Surveys, interviews, market trends	Surveys, government records, farm associations	XLSX, CSV, PDF	~500 GB over project duration

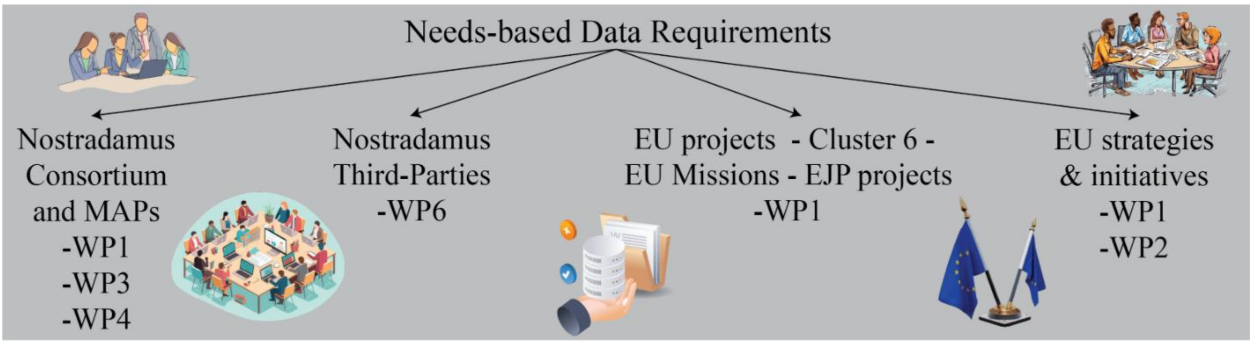

FIGURE 1. NEEDS BASED DATA CONCEPT

4.2 DATA SOURCES & ACQUISITION METHODS

The Nostradamus project integrates data from various public, proprietary, and real-time sources to enhance its predictive capabilities. Data acquisition from various sources, based on needs and aligned with EU strategies is outlined in Figure 2.

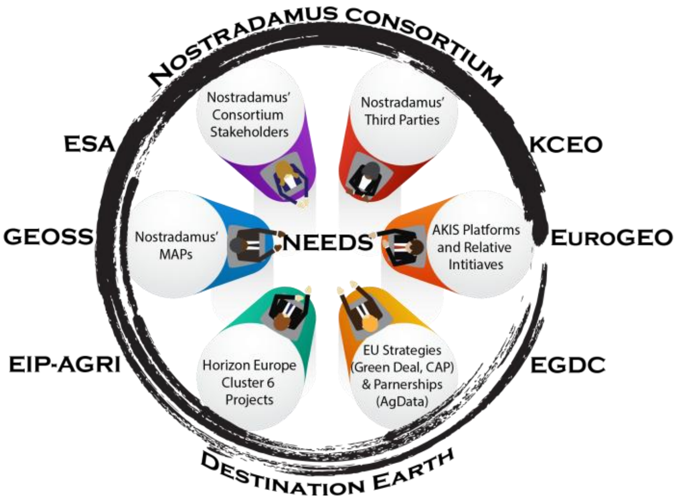

FIGURE 2. NOSTRADAMUS NEEDS BASED DATA ACQUISITION

4.2.1 EARTH OBSERVATION (EO) DATA SOURCES & IOT SENSOR NETWORKS

Earth Observation (EO data sources include Copernicus Sentinel-2 & Sentinel-1 for multi-spectral satellite imagery for crop monitoring, Landsat 8 & 9 for High-resolution EO imagery for historical analysis, European Space Agency (ESA) Climate Data Store (CDS): Climate datasets relevant to agricultural modeling, as well as drone & UAV Sensors capturing hyperspectral imagery for precision agriculture.. Data from IoT sensor networks incorporate data from Soil Moisture Probes capturing real-time soil conditions, weather Stations collecting temperature, precipitation, and humidity.

4.2.2 AI & MACHINE LEARNING TRAINING DATA

In Nostradamus, AI & Machine learning training data such as Historical crop yield datasets from EU agricultural research projects, Climate risk assessment models trained on EO data, as well as Synthetic datasets generated for model validation will also be used.

4.2.3 SOCIOECONOMIC & MARKET DATA

Furthermore, Socioeconomic and Market data such as the EUROSTAT & FAOSTAT Official European agricultural and market statistics, Stakeholder Surveys & Farm Management Logs: Direct input from farmers and policymakers, as well as policy data from the Common Agricultural Policy (CAP) Reports will be used for model calibration.

Regarding all data cases, Nostradamus ensures alignment with the European Green Deal, other EU strategies, and international initiatives by organizing stakeholder consultations, workshops, and ongoing collaborations with various projects and platforms. Examples of strategies/initiatives considered by Nostradamus, Actions and Data for supporting them are presented in Table 3.

TABLE 3. DATA REQUIREMENMTS FOR DATACUBES

Strategy/ Initiative	Objectives	Actions	Data for supporting actions
		EU Green Deal	
Farm-to-fork strategy	To facilitate a rapid transition to a sustainable food system that: Has a neutral or positive impact on the environment Aids in the mitigation of climate change and adapts to its repercussions Contributes to reversing the trend of biodiversity loss Ensures food security, nutrition, and public health, guaranteeing that all individuals have access to adequate, safe,	Proposal for a legislative framework for sustainable food systems. Develop a contingency plan for ensuring food supply and food security. Ensure sustainable food production. Stimulate sustainable food processing, wholesale, retail, hospitality and food services practices. Promote sustainable food consumption, facilitating the shift towards healthy, sustainable diets. Reduce food loss and waste.	Agricultural Management practices-techniques and their percentage per country. Crop production Crop arable land areas Organic Agriculture Share of Smart Agriculture entities per country Chemicals their Distribution and their economic accounts Irrigated / Non – irrigated land areas Water resources utilized and total water consumption in Agriculture. Imported / Exported Crop products

	nutritious, and sustainable food Maintains food affordability while generating equitable economic returns, enhancing the competitiveness of the EU supply sector, and promoting fair trade initiatives.		Imports / Exports of raw materials Food loss and Waste per capita / Country. Indicator of traceability of Agricultural products
Zero Pollution Action Plan	The objective is to reduce air, water, and soil pollution to levels that are no longer harmful to public health and natural ecosystems. The targets consist of: Improving water quality will reduce waste and plastic litter in marine environments by 50% and decrease the release of microplastics into the environment by 30%. Soil quality will be improved through a 50% reduction in nutrient losses and the use of chemical pesticides. Reduce the number of EU ecosystems at risk from air pollution's impact	Revise the Environmental Quality Standards Directive and the Groundwater Directive. Review and, if necessary, revise the Marine Strategy Framework Directive. Review and, if necessary, revise the Bathing Water Directive. Revise the Urban Wastewater Treatment Directive in synergy with reviewing the Industrial Emissions Directive and evaluating the Sewage Sludge Directive. Identify and remedy contaminated sites. Evaluation of the Environmental Liability Directive. Compile and make accessible in a digital format all primary obligations on nutrient management stemming from EU law to limit the environmental footprint of farming activities. Create a Zero Pollution contribution to the European Green Deal Dataspace to improve data availability.	Pollution levels per country (carbon monoxide, ozone, nitrogen dioxide, Sulphur dioxide average levels) Eutrophication levels per country Soil contamination levels (heavy metals, pesticides). Chemical exposure levels in consumer goods. Biodiversity indices. Ecosystem service assessments (water filtration, carbon sequestration rates). Availability of datasets on pollution levels across different regions.

	on biodiversity by 25%. Reduction in waste generation is anticipated, aiming to halve residual municipal waste.	Create Destination Earth to develop a very high-precision digital model of the Earth, using Copernicus data as a key building block to monitor the state of air, waters, seas, and soil. Improved training and educational support on environmental risks, including pharmaceuticals. Develop a 'European Environment and Health Atlas'. Implement nature-based solutions to mitigate pollution. Integrate digital tools (e.g., IoT, AI) for real-time pollution monitoring.	
EU chemicals strategy for sustainability	The European Union's chemicals strategy seeks to: improve the safeguarding of citizens and the environment. promote innovation in creating safe and sustainable chemical products.	Banning the most harmful chemicals in consumers. products - allowing their use only where it is essential. Account for the cocktail effect of chemicals when assessing risks from chemicals. Phasing out the use of per- and poly-fluoroalkyl substances (PFAS) in the EU unless their use is essential. Boosting the investment and innovative capacity for the production and use of chemicals that are safe and sustainable by design and throughout their life cycle Promoting the EU's resilience in the supply and sustainability of critical chemicals. Establishing a more straightforward "one substance, one assessment" process for chemical risk and hazard assessment. Playing a leading role globally by championing and promoting high	Number of substances identified as carcinogenic, mutagenic or toxic per country Production and consumption of chemicals by hazard class per country Total GHG in the chemical industry per country Total use of chemical pesticides Total use of more hazardous pesticides Industrial chemical releases to air Industrial chemical releases to water Persistent organic pollutant emissions in Europe Concentrations of pesticides in rivers, lakes and groundwater in Europe

		standards and not exporting chemicals banned in the EU.	Population connected to at least secondary wastewater treatment Production and consumption of chemicals by hazard class Progress in the management of contaminated sites
EU Biodiversity Strategy for 2030	The biodiversity strategy aims to set Europe's biodiversity on a recovery trajectory by 2030, benefiting individuals, the climate, and the Earth. Considering the post-COVID-19 environment, the strategy focuses on fortifying societal resilience against potential threats, such as: climate change repercussions, forest fires, food scarcity, disease outbreaks. safeguarding wildlife and addressing the illegal wildlife trade.	Establishing a more extensive EU-wide network of protected land and sea areas. Introducing the EU's first Nature Restoration Law, which sets forth a broad restoration objective focused on the long-term revitalization of nature within the EU's land and marine territories, incorporating legally binding restoration targets for habitats and species. Unlocking funding for biodiversity and setting in motion a new, strengthened governance framework. Introducing measures to tackle the global biodiversity challenge.	Land Cover - Land Use maps Terrestrial protected areas coverage Natura 2000 terrestrial protected areas coverage Marine protected areas coverage Forest Maps Forest Fires indicators Drought predictor Indices Areas under organic farming Percentage of remediated sites Climatic predictor variables Vegetation Maps Ecological Habitat Maps Ecological Degradation Maps (17 indicators from ESDAC) Land Surface Temperature (LST) Indicator Biotic predictor variables (NDVI, Leaf area index, Chlorophyll Content, Soil Moisture Indices) Biomass accumulator data

Soil Monitoring Law	The Directive seeks to implement a well-structured and integrated soil monitoring framework across the EU to improve soil health continuously. The overarching objective is to achieve healthy soils by 2050 and to maintain them in a state that allows for the provision of multiple ecosystem services, thereby meeting the environmental, societal, and economic needs of the Union. Mitigate the effects of climate change and biodiversity loss, enhance resilience to natural disasters, ensure food security, and lower soil contamination to levels no longer considered detrimental to human health and the environment.	Soil districts in MS (member states). Competent authorities by all MS. Soil health and land take monitoring framework by all MS. Soil descriptors, criteria for healthy soil condition, land take and soil sealing indicators for monitoring and assessment. Measurements and methodologies. Assessment of the soil health. Sustainable soil management. Land take mitigation principles. Risk-based approach on potentially contaminated sites Identification of potentially contaminated sites. Investigation of potentially contaminated sites. Risk assessment and management of contaminated sites. Reporting by MS.	Soil Mechanical Properties Data Soil Chemical Properties Data Soil Composition Index Normalized Difference Salinity Index Ferric iron, Fe^{3+} Ferric iron, Fe^{2+} Soil Erosion Maps / Rates per country Maps of Soil Degradation per country Soil Organic Content Maps per country Soil pH Data / Maps per country Bulk density Data / Maps per country Imperviousness maps per country per land use Percentage of remediation per country Additional Soil Health Descriptors based on the Directive
EU Strategy for Data	The European data strategy seeks to establish: A unified data market that enhances Europe's	The actions are based on four pillars: A cross-sectoral governance framework for data access and use. Investments in data and strengthening of Europe's	Metadata standards for connected devices and services. Inventories of datasets

	global competitiveness and upholds data sovereignty. To defend and promote European values and rights in the digital world.	capabilities and infrastructures for hosting, processing, and using data, as well as interoperability. Empower individuals by investing in skills and in SMEs. Creation of European data spaces in strategic sectors and domains of public interest.	Energy efficiency for hosting infrastructure. Metadata and taxonomies for harmonizing datasets Cybersecurity reports.

4.3 DATA FORMATS & INTEROPERABILITY

To ensure seamless integration across platforms, Nostradamus adopts open, standardized file formats based on FAIR Data Principles. A summary of DATA types used, their formats and Compatibility is provided in Table 4. As an example, farmers using IoT weather sensors will store real-time data in MQTT format, which will be converted into CSV and JSON for integration with AI models.

TABLE 4. DATA COMPATIBILITY WITH FAIR PRINCIPLES

Data Type	**Format(s)**	**Compatibility**
Geospatial EO Data	GeoTIFF, NetCDF, HDF5	QGIS, Google Earth Engine, Copernicus DIAS
Tabular & Statistical Data	CSV, JSON, XLSX	Python (Pandas), R, Excel, OpenAIRE
AI Model Outputs	ONNX, HDF5, TFRecord	TensorFlow, PyTorch, Jupyter Notebooks
IoT Sensor Data	MQTT, XML, JSON	OpenIoT, OGC SensorThings API, Edge AI Pipeline

4.4 DATA STORAGE & PROCESSING METHODS

The Nostradamus project manages large-scale datasets using a hybrid cloud and on-premises storage approach.

4.4.1 PRIMARY STORAGE LOCATIONS

Primary storage locations include Cloud-Based Storage (Google Cloud, AWS S3, Azure Blob Storage), used for large-volume EO and AI model data; On-Premise High-Performance Computing (HPC) Servers (UNIGE, ERATOSTHENES CoE), used for secure, controlled data processing, and Data Cube Infrastructure, used to host structured, multi-dimensional EO and IoT datasets.

4.4.2 DATA PROCESSING PIPELINES

The Nostradamus project ensures that all collected data undergoes structured transformation processes to maintain consistency, quality, and interoperability across multiple research domains. Data transformation plays a crucial role in enabling seamless integration, efficient analysis, and long-term usability. By applying FAIR (Findable, Accessible, Interoperable, and Reusable) principles, Nostradamus ensures that datasets follow open, machine-readable formats, are standardized according to EU and international metadata specifications, and are preprocessed for optimal usability in AI, GIS, and statistical models.

To maximize interoperability, Nostradamus adopts widely recognized open data formats, including GeoTIFF, NetCDF, and HDF5 for geospatial Earth Observation (EO) data, CSV, JSON, and XML for IoT sensor data, and ONNX, HDF5, and TFRecord for AI model outputs. These formats ensure compatibility with platforms such as QGIS, Google Earth Engine, Copernicus DIAS, TensorFlow, and PyTorch. Prior to storage, all datasets undergo structured preprocessing workflows to enhance data integrity, including outlier detection, normalization, noise reduction, and temporal/spatial alignment. For example, Sentinel-2 images affected by cloud cover are automatically processed using cloud-masking algorithms, ensuring only valid pixels are retained for analysis. Similarly, IoT sensor readings are filtered for anomalies before ingestion into real-time databases.

To support AI-driven analytics, raw datasets are transformed into ML-compatible formats. This includes dimensionality reduction, feature engineering, label encoding, and tensor conversion into structured deep-learning files like HDF5 and TFRecord. A deep learning model predicting crop yield trends will utilize these optimized formats for high-performance training. In addition, data compression techniques such as GeoTIFF deflate compression, Parquet storage for tabular data, and HDF5 chunking are employed to reduce storage overhead while ensuring efficient retrieval. These optimization strategies enhance data accessibility while maintaining high fidelity.

All datasets are richly documented with standardized metadata to ensure transparency and discoverability. ISO 19115 metadata standards are applied to Earth Observation datasets, while Dublin Core (DCAT-AP) metadata schemas ensure integration with OpenAIRE and Zenodo repositories. AGROVOC semantic tagging is used for agricultural datasets, enhancing structured classification and enabling advanced searchability. For example, an AI model processing irrigation data will use AGROVOC-tagged metadata to filter relevant crop-related datasets.

By ensuring all datasets undergo structured transformation and format standardization, Nostradamus guarantees that its research outputs remain machine-readable, AI-compatible, and highly interoperable across diverse scientific and policy-making domains. Through these structured methodologies, data remains fully FAIR-compliant, accessible, and efficiently processed for future applications, ensuring maximum impact and reusability across research communities.

Data processing is performed using Distributed Computing (Kubernetes, Apache Spark) for scalable AI model training; Geospatial Processing Pipelines (GDAL, Rasterio, QGIS) for EO and GIS dataset transformations; and ETL (Extract, Transform, Load) Pipelines: Ensuring efficient data ingestion from IoT devices.

4.5 EXPECTED DATA VOLUME & GROWTH

The Nostradamus project is expected to generate large volumes of data across its lifecycle. Estimated storage needs and growth rate per data source is outlined in Table 5. It is accounted that in order to accommodate growth, automated storage scaling policies will be applied.

TABLE 5. ESTIMATED STORAGE NEEDS

Data Source	Estimated Storage Needs
Earth Observation (EO) Data	~20 TB
IoT Sensor Data	~5 TB
AI Model Training & Outputs	~3 TB
Socioeconomic Data	~500 GB

4.6 DATA BACKUP & REDUNDANCY

The project integrates EO, IoT, AI, and socioeconomic data sources for agricultural innovation. To prevent data loss, Nostradamus employs a multi-layered backup strategy, incorporating Incremental backups every 24 hours, full backups every 7 days; Geo-redundancy with copies stored in two separate cloud data centers; and Checksum-based integrity verification to prevent data corruption. For example, if a dataset is accidentally deleted, a versioned backup from the ERATOSTHENES CoE cloud archive will be restored using automated scripts.

Overall, all datasets follow FAIR-compliant open formats (GeoTIFF, CSV, ONNX), data is processed using cloud-HPC hybrid infrastructure, ensuring efficiency, and backup redundancy prevents data loss and ensures long-term preservation.

By structuring data acquisition, storage, and processing under a scalable, FAIR-compliant architecture, Nostradamus ensures that its research outputs are robust, reusable, and accessible to stakeholders across disciplines.

5 DATA COLLECTION, PROCESSING & STORAGE

The Nostradamus project collects, processes, and stores large-scale, multi-source datasets from Earth Observation (EO) platforms, Internet of Things (IoT) networks, AI-driven analytics, and stakeholder-generated socioeconomic data. To ensure efficient data management, security, and accessibility, this section outlines the data acquisition methods, preprocessing workflows, storage infrastructure, and backup strategies used in the project. By establishing robust and scalable data architecture, the project ensures that all datasets are: Accurately collected from trusted sources; Properly processed and cleaned for analysis; Stored securely with high availability for long-term use; Structured for interoperability and reuse across research disciplines.

5.1 DATA COLLECTION STRATEGIES

Data collection in Nostradamus follows three main approaches:

a) Automated Data Collection – Using satellite APIs, IoT networks, and AI-powered pipelines to capture high-frequency, real-time data.

b) Survey-Based & Stakeholder Data – Gathering qualitative and quantitative insights from farmers, policymakers, and industry experts.

c) Pre-Existing Datasets & Open Data Sources – Leveraging historical climate records, crop yield data, and economic reports from EU repositories.

5.1.1 EARTH OBSERVATION (EO) DATA COLLECTION

Earth Observation (EO) data are collected from satellites, e.g. Copernicus Sentinel-1 & Sentinel-2, Landsat 8 & 9, and MODIS. API Access is provided via European Space Agency (ESA) Copernicus Open Access Hub, and Google Earth Engine APIs. The Geospatial Resolution is 10m-30m per pixel, depending on the satellite source with the Data Acquisition Frequency of updating every 5-7 days. Particularly for the Nostradamus project needs, required EO data categories are identified and presented in Table 6, where the index/data, source, purpose, application and strategies or initiatives aligned with are presented per data category.

TABLE 6. EARTH OBSERVATION (EO) DATA

Category	Index/Data	Source	Purpose	Application	Strategies/Initiatives
Soil Health Monitoring	Bare Soil Index (BSI)	Sentinel-2, Landsat	Identifies exposed soil areas, distinguishes bare soil from vegetation	Maps erosion risks, plans cover cropping, and identifies degraded soils	CAP, Farm-to-Fork, Biodiversity
	Soil Adjusted Vegetation Index (SAVI)	Sentinel-2, Landsat	Reduces soil reflectance impacts in sparsely vegetated areas	Improves monitoring of semi-arid agricultural systems	Bioeconomy, Farm-to-Fork, Climate Law, CAP, Soil Monitoring Law
	Normalized Difference Salinity Index (NDSI)	Sentinel-2, MODIS	Detects soil salinity stress	Identifies salinity-prone areas for reclamation or salt-tolerant crops	Climate Law, Soil Monitoring Law
	Soil Organic Carbon (SOC) Maps	ISRIC SoilGrids, Sentinel	Tracks organic carbon levels in soils	Guides carbon farming strategies, ensures soil health sustainability	Bioeconomy, Farm-to-Fork, Climate Law, CAP, Soil Monitoring Law,
	Soil Erosion Index (SEI)	Copernicus, ESA	Maps soil erosion based on land	Supports erosion control measures such	Soil Monitoring Law, CAP, Biodiversity

			cover, slope, and rainfall	as terracing and buffer zones	
	Bulk Density Index (BDI)	Derived from EO + in situ data	Measures soil compaction / can be used for pedrotransfer functions	Guides reduced tillage practices and soil aeration strategies	CAP, Farm-to-Fork
	Topographic Wetness Index (TWI)	Sentinel-2, DEM (Digital Elevation Models)	Evaluates soil water retention and drainage	Reduces waterlogging in flood-prone farming areas	Soil Monitoring Law, Farm-to-Fork
	Surface Roughness Index	Sentinel-1 SAR	Monitors soil preparation and micro-topography	Tracks plowing quality and suitability for seeding	Soil Monitoring Law, Biodiversity
	Soil Temperature Anomalies	Sentinel-3	Monitors thermal impacts on soils from industrial or agricultural processes	Identify areas impacted by waste heat or industrial activity	Soil Monitoring Law, Climate Law
	Soil Moisture Index (SMI)	Sentinel-1 SAR, SMAP	Monitors soil moisture dynamics across landscapes	Support drought resilience planning, irrigation scheduling, and water conservation	CAP, Farm-to-Fork, Climate Adaptation
Water Quality Monitoring	Normalized Difference Water Index (NDWI)	Sentinel-2, MODIS	Monitor surface water availability and dynamics	Track water body extent and seasonal variations	Zero Pollution Action Plan, SDG6, CAP, Biodiversity Strategy
	Chlorophyll-a Concentration	Sentinel-2, MODIS	Monitor algal blooms caused by nutrient pollution	Identify eutrophication risks in freshwater and coastal ecosystems	Farm-to-Fork, Zero Pollution Action Plan

	Eutrophication Risk Index	Copernicus Marine Monitoring Service	Assess nutrient loading and algal bloom potential	Prioritize regions for nutrient runoff control	CAP, Farm-to-Fork
	Suspended Sediment Index (SSI)	Sentinel-2, Landsat	Monitor sediment loads in rivers, reservoirs, and coasts	Identify erosion sources and industrial discharge hotspots	Biodiversity Strategy, SDG6
Crop and Vegetation Monitoring	Normalized Difference Vegetation Index (NDVI)	Sentinel-2, MODIS	Tracks vegetation greenness and health	Monitor crop growth, productivity, and stress	Farm-to-Fork, CAP, Biodiversity
	Enhanced Vegetation Index (EVI)	MODIS	Provide better sensitivity in high-biomass areas	Track dense vegetation growth in forests and agricultural fields	Farm-to-Fork, CAP, Biodiversity
	Modified Soil Adjusted Vegetation Index (MSAVI)	Sentinel-2, Landsat	Reduce soil reflectance in sparse vegetation	Identify areas with poor vegetation cover requiring intervention	Farm-to-Fork
	Soil Adjusted Vegetation Index (SAVI)	Sentinel-2	Monitor vegetation in areas with significant soil brightness	Guide sustainable crop management in semi-arid and arid regions	CAP, Farm-to-Fork
	Fraction of Absorbed Photosynthetically Active Radiation (fPAR)	Sentinel-2, MODIS	Measure plant photosynthetic activity	Track productivity trends and drought impacts on crops	Climate Law, CAP

	Vegetation Productivity Index (VPI)	Sentinel-2, MODIS	Monitor aboveground biomass productivity	Support biomass and yield estimation	Farm-to-Fork, Biodiversity Strategy
	Leaf Area Index (LAI)	Sentinel-2, MODIS	Monitor vegetation canopy density	Track crop density and forest cover changes	Biodiversity, CAP
	Red-Edge Chlorophyll Index	Sentinel-2	Track chlorophyll content to monitor nutrient and water availability	Identify nutrient deficiencies and crop stress	Farm-to-Fork
	Crop Water Stress Index (CWSI)	Sentinel-3 SLSTR	Detect water stress in agricultural areas	Supports irrigation scheduling and water use optimization	CAP, Farm-to-Fork
	Land Cover Dynamics	Sentinel-2, Copernicus Land Monitoring Service	Track transitions in agricultural and natural vegetation	Monitors land-use changes and their impacts on vegetation	Biodiversity, Climate Adaptation
	Habitat Degradation Index	Sentinel-2, MODIS	Track loss of vegetation cover due to human activity	Identify degraded habitats for restoration	Biodiversity Strategy, Climate Adaptation
	Aboveground Biomass Monitoring	LiDAR, Sentinel-2	Track biomass in forests, grasslands, and agricultural fields	Evaluates carbon stocks and ecosystem productivity	Biodiversity Strategy, Climate Law
	Fire Severity Index	Sentinel-3, MODIS	Track post-fire vegetation recovery and damage severity	Guides forest and ecosystem restoration after fires	Forest Strategy, Biodiversity Strategy

	Drought Severity Index (DSI)	Sentinel-2, MODIS, ERA5 Climate Data	Tracks drought impact on vegetation productivity	Identifies areas under severe drought stress for intervention planning	CAP, Farm-to-Fork, Biodiversity Strategy, Climate Adaptation
	Vegetation Condition Index (VCI)	Sentinel-2, MODIS	Monitor vegetation stress caused by drought or environmental variability	Detect deviations in vegetation health for early warning systems	CAP, Farm-to-Fork, Biodiversity Strategy, Climate Adaptation
	NDRE (Normalized Difference Red Edge)	Sentinel-2	Monitor chlorophyll content and nutrient status in crops	Detect nutrient deficiencies and crop stress for precision agriculture	CAP, Farm-to-Fork, Biodiversity Strategy, Climate Adaptation
	LCC (Leaf Chlorophyll Content)	Sentinel-2	Track chlorophyll concentration as an indicator of crop photosynthesis	Guide fertilization and irrigation planning	CAP, Farm-to-Fork, Biodiversity Strategy, Climate Adaptation
Meteorological data, DEM, and others	Precipitation Data	ERA5, GPM	Monitor rainfall patterns affecting crop and vegetation health	Guide irrigation schedules and drought response planning	CAP, Biodiversity Strategy, Climate Adaptation
	Temperature Data	ERA5	Track seasonal temperature variations impacting crop growth	Assess heat stress and frost risks in crops	CAP, Biodiversity Strategy, Climate Adaptation
	Solar Radiation Data	ERA5	Track solar energy availability for photosynthesis	Support yield predictions and energy balance modeling	CAP, Biodiversity Strategy, Climate Adaptation
	Wind Speed and Direction	ERA5	Monitor wind effects on soil	Guide planting strategies and erosion control	CAP, Biodiversity Strategy, Climate Adaptation

		erosion and pollination		
Soil Moisture Data	SMOS, ERA5, Copernicus Land Monitoring Service	Monitor soil water availability	Support crop water stress assessments and irrigation efficiency	CAP, Farm-to-Fork, Biodiversity Strategy, Climate Adaptation
DEM (Digital Elevation Model)	Copernicus DEM	Track elevation-related habitat changes	Identify species ranges and water drainage patterns	CAP, Farm-to-Fork, Biodiversity Strategy, Climate Adaptation
Land Surface Temperature (LST)	Sentinel-3, MODIS	Track heat stress in crops and habitat area	Inform heat-tolerant crop selection	CAP, Farm-to-Fork, Biodiversity Strategy, Climate Adaptation
10m Annual Land Use Land Cover (9-class) Sentinel Hub	Sentinel-2	Classifies land into nine categories: Built, Crops, Trees, Water, Rangeland, Flooded Vegetation, Snow/Ice, Bare Ground, and Clouds	Monitors annual land cover changes for environmental and urban planning	All
Impact Observatory LULC Map	Deep learning-derived land cover classifications from Sentinel-2 imagery		Tracks human impacts on land cover for policy and restoration planning	All

5.1.2 IOT SENSOR-BASED DATA COLLECTION

IoT sensor data will be obtained throughout the project duration, for example via Soil Moisture Sensors – Measuring water retention and drought risk in agricultural fields; Weather Stations – Collecting temperature, humidity, wind speed, and precipitation data, i.e., a smart irrigation system will use IoT soil moisture sensors to trigger automated water distribution, optimizing resource efficiency.

The Nostradamus project Data needs from ground measurements are summarized in Table 7, listing each parameter needed, purpose of data acquisition and application within the project, as well as its relevant alignment with EU and international initiatives and/or strategies.

TABLE 7. GROUND MEASUREMENTS

Ground Measurements				
Parameter	**Measurement**	**Purpose**	**Application**	**Strategies/ Initiatives**
Soil Organic Carbon	Carbon content in soil (%)	Monitor soil health and carbon sequestration potential	Support carbon accounting and sustainable land management	Soil Monitoring Law, Climate Law, CAP
Soil pH	Acidity/alkalinity levels	Track chemical properties influencing crop and microbial activity	Guide liming and fertilization practices	Soil Monitoring Law, Farm-to-Fork, CAP
Bulk Density	Soil compaction (g/cm^3)	Evaluate soil structure and its impact on root penetration and water infiltration	Detect areas requiring decompaction or conservation tillage	Soil Monitoring Law, CAP
Soil Moisture	Volumetric water content (%)	Monitor water availability for crops and vegetation	Guide irrigation scheduling and drought resilience planning	CAP, Farm-to-Fork, Climate Adaptation
Nutrient Levels	Nitrogen (N), Phosphorus (P), Potassium (K)	Track soil fertility levels	Guide precision agriculture and sustainable fertilization practices	CAP, Farm-to-Fork
Heavy Metals	Cadmium (Cd), Lead (Pb), Mercury (Hg), etc.	Detect soil contamination from industrial and	Identify hotspots for remediation and pollution control	Zero Pollution Action Plan, Chemicals Strategy

		agricultural sources		
Soil Salinity	Electrical conductivity (EC)	Track soil salinity affecting crop growth	Guide reclamation and irrigation practices	CAP, Farm-to-Fork
Soil Erosion	Erosion rates (t/ha/year)	Evaluate topsoil loss and its environmental impacts	Guide soil conservation and restoration strategies	Soil Monitoring Law, Biodiversity Strategy
Vegetation Biomass	Aboveground biomass (kg/ha)	Track productivity of crops and natural vegetation	Support yield estimation and biomass energy planning	Farm-to-Fork, Biodiversity Strategy
Canopy Cover	Fractional vegetation cover (%)	Monitor vegetation density and land degradation	Identify areas requiring reforestation or crop intensification	Biodiversity Strategy, CAP
Crop Phenology	Plant growth stages (emergence, flowering, maturity)	Monitor seasonal crop development	Optimize planting and harvesting schedules	CAP, Farm-to-Fork
Soil Respiration	CO_2 flux from soil (μmol CO_2 m^{-2} s^{-1})	Measure soil microbial activity and carbon cycling	Track soil health, carbon sequestration potential, and impacts of management practices	Soil Monitoring Law, Climate Law, CAP
Rainfall , Temperature, Wind Speed	Precipitation levels (mm), Air temperatures (°C), Wind velocity (m/s)	Monitor climate impacts on crops and ecosystems, Track water availability for ecosystems and agriculture	Support drought resilience and irrigation planning, Guides crop variety selection and climate adaptation strategies	Climate Law, Farm-to-Fork, CAP, Biodiversity Strategy
Groundwater Levels	Depth to water table (m)	Monitor groundwater availability for agriculture and ecosystems	Guide sustainable water use policies	Climate Adaptation, CAP

Energy Input	Energy use in farming practices (MJ/ha)	Evaluate energy efficiency in agricultural systems	Identify energy-intensive practices to optimize and reduce energy consumption	Farm-to-Fork, Bioeconomy Strategy
Energy Efficiency	Ratio of energy output to input	Track efficiency of agricultural production	Guide transitions to energy-efficient technologies and renewable energy use	Farm-to-Fork, Zero Pollution Action Plan
Fuel Consumption	Diesel or electricity use per hectare	Measure energy use in machinery and irrigation	Identify high-energy practices for optimization	Climate Law, Bioeconomy Strategy, CAP

5.1.3 SOCIOECONOMIC & STAKEHOLDER DATA COLLECTION

Socioeconomic & Stakeholder data is collected from Farmer Surveys & Interviews, capturing agricultural practices, challenges, and decision-making factors, from EUROSTAT & FAOSTAT Market Reports, which provide historical economic trends in agriculture, while also using data from the Common Agricultural Policy (CAP) Compliance Data, to assess policy effectiveness on farm management. As an example, a policy analyst evaluating CAP subsidies will combine survey results from farmers with EUROSTAT agricultural trade data.

For the case of this project, identified Open-Source Databases useful to Nostradamus are presented in Table 8, along with the purpose of usage, type of data to be retrieved and relevance to strategies/initiatives.

TABLE 8. OPEN-SOURCE DATABASES

Open-Source Databases			
Database	**Purpose**	**Type of Data**	**Relevance to Strategy**
FAOSTAT	Track global agricultural statistics and environmental impacts	Crop yields, livestock numbers, fertilizer use, emissions	Farm-to-Fork, Biodiversity, CAP

ISRIC SoilGrids	Provide global soil property maps	Organic carbon, pH, bulk density, salinity	Soil Monitoring Law, Farm-to-Fork, Climate Law
Copernicus Land Monitoring Service (CLMS)	Map land cover, land use, vegetation, and erosion risks, bio-geophysical parameters	Habitat maps, land degradation, agricultural trends	Biodiversity, CAP, LULUCF
Copernicus Marine Monitoring Service (CMEMS)	Monitor marine ecosystems and coastal impacts	Eutrophication indices, sediment movement, algal bloom data	Zero Pollution Action Plan, Biodiversity Strategy
Copernicus Atmosphere Monitoring Service (CAMS)	Provide emissions and air quality data	NO_2, NH_3, CO_2, CH_4, O_3, PM2.5	Zero Pollution Action Plan, Chemicals Strategy
Global Biodiversity Information Facility (GBIF)	Track biodiversity and ecosystem health	Species occurrence, habitat suitability	Biodiversity Strategy
Natura 2000 Database	Track protected areas in Europe	Spatial data on biodiversity hotspots	Biodiversity Strategy, Climate Adaptation
AQUASTAT (FAO)	Analyze global water use and irrigation	Water availability, irrigation efficiency	Farm-to-Fork, CAP
Eurostat	Track EU socio-economic and agricultural trends	Rural employment, food production, trade	Farm-to-Fork, CAP, Climate Adaptation
European Soil Data Centre (ESDAC)	Offer soil health and contamination data	Heavy metals, pesticide residues, organic carbon, salinity	Zero Pollution Action Plan, Soil Monitoring Law
EarthData (NASA)	Provide global Earth observation datasets	Land cover, soil moisture, vegetation indices	Farm-to-Fork, CAP, Biodiversity
World Bank Open Data	Provide global socio-economic data	GDP, trade, energy consumption, rural livelihoods	Farm-to-Fork, Climate Adaptation, Bioeconomy Strategy
OpenStreetMap (OSM)	Offer detailed land-use and infrastructure data	Land cover, infrastructure, roads	Farm-to-Fork, CAP

European Environment Agency (EEA)	Provide environmental data and reports	Air quality, biodiversity, climate, pollution levels	Zero Pollution Action Plan, Chemicals Strategy
Global Forest Watch	Track deforestation and forest health	Forest cover loss, restoration data	EU Forest Strategy, Biodiversity Strategy
Sentinel Hub (Copernicus)	Provide access to Sentinel EO datasets	Land cover, soil moisture, vegetation indices	Farm-to-Fork, CAP, LULUCF
Desertification Information System for the Mediterranean (DISMED)	Monitor desertification risks	Soil erosion, salinity, vegetation degradation	Climate Adaptation, Biodiversity Strategy
Emissions Database for Global Atmospheric Research (EDGAR)	Track global emissions	CO_2, CH_4, NOx, SO_2, PM2.5	Climate Law, Zero Pollution Action Plan

5.2 DATA PROCESSING & STANDARDIZATION

Raw data collected from EO platforms, IoT sensors, and socioeconomic sources as described in the previous sections of this Deliverable, must be cleaned, transformed, and structured for analysis and integration, as outlined in the following paragraphs.

5.2.1 DATA PREPROCESSING & CELANING PIPELINES

Data processing and cleaning pipelines include Outlier Detection & Removal, where filtering sensor errors and misclassifications are conducted; Data Normalization & Resampling, e.g. Standardizing temporal and spatial resolutions; Noise Reduction in Satellite Data, where cloud mask algorithms are applied to remove obstructions; and AI-Powered Data Imputation, such as filling gaps in datasets using machine learning models. As an example, for Sentinel-2 images affected by cloud cover, processing using cloud-masking algorithms will be employed, ensuring that only valid pixels are used in analysis.

5.2.2 DATA TRANSFORMATION & STANDARDIZED FORMATS

To ensure interoperability, all datasets will be converted into open, machine-readable formats using the respective processing tools for each data type. For example, in the case of Geospatial EO Data, GDAL, Rasterio, QGIS will be used to convert to GeoTIFF, NetCDF output formats. For IoT sensor data, processing tools such as Python (Pandas, NumPy), Apache Kafka will be used to convert IoT data to CSV, JSON, XML formats. Furthermore, AI Model outputs will be converted to ONNX, HDF5, TFRecord formats via processing using tools such as TensorFlow and PyTorch. Survey & Market data will be processed using Excel, and PostgreSQL to convert to output types XLSX, CSV, and SQL. Accordingly, AI-generated crop risk assessments will be exported as ONNX models, ensuring compatibility with other machine learning platforms.

5.3 DATA STORAGE INFRASTRUCTURE

5.3.1 PRIMARY DATA STORAGE LOCATIONS

The Nostradamus project adopts a hybrid cloud-HPC (High-Performance Computing) storage architecture to ensure scalability, security, and real-time accessibility for handling large-scale Earth Observation (EO), IoT, and AI-driven datasets. The project's infrastructure is designed to support efficient data ingestion, processing, and retrieval, ensuring data integrity and long-term preservation. The primary data storage locations for the project include cloud storage, on-premise HPC servers, and Data Cube infrastructures, each serving a specific function in the data lifecycle.

Nostradamus leverages cloud-based storage solutions for handling high-volume datasets and real-time data ingestion from EO satellites and IoT networks. The cloud infrastructure ensures scalability, redundancy, and efficient disaster recovery mechanisms. The selection of cloud providers is based on their ability to support virtualization technologies (e.g., Docker, Kubernetes) and facilitate access to satellite Analysis Ready Data (ARD).

Primary Cloud Storage Providers such as Google Cloud Storage & AWS S3 will be used for hosting AI models, EO datasets, and IoT sensor data. Furthermore, Microsoft Azure Blob Storage will be used for supporting raw and processed geospatial datasets. Key Features of this approach are the possibilities for Redundant data replication across multiple geographic regions; Object storage for scalable, unstructured data access; and Support for Cloud-Optimized GeoTIFFs (CoG) to facilitate efficient EO data retrieval.

For real-time data processing and AI model training, the project utilizes on-premise High-Performance Computing (HPC) servers. These dedicated HPC clusters enable large-scale parallel processing of complex geospatial and machine learning datasets. The Nostradamus HPC Storage Features Low-latency access for AI training models; Secure data processing for sensitive datasets (e.g., farmer surveys, economic models); and Support for GPU-accelerated AI applications.

A key innovation in Nostradamus is the Data Cube infrastructure, which allows multi-dimensional storage and retrieval of geospatial and temporal datasets. Data Cube Technologies Implemented are: Open Data Cube (ODC) Framework for supporting on-demand spatiotemporal data analysis; and Data Cube on Demand (DCoD) for enabling dynamic creation of case-specific data cubes;

Example Use Cases for Nostradamus Data Cubes are the following: Aggregation of multi-sensor satellite imagery; Integration of IoT sensor networks for real-time environmental monitoring; and Spatial-temporal querying for agricultural risk assessments. For example, a policymaker can analyze drought stress trends queries soil moisture and NDVI indices from a spatiotemporal Data Cube, extracting relevant datasets without manual processing.

In summary, Nostradamus uses cloud storage (AWS, Azure, Google Cloud) for scalability and redundancy, HPC servers enable real-time AI model training and secure processing of sensitive data. Furthermore, Data Cubes provide structured, multi-dimensional access to EO, IoT, and economic datasets, while Geo-redundant backup strategies ensure data security and long-term availability.

By integrating cloud, HPC, and Data Cube infrastructures, Nostradamus ensures that its datasets remain highly accessible, scalable, and secure, supporting a wide range of scientific, policy-driven, and commercial applications.

5.4 BACKUP, REPLICATION & DISASTER RECOVERY

To ensure data security and resilience and mitigate data loss risks Nostradamus employs a multi-tier backup strategy through Incremental backups every 24 hours and full backups weekly; Geo-redundant storage replication in two separate cloud regions; Checksum-based integrity verification to prevent corruption; and Cold Storage for Long-Term Archival – Moving datasets to tape backups after 5 years.

5.5 DATA ACCESS, RETRIEVAL & QUERYING

All datasets will be easily accessible through RESTful APIs & OGC Web Services, for querying geospatial and IoT datasets; GraphQL APIs for AI Model Access for allowing integration with AI-driven decision-support tools; SQL-Based Data Querying for using PostGIS-enabled PostgreSQL databases for structured data retrieval.

Overall, Nostradamus collects multi-source data from EO platforms, IoT sensors, and socioeconomic surveys. Data processing includes noise reduction, outlier filtering, and AI-driven imputation, while Datasets follow open, standardized formats (GeoTIFF, CSV, ONNX) for interoperability. In addition, a hybrid cloud-HPC storage infrastructure ensures scalability and security, while automated backups, disaster recovery, and API-based querying enhance data reliability. By structuring data collection, processing, and storage under a secure, FAIR-compliant framework, Nostradamus ensures that its research outputs are highly accurate, reliable, and accessible for future scientific and policy-driven applications.

6 SECURITY, PRIVACY & CONFIDENTIALITY

The Nostradamus project prioritizes data security, privacy, and confidentiality to protect sensitive information, comply with GDPR regulations, and ensure the integrity of research datasets. As the project collects and processes large volumes of Earth Observation (EO), IoT, AI-driven, and socioeconomic data, it follows strict security protocols to prevent unauthorized access, data breaches, and misuse. This section outlines the relevant Encryption standards and access control mechanisms to protect sensitive data; GDPR compliance and ethical handling of personal data; and

Data security audits and risk mitigation strategies. By embedding security measures at every stage of data management, Nostradamus ensures that datasets remain protected, accessible only to authorized users, and ethically managed.

6.1 DATA SECURITY MEASURES

6.1.1 ENCRYPTION STANDARDS & SECURE STORAGE

The following Data encryption standards and secure storage tools are used in Nostradamus:

- AES-256 encryption for data at rest – All datasets stored in cloud and on-premise servers are encrypted using Advanced Encryption Standard (AES-256).
- TLS 1.3 encryption for data in transit – Secure transmission of datasets between partners and repositories using Transport Layer Security (TLS 1.3).
- Hashing & Digital Signatures – Secure hashing (SHA-256) and digital signatures will be applied to ensure data integrity.

6.1.2 ROLE-BASED ACCESS CONTROL (RBAC)

User roles define data access permissions based on responsibilities (e.g., Researcher, Administrator, External Partner). Furthermore, Multi-Factor Authentication (MFA) is required for restricted datasets, while Access logs track all data retrieval and modifications to prevent unauthorized actions.

6.1.3 SECURE APIS & DATA TRANSFER PROTOCOLS

The Nostradamus project utilizes the following secure APIs and Data transfer protocols, e.g. OAuth 2.0 Authentication for ensuring secure API-based access to Nostradamus datasets; Rate Limiting & Logging for preventing unauthorized bulk data downloads; and OGC Web Services for providing secure WMS/WFS endpoints for geospatial data sharing.

6.2 ETHICAL CONSIDERATIONS & GDPR COMPLIANCE

6.2.1 COMPLIANCE WITH GDPR & DATA PROTECTION LAWS

The Nostradamus project fully complies with EU General Data Protection Regulation (GDPR - Regulation (EU) 2016/679) to protect personal data and ensure lawful processing. Lawfulness, Fairness & Transparency is practiced in all data cases, therefore subjects (e.g., survey participants) will be informed of how their data is used. Furthermore, Data Minimization is employed, e.g. only necessary data will be collected, avoiding excessive personal data storage. In addition, the Right to

Access & Deletion is maintained, meaning that individuals can request access to or deletion of their personal data at any time.

6.2.2 ETHICAL REVIEW & DATA PROTECTION

The Nostradamus Ethics & Data Protection Advisors oversee compliance with GDPR and ethical AI standards. In addition, Annual Ethical Review Reports are employed to assess risks in data privacy, AI bias, and consent management.

6.3 DATA SECURITY AUDITS & RISK MANAGEMENT

6.3.1 REGULAR SECURITY AUDITS

Quarterly security reviews are conducted to assess encryption effectiveness, system vulnerabilities, and access control policies. Penetration Testing & Cybersecurity Drills are also employed as simulated cyberattacks to identify weaknesses. Furthermore, verification of Data Integrity is done using checksums and blockchain-based validation. For example, a penetration test simulating an unauthorized data access attempt ensures that RBAC and MFA mechanisms are functioning properly.

6.3.2 DATA BREACH RESPONSE & DISASTER RECOVERY

A Data Breach and Incident Response Plan outlines the proper steps to mitigate and report security breaches. In addition, Redundant Cloud Backups (stored across two separate geographic regions for resilience) are utilized. Furthermore, Cold Storage for Long-Term Archival is employed, e.g. Datasets older than 5 years moved to tape storage.

In summary, AES-256 and TLS 1.3 encryption ensure secure data storage and transmission, RBAC and MFA protect sensitive datasets from unauthorized access, and GDPR compliance ensures ethical data collection, processing, and storage. Quarterly security audits and penetration testing prevent cybersecurity threats, while Disaster recovery planning ensures resilience against data loss or breaches. By embedding strong security, privacy, and compliance mechanisms, Nostradamus guarantees that research datasets remain safe, ethically managed, and legally compliant with EU regulations.

7 DATA ACCESS, SHARING & REUSE

The Nostradamus project promotes responsible and FAIR-compliant data sharing to ensure that research outputs are widely accessible, reusable, and impactful. This section outlines data accessibility, sharing mechanisms, licensing frameworks, and reuse policies to maximize the scientific, economic, and policy impact of project datasets.

By implementing a structured data access model, Nostradamus ensures that public research data is freely available under open-access policies; Restricted datasets follow strict access control mechanisms (RBAC, MFA, OAuth); and Reusable data follows structured documentation and machine-readable licensing.

7.1 DATA ACCESS & LICENSING

7.1.1 OPEN DATA REPOSITORIES & PUBLIC ACCESSIBILITY

All publicly shareable datasets will be deposited in:

- Zenodo (Horizon Europe Open Science repository).
- OpenAIRE (FAIR-compliant EU research database).
- EU Open Data Portal (European Commission-funded projects).

Licensing Models to be utilized for Open Data include:

- CC0 (Public Domain Dedication) – No restrictions on reuse.
- CC-BY (Attribution Required) – Users must credit dataset creators.
- CC-BY-NC (Non-Commercial Use Only) – Ensuring research-based applications only.

7.1.2 RESTRICTED DATA ACCESS & AUTHENTICATION

Some datasets require controlled access due to privacy or commercial concerns. Relevant authentication-based access controls include: Role-Based Access Control (RBAC) – Permissions based on user roles; Multi-Factor Authentication (MFA) – Enhanced security for restricted datasets; and Data Use Agreements (DUAs) – Ensuring ethical dataset handling.

7.2 SHORT-TERM & LONG-TERM DATA PRESERVATION

All datasets will be stored for a minimum of 5 years post-project completion. Furthermore, high-value datasets will be archived in long-term institutional repositories (e.g., ERATOSTHENES CoE, UNIGE data centers). Data versioning & checksum validation will ensure dataset integrity over time.

8 RISK MANAGEMENT & DATA INTEGRITY

Risk management in Nostradamus is critical for ensuring data reliability, compliance, and security. The project implements a structured risk assessment framework to identify and mitigate threats related to data loss, cybersecurity breaches, ethical non-compliance, and interoperability issues.

By implementing proactive monitoring, Nostradamus ensures that Data security risks are continuously assessed and mitigated; Ethical and regulatory compliance remains a priority; and Data integrity is protected through redundancy and validation mechanisms.

8.1 RISK IDENTIFICATION& MITIGATION

8.1.1 KEY DATA-RELATED RISKS

Key data-related risks identified for the Nostradamus project are summarized in Table 9.

TABLE 9. SUMMARY OF DATA RELATED RISKS

Risk Type	Potential Impact	Mitigation Strategy
Data Breach	Unauthorized access to sensitive datasets	AES-256 encryption, MFA, and access control policies
Data Loss	Accidental deletion or corruption of datasets	Automated backups, checksum verification, geo-redundant storage
Ethical Compliance Violation	Misuse of personal or sensitive data	Ethics Board review, GDPR compliance audits
Data Interoperability Issues	Incompatibility across systems	Standardized formats (GeoTIFF, JSON, HDF5)

8.2 DISASTER RECOVERY & DATA BACKUP

Full backups are conducted weekly, with incremental backups every 24 hours. Geo-redundant storage ensures failover mechanisms for data restoration, while Data loss prevention measures include immutable storage snapshots and blockchain-based validation. For example, in case an IoT weather sensor dataset is corrupted, an earlier version stored in the disaster recovery vault is retrieved.

9 FINANCIAL PLANNING FOR DATA MANAGEMENT

Ensuring sustainable data management requires adequate financial planning for storage, security, compliance, and Open Science mandates. This section outlines the budget allocation for FAIR data implementation and long-term sustainability. By securing adequate funding, Nostradamus ensures that High-performance storage solutions remain operational; Open-access publication costs are covered; and Long-term dataset preservation remains financially viable.

9.1 LONG-TERM SUSTAINABILITY PLANNING

Post-project funding strategies will be explored to maintain dataset accessibility. It is expected that Institutional commitments (ERATOSTHENE CoE, UNIGE) will ensure that data repositories will remain active beyond the project's lifecycle. Furthermore, this can be established through Collaboration with Horizon Europe's OpenAIRE initiative for long-term digital archiving.

10 CONCLUSION & FUTURE UPDATES

The Nostradamus project Data Management Plan (DMP) establishes a comprehensive, secure, and FAIR-compliant framework for handling the project's multi-source datasets, including Earth Observation (EO), IoT sensor data, AI-driven analytics, and socioeconomic information. By integrating state-of-the-art data governance strategies, scalable storage solutions, and Open Science policies, this document ensures that all data collection, processing, storage, access, and sharing mechanisms align with Horizon Europe requirements, GDPR, and international best practices.

A key strength of Nostradamus is its hybrid cloud-HPC infrastructure, which supports scalable, real-time data ingestion and processing, ensuring that datasets remain highly available and computationally efficient. The use of Data Cubes and machine-readable APIs enhances data integration and interoperability, allowing stakeholders—including researchers, policymakers, industry leaders, and farmers—to efficiently access and analyze multi-source data in a structured manner.

The DMP outlines robust security protocols, including AES-256 encryption, TLS 1.3 secure data transmission, Role-Based Access Control (RBAC), and Multi-Factor Authentication (MFA), ensuring that all sensitive and restricted datasets are protected from unauthorized access. Ethical compliance is embedded throughout the plan, with GDPR-aligned privacy policies, anonymization of personal data, and continuous ethics reviews to maintain trust and transparency in research activities.

The risk management and compliance framework in Nostradamus ensures ongoing monitoring of data quality, security, and accessibility. Quarterly FAIR Data Audits, annual Open Science compliance reports, and penetration testing for cybersecurity provide proactive risk mitigation strategies to protect project data assets.

Moving forward, future DMP updates submitted throughout the project's 48-month duration will reflect technological advancements, evolving EU regulations, and lessons learned from the project's implementation. Updates will focus on enhancing metadata completeness, expanding AI-driven data cataloging, and improving cross-platform interoperability. Additionally, long-term sustainability strategies will be outlined, ensuring that project datasets remain accessible beyond the project's completion through institutional and open-access repositories (Zenodo, OpenAIRE, EU Open Data Portal).

By ensuring scientific reproducibility, open-access data sharing, and structured long-term preservation, Nostradamus will contribute high-value research outputs to the European scientific and agricultural communities, fostering innovation, sustainability, and informed policy-making.

11 REFERENCES

[1] European Commission Research Executive Agency. Grant Agreement GA No. 101134888. Project: 101134888 — Nostradamus — HORIZON-CL6-2023-GOVERNANCE-01 (2024)

[2] Nostradamus Consortium. Nostradamus Project Consortium Agreement (2024)

[3] Nostradamus Consortium. Nostradamus Deliverable - D7.11 General Ethical Guidelines (2024)

[4] European Parliament. 2023. "EU AI Act: First Regulation on Artificial Intelligence." European Parliament, June 1, 2023. *https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence*.