ENVRI-Hub
NEXT

D11.1 Legal Framework for Cross-Domain AAI System

Recommendations for Standardised Data and Metadata Policies

Status: Under EC Review

Dissemination Level: Public

Funded by
the European Union

Disclaimer: Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them

Abstract

Keywords Cross-domain Authentication and Authorisation Infrastructure (AAI); Data and Metadata Licensing; Legal Framework; FAIR Principles; Research Infrastructures (RIs); GDPR Compliance

This deliverable, part of ENVRI-Hub NEXT Work Package 11 (WP11), focuses on defining recommendations pertaining to a legal framework for cross-domain Authentication and Authorisation Infrastructure (AAI) systems, ensuring secure and seamless access to data and services across Research Infrastructures (RIs). Based on landscape analyses of current data and metadata licensing practices, as well as AAI implementations, it provides recommendations to standardize policies while addressing legal and technical boundary conditions.

Findings highlight diverse licensing practices, with Creative Commons Attribution 4.0 (CC-BY-4.0) and Creative Commons Zero 1.0 (CC0-1.0) being predominant. The analysis also underscores the importance of aligning metadata practices with FAIR principles and European standards. Nine RIs have implemented AAI systems using federated identity protocols, emphasizing role-based access control and interoperability with platforms like the European Open Science Cloud (EOSC).

The deliverable recommends harmonizing data and metadata policies, adopting federated AAI standards, and ensuring GDPR compliance. This framework adopts open science, interoperability, and secure cross-domain collaboration, supporting ENVRI-Hub NEXT's goal of enabling interdisciplinary, science-based services.

Revision History			
Version	**Date**	**Description**	**Author/Reviewer**
V 0.1	27/11/2024	ToC and First Draft	Shridhar D. Jawak (NILU) Markus Fiebig (NILU) Claudio Dema (CNR)
V 0.2	14/01/2025	Review Comments	Nicolas Liampotis (GRNET S.A.) Alessandro Turco (EPOS ERIC)
V 0.3	28/01/2025	Revision	Shridhar D. Jawak (NILU) Markus Fiebig (NILU) Claudio Dema (CNR)
V 0.4	29/01/2025	Confirmation on Revision	Nicolas Liampotis (GRNET S.A.) Alessandro Turco (EPOS ERIC)
V 1.0	**31/01/2025**	**Final**	Shridhar D. Jawak (NILU) Markus Fiebig (NILU) Claudio Dema (CNR)

Document Description			
Legal Framework for Cross-Domain AAI System Recommendations for Standardised Data and Metadata Policies			
Work Package Number 11			
Document Type	Deliverable		
Document Status	Under EC Review	**Version**	V 1.0
Dissemination Level	Public		
Copyright Status	This material by Parties of the ENVRI-Hub NEXT Consortium is licensed under a Creative Commons Attribution 4.0 International License.		
Lead partner	NILU		
Document Link	**https://documents.egi.eu/document/4034**		
DOI	**https://zenodo.org/records/14780481**		
Author(s)	- Shridhar D. Jawak (NILU) - Markus Fiebig (NILU) - Claudio Dema (CNR)		
Reviewers	- Nicolas Liampotis (GRNET - Alessandro Turco (EPOS)		
Moderated by:	- Matteo Agati (EGI Foundation)		
Approved by:	Development Steering Board (DSB)		

Terminology / Acronyms	
Term/Acronym	**Definition**
AAI	Authentication and Authorisation Infrastructure
RIs	Research Infrastructures
FAIR	Findable, Accessible, Interoperable, and Reusable
GDPR	General Data Protection Regulation
EOSC	European Open Science Cloud
CC-BY-4.0	Creative Commons Attribution 4.0 International
CC0-1.0	Creative Commons Zero v1.0 Universal
M2M	Machine-to-Machine

Table of Contents

Table of Figures

Table of Tables

Executive Summary

The deliverable **"Legal Framework for Cross-Domain AAI Systems"** addresses the critical need for secure, interoperable, and legally compliant Authentication and Authorisation Infrastructure (AAI) systems within the ENVRI-Hub NEXT project. This framework is essential to enable seamless access to data and services across diverse Research Infrastructures (RIs), aligning with FAIR principles and European Open Science Cloud (EOSC) standards.

The report begins with an analysis of the current landscape, revealing diverse practices in data and metadata licensing among RIs. Creative Commons Attribution 4.0 (CC-BY-4.0) and Creative Commons Zero 1.0 (CC0-1.0) are the predominant licenses, demonstrating a commitment to openness and accessibility. However, inconsistencies in licensing practices highlight the need for harmonised policies to enhance interoperability and compliance.

The document further explores the implementation of AAI systems. Nine (out of 13) RIs have adopted federated identity solutions, leveraging solutions such as eduGAIN and ORCID for secure user authentication and authorisation. Role-based and attribute-based access controls dominate, with an emphasis on ensuring compatibility with EOSC interoperability guidelines. However, challenges persist, including gaps in GDPR compliance, cross-border data sharing, and equitable access for non-EU or unaffiliated users.

Key recommendations focus on:

1. Adopting standard licensing frameworks to promote consistency and support open science initiatives.
2. Enhancing metadata discoverability through machine-readable formats aligned with European standards.
3. Developing a unified legal framework to address GDPR compliance and cross-border data sharing.
4. Encouraging interoperability through the adoption of the AARC Blueprint Architecture (BPA) and federated AAI systems.

The deliverable underscores the importance of integrating technical, legal, and governance measures to create a robust, FAIR-compliant AAI ecosystem. This framework provides a foundation for adopting interdisciplinary collaboration, enhancing resource accessibility, and supporting ENVRI-Hub NEXT's mission to enable science-based services across research domains.

1. Introduction

1.1. Background

The ENVRI-Hub NEXT project serves as a foundation for advancing interdisciplinary research by enabling seamless access to data, tools, and services across diverse RIs. As modern scientific endeavors increasingly span multiple domains, a unified approach to data access, management, and sharing is critical to address the challenges posed by fragmented infrastructures and inconsistent access policies. Central to this mission is the implementation of a cross-domain AAI system, underpinned by a robust legal framework.

A cross-domain AAI system is essential for facilitating secure and interoperable access to resources across RIs. However, technical solutions alone are insufficient to address the broader challenges of governance, compliance, and inclusivity. A comprehensive legal framework is necessary to define standardized policies and practices that align with national and international laws, such as the General Data Protection Regulation (GDPR). The legal framework also ensures that the diverse needs of stakeholders—ranging from researchers and citizen scientists to industrial partners—are met while maintaining the integrity and security of shared data and services.

The legal framework for a cross-domain AAI system addresses several critical aspects:

- Data and metadata licensing: Establishing clear and consistent licensing policies to facilitate open access while protecting intellectual property rights.
- Privacy and security: Ensuring compliance with data protection laws, safeguarding user information, and maintaining trust.
- Interoperability: Harmonizing authentication and authorisation protocols across RIs to enable seamless collaboration and integration with platforms such as the European Open Science Cloud (EOSC).
- Inclusivity: Providing equitable access to users from non-EU countries or non-federated institutions, thereby broadening the scope and impact of ENVRI-Hub NEXT.

The ENVRI-Hub NEXT project recognises the importance of aligning legal and technical considerations to create a secure, FAIR-compliant, and user-friendly environment. This deliverable focuses on developing recommendations for a legal framework that not only supports the implementation of a cross-domain AAI system but also ensures that it operates within defined legal and ethical boundaries. By addressing issues such as data ownership, user accountability, and cross-border access, the framework aims to foster trust, collaboration, and innovation across the research community.

1.2. Relevance

ENVRI-Hub NEXT project's Work Package 11 (WP11) and WP12 play a critical role in advancing the interdisciplinary objectives of the project by focusing on enabling essential services that enhance data accessibility, interoperability, and security. A key aim of this work package is to establish a standardised process for harmonising vocabularies and metadata schemas across environmental domains, thereby improving data discovery and fostering interoperability. It also involves the full implementation of a cross-domain AAI system. In addition, WP11 is tasked with developing legal framework recommendations for multi-RI AAI systems. This includes defining recommendations for a machine-to-machine (M2M) legal framework that is vital for enabling cross-domain access to RI data and digital products. Furthermore, the work package seeks to identify a unified AAI system that can serve the entire ENVRI RI landscape while ensuring seamless integration with broader digital ecosystems, such as the European Open Science Cloud (EOSC). By achieving these objectives, WP11 contributes significantly to the overarching mission of ENVRI-Hub NEXT to deliver robust, interdisciplinary, and science-based services. This will complement the work associated with the central catalogue in WP7/8.

This deliverable will be based on outcomes of Task 11.4: Machine-to-Machine (M2M) legal framework, milestone (M16) on RIs landscape analysis on metadata schema and services, and milestone (M17) relevant to landscape analysis on RIs AAI systems usage, maturity, and compatibility with the EOSC AAI interoperability guidelines. This deliverable provides the current state of data and metadata standards within the consortium with legal boundary conditions. Besides, it provides recommendations for standardised data and metadata policies meeting RIs legal boundary conditions for cross-domain AAI system.

1.3. Status from the Landscape Analysis

To provide the basis and relevance to the recommendations for standardised data and metadata policy, here, we summarise results from landscape analysis (M16) on AAI and data/metadata policy across RIs to provide the current status of Authentication and Authorisation infrastructure (AAI) and data/metadata licensing across RIs.

1.3.1. Licenses (Data)

A total of 13 (out of 14) RIs responded to the landscape analysis survey. Most of the questions under this section of the survey were answered showing the robustness of the current analysis.

An analysis of RI data licensing practices reveals a diverse approach across institutions. Eleven RIs currently employ one or more licensing models to govern the use of their data. One RI plans to implement licensing within the next two years. Notably, eleven RIs have publicly accessible data policy documents, with one additional institution planning to make their policy public soon. While one RI currently operates without a specific license and has no immediate plans to adopt one, the overall trend indicates a growing recognition of the importance of data licensing and open data practices within the RI community.

The most commonly adopted license among these RIs is the Creative Commons Attribution 4.0 International (CC-BY-4.0), followed by the Creative Commons Zero v1.0 Universal (CC0-1.0).

While the majority of RIs utilise a single license, four institutions employ multiple licenses to address diverse data types and usage scenarios.

1.6.3.1. Creative Commons Attribution 4.0 International (CC-BY-4.0)

The primary rationale for adopting CC-BY-4.0 is its balance between ensuring proper attribution to data originators and promoting open data sharing. This license is widely recognised and recommended within the ENVRI community, making it a suitable choice for RIs. Additionally, it aligns with the preferences of many research groups and individual researchers who contribute data. In some cases, this license type is requested by data providers.

1.6.3.2. Creative Commons Zero v1.0 Universal (CC0-1.0)

CC0-1.0 is often employed in situations where maximum data accessibility and reuse are desired, such as in citizen science projects. Its simplicity and lack of restrictions make it a convenient choice for receiving data from diverse sources, particularly when preserving the anonymity of humans, e.g. under European GDPR law, is a concern. Due to the lack of an attribution requirement, the corresponding metadata doesn't need to be provided either.

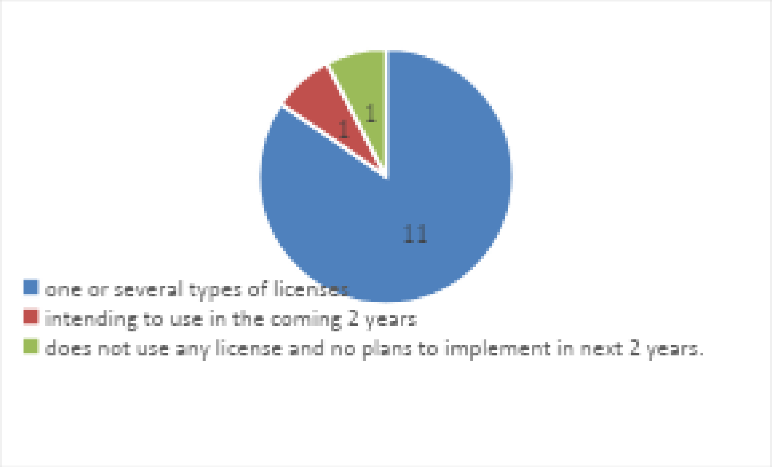

Figure 1: Status of usage of licenses for the data produced by RIs within ENVRI-Hub NEXT.

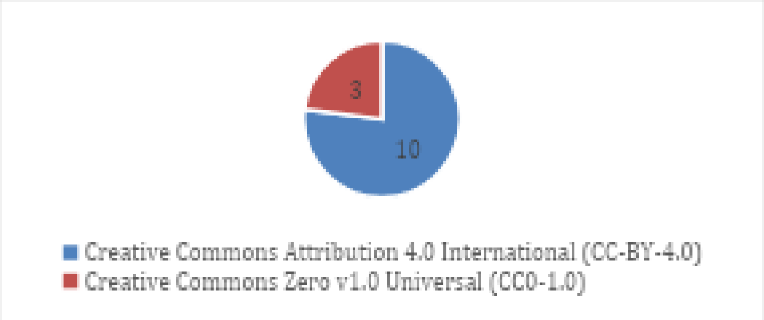

Figure 2: Comparison of usage of licenses within ENVRI-Hub NEXT.

Table 1: Comparison of licensing strategies across RIs

Licensing strategy	**# RIs adopting***	**Primary benefits**	**Challenges**
Single License (CC-BY-4.0)	7	Ensures proper attribution, aligns with FAIR	Limited flexibility for diverse use cases
Single License (CC0-1.0)	1	Maximises accessibility, supports citizen science	May not meet all stakeholder needs
Multiple Licenses	4	Tailored to data types and stakeholder requirements	Increases complexity in policy management
No License	1	Simplicity (if data is openly available)	Lack of clarity and legal risks

1.6.3.3. Multiple Licenses

The use of multiple licenses is often driven by the specific needs of data providers and the nature of the data itself. CC-BY-4.0 is frequently used for research-driven data, while CC0-1.0 is employed for data where maximum accessibility and minimal restrictions are prioritised. This flexibility allows RIs to tailor their licensing strategies to different data sets and usage scenarios. Another rational for multiple licensing is to provide unrestricted access to scientific and academic communities or stakeholders while other options for industrial stakeholders.

1.6.3.4. Brief Findings (Data Licenses)

- Diverse licensing practices: RIs exhibit a variety of data licensing practices. Eleven RIs currently use one or more licensing models, with one planning to adopt licensing soon. One RI operates without a license and has no immediate plans to implement one;
- Public accessibility of policies: Eleven RIs have publicly accessible data policy documents, with one additional institution planning to publish its policy in the near future;
- Predominant licenses: The Creative Commons Attribution 4.0 International (CC-BY-4.0) is the most commonly adopted license, followed by Creative Commons Zero v1.0 Universal (CC0-1.0). These licenses align with the needs of open data practices and research communities;
- Use of multiple licenses: Four RIs employ multiple licensing strategies to cater to diverse data types and usage scenarios. This approach provides flexibility for both open access and more restrictive usage depending on stakeholder requirements;
- Rationale for license choices:
    - **CC-BY-4.0** ensures proper attribution while promoting data sharing, making it suitable for research-driven data;
    - **CC0-1.0** maximizes accessibility and reuse, particularly for citizen science projects privacy issues, and data requiring minimal restrictions.

1.6.3.5. Implications (Data Licenses)

- Growing recognition of licensing importance: The widespread adoption of licensing reflects an increasing acknowledgment of its role in fostering data sharing and openness within the RI community;
- Alignment with open science goals: The use of CC-BY-4.0 and CC0-1.0 supports open science initiatives by facilitating proper attribution, accessibility, and reuse of data;
- Enhanced interoperability and collaboration: Publicly accessible data policies and standardised licenses improve interoperability and encourage cross-institutional collaboration;
- Adaptability to stakeholder needs: Employing multiple licenses allows RIs to address the diverse requirements of scientific, academic, and industrial stakeholders;
- Encouragement for non-adopting RIs: The trend towards licensing could serve as an incentive for the remaining RIs to adopt and publish their data policies, ensuring consistency across the RI landscape.

1.3.2. Licenses (Metadata)

The analysis of metadata licensing practices among RIs reveals a strong emphasis on open data and open metadata principles. The most commonly adopted licenses for both data and metadata are Creative Commons Attribution 4.0 International (CC-BY-4.0) and Creative Commons Zero v1.0 Universal (CC0-1.0).

The primary reasons for using CC-BY-4.0 for metadata are:

- Community alignment: Adhering to the recommendations of the ENVRI community and promoting FAIR data principles;
- Maximising accessibility: Ensuring that metadata is easily and freely accessible to a wide range of users;
- Encouraging openness: Promoting the use of open licenses for metadata to facilitate data sharing and reuse;
- Improving discoverability: Making metadata machine-readable to enhance its discoverability through search engines and data portals;
- Assure visibility of RIs: the attribution requirement assures that the role and contribution of RIs are visible, e.g. in data discovery portals.

The primary reasons for using CC0-1.0 for metadata are:

- Community alignment: Aligning with the licensing practices of other RIs;
- Citizen science: In the case of citizen science contributions, CC-0 is commonly used to preserve anonymity;
- Maximising accessibility: Making metadata freely available without any copyright restrictions;

- Simplifying licensing: Providing a simple and straightforward licensing option for data received from multiple sources.

Two RIs plan to adopt CC-BY-4.0 or CC0-1.0 for their metadata in the next two years. The goal is to ensure proper attribution to the RIs, maximize metadata accessibility, and align with European-level standards. By using these licenses, RIs aim to simplify metadata management, improve data discoverability, and facilitate data sharing and reuse.

Some RIs do not use licensing as metadata are openly available through standard endpoints which do not commonly have a license.

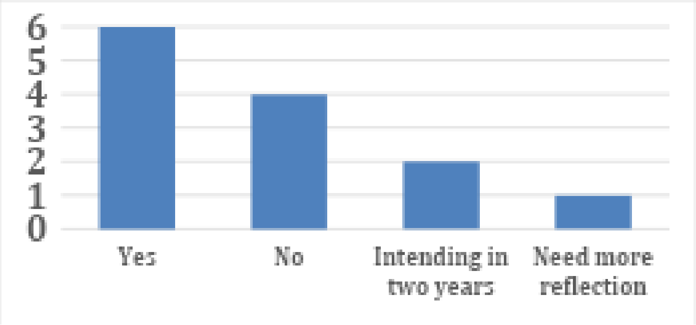

Figure 3: Status of usage of licenses for the metadata produced by RIs within ENVRI-Hub NEXT.

1.3.2.1. Brief Findings (Metadata Licenses)

- Strong emphasis on open metadata: RIs prioritise open data and metadata principles, with Creative Commons Attribution 4.0 International (CC-BY-4.0) and Creative Commons Zero v1.0 Universal (CC0-1.0) being the most commonly adopted licenses;
- Key drivers for CC-BY-4.0 usage:
  - Alignment with ENVRI community recommendations and FAIR data principles:
  - Maximised accessibility for a broad user base;
  - Encouragement of openness to promote sharing and reuse;
  - Enhanced discoverability through machine-readable metadata;
  - Visibility of RIs in data discovery portals.
- Key drivers for CC0-1.0 usage:
  - Alignment with other RIs' practices;
  - Utility in citizen science projects to preserve contributor anonymity;
  - Removal of copyright restrictions for unrestricted accessibility;
  - Simplification of licensing for data from diverse sources.
- Adoption plans: Two RIs plan to implement CC-BY-4.0 or CC0-1.0 licenses within two years to improve metadata accessibility, ensure proper attribution, and align with European standards;

- Licensing exceptions: Some RIs do not apply for specific licenses as their metadata is openly available through standard endpoints without restrictions.

1.3.2.2. Implications (Metadata Licenses)

- Enhanced interoperability and accessibility: The widespread adoption of open metadata licenses promotes interoperability and ensures metadata is freely accessible to a global audience;
- Alignment with FAIR and European standards: Using CC-BY-4.0 and CC0-1.0 supports FAIR principles and aligns RIs with European-level metadata sharing standards;
- Simplification of metadata management: The use of standardized licenses simplifies licensing processes and reduces administrative complexity, particularly for metadata from diverse contributors;
- Encouragement for licensing adoption: The trend towards open metadata licensing may encourage non-licensed RIs to formalize their practices, fostering consistency across the RI landscape;
- Increased discoverability: Machine-readable metadata under open licenses enhances visibility through search engines and portals, benefiting both data providers and users.

1.3.3. Authentication and Authorisation Infrastructure (AAI)

Of the 14 surveyed RIs, 12 responded to the AAI section of the survey, providing insights into their approaches to implementing AAI.

1.3.3.1. Availability and Rationale behind Establishment of AAI for RIs

Nine of the surveyed RIs have implemented their own AAI systems. The primary reasons for establishing AAI include:

- Resource management: Tracking resource usage, especially computational power consumption, to ensure efficient allocation and cost recovery;
- Access control: Authorising specific users to perform administrative tasks, such as uploading data and managing user accounts;
- Security and compliance: Enforcing security measures, protecting sensitive data, and ensuring compliance with relevant regulations;
- User management: Streamlining user registration, authentication, and authorisation processes, and managing user roles and permissions;
- Policy relevance: By requiring user registration and acceptance of data licenses, RIs can ensure compliance with data policies and ethical guidelines. Furthermore, AAI enables granular control over access permissions, allowing RIs to manage user roles and enforce specific policies.

While three of the surveyed RIs have opted not to implement an AAI system due to the open nature of their data and metadata, the trend is towards adopting AAI to enhance security, efficiency, and compliance.

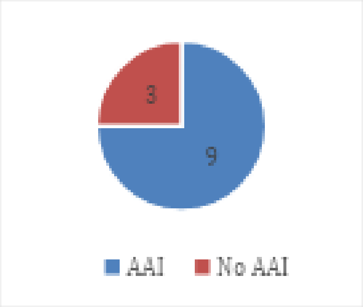

Figure 4: Status of usage of AAI by RIs within ENVRI-Hub NEXT

1.3.3.2. Access to AAI for Users of RI:

The majority of RIs surveyed do not have users who lack access to an AAI. However, some challenges exist for specific user groups:

- Small Institutions: Institutions that are not part of a federation may struggle to provide identity-provider services for their users;
- Non-EU Users: Users from outside the EU may face difficulties accessing RI services if their institutions are not part of federated authentication systems.

For RIs that currently do not have an AAI, the primary reason is the open nature of their data. However, they anticipate the need for an AAI in the future to regulate access to specific services.

1.3.3.3. AAI Authentication:

The most common authentication options available within RI AAI systems are:

- Institutional accounts (eduGAIN): Leveraging federated identity providers to provide seamless access for researchers affiliated with academic institutions;
- ORCID: Using ORCID IDs to identify researchers and track their contributions;
- Social media accounts: Allowing users to log in with their social media credentials (e.g., Google, Facebook, Twitter);
- Local accounts: Creating individual accounts directly within the RI's system.

Based on the survey responses, the following insights can be drawn regarding the required user attributes for AAI identification:

- Name and Email: This is the most common combination of attributes, required by a significant number of RIs. It provides basic identification information and is sufficient for many use cases;
- Email: Some RIs rely solely on email addresses for identification, which can be sufficient if email addresses are unique and reliable;
- Name, Email, and Affiliation: This combination is used by several RIs to identify users and determine their access privileges based on their institutional affiliation;

- Globally Unique Identifier (GUID): A few RIs require a GUID, such as an ORCID ID, to ensure the unique identification of users, especially in cases where email addresses may not be unique or reliable;
- EGI System: One RI relies on the EGI system for user identification, leveraging its existing AAI.

While email addresses are commonly used for identification, it is important to note their limitations as reliable user identifiers. Emails are not inherently stable; they can be reassigned, deactivated, or changed by users. Additionally, relying on email alone does not account for scenarios such as shared role-based accounts (e.g., emails (re)assigned to employees with specific roles). The same email can sometimes correspond to different people, such as in cases where shared or generic email addresses (e.g., *admin@university.edu*) are used, or when institutions recycle email addresses for new users. In such situations, relying solely on email does not ensure unique identification. On the other hand, it's common for one person to have multiple email addresses, and this typically isn't an issue as long as systems can effectively link those emails to the same individual. These scenarios underscore the importance of combining email with other attributes for reliable user identification.

Recommendation: To address these limitations, the project should emphasize the need for more robust identifiers, such as the OASIS Subject Identifier. This identifier is the standard for academic identity providers from eduGAIN as it provides greater assurance of uniqueness and persistence over time. More details can be found at https://refeds.org/category/personalized.

1.3.3.4. Handling of AAI Authentication:

RIs employ a variety of authorisation mechanisms to control access to resources and services. Common approaches include:

- Role-based access control: Assigning specific roles to users based on their capabilities and permissions;
- Affiliation-based access control: Granting access based on the user's affiliation with a specific institution or organization;
- Group-based access control: Using groups to define access permissions for specific sets of users;
- Identity assurance-based access control: Considering factors like identity proofing and affiliation freshness to determine access levels.

While some RIs use multiple mechanisms in combination, others rely on simpler approaches. The choice of authorisation mechanism depends on the specific needs of the RI, the complexity of its services, and the security requirements.

In most cases, the management of user capabilities and permissions is handled within the RI itself. However, some RIs may rely on external identity providers or federation services to manage certain aspects of authorisation.

1.3.3.5. Brief Findings (AAI)

- Diverse authorisation approaches: RIs employ a variety of authorisation mechanisms to control access to resources and services.
- Role-based access control dominance: It is the most prevalent method, with many RIs assigning specific roles to users based on their capabilities and permissions.
- Affiliation-based access control: This method is less common but is used by some RIs to grant access based on users' institutional affiliations.
- Group-based access control: A smaller number of RIs utilise group-based access control to manage permissions for specific groups of users.
- Identity assurance: While less common, some RIs consider identity assurance factors like strong authentication and affiliation freshness to enhance security.
- Centralized management: Most RIs manage user capabilities and permissions within their own infrastructure, although some may rely on external identity providers or federation services.

1.3.3.6. Implications (AAI):

- Security and privacy: A diverse range of authorisation mechanisms can enhance security but also introduces complexity. RIs must carefully balance security requirements with user experience;
- Interoperability: As RIs increasingly collaborate, interoperable authorisation mechanisms are crucial for seamless access to shared resources;
- User experience: User-friendly authentication and authorisation processes are essential for maximising user adoption and satisfaction;
- Future trends: As technology evolves, RIs may explore advanced authorisation techniques, such as attribute-based access control and policy-based access control.

By understanding the various authorisation mechanisms and their strengths and weaknesses, RIs can make informed decisions about how to best protect their resources and services while providing a positive user experience.

1.3.3.7. AAI for various Service Types offered by RIs

Based on the survey responses, the most common service types for RIs are:

- Browser-accessible services: These services can be accessed directly through a web browser, providing a user-friendly interface;
- API-based services: These services rely on APIs to interact with other services or systems. They can be consumed by users directly or by other services on behalf of users;
- Client-based services: These services use client applications to interact with service APIs, often using delegated user identities.

The survey responses reveal most appropriate service type for services in RIs listed with decreasing popularity is as follows (see also **Figure 5**): (1) 10 RIs: Browser Accessible Service:,

(2) 9 RIs: API Consumed by or on behalf of Users, (3) 8 RIs: API Consumed by Services, (4) 5 RIs: Client consuming other Service APIs using its own client identity and (5) 4 RIs: Client consuming Service APIs using delegated user identities:

1.3.3.8. Offline Access and Authentication Protocols:

- Limited offline access: A small number of RIs require user authentication and authorisation even when users are not actively logged in. This is typically for specific services or datasets that require access control;
- Dominant authentication protocols: OpenID Connect/OAuth 2.0 and SAML 2 are the most commonly used authentication protocols, providing secure and standardized access to RI resources.

1.3.3.9. AARC Blueprint Architecture (BPA):

- Adoption and implementation: Several RIs are either currently implementing or considering the adoption of the AARC BPA, which provides a reference architecture for AAI systems;
- Custom implementations: Some RIs have developed their own AAI architectures, often tailored to their specific needs.

Overall, RIs are increasingly adopting modern authentication and authorisation protocols and aligning their AAI implementations with industry standards.

1.3.3.10. AAI Interoperability

Several RIs have established federated access with other RIs, allowing users to seamlessly access services across different platforms. For example, EPOS users can access EMSO and AnaEE services, and LifeWatch users can access services through the EGI Federation. SeaDataNet also leverages Marine-ID to provide access to multiple services.

Additionally, some RIs, like AnaEE and EPOS, have reciprocal federations, allowing users from both RIs to access each other's services. This interoperability enhances collaboration and data sharing among researchers.

1.3.3.11. *Data Protection and Security*

Some RIs have implemented robust data protection and security measures. A majority of RIs have designated a GDPR Data Controller to ensure compliance with data privacy regulations. Additionally, many RIs have a designated security contact to handle security incidents and maintain the security posture of the AAI. While some RIs adhere to recognised security frameworks like SIRTFI, others are still evaluating their security practices.

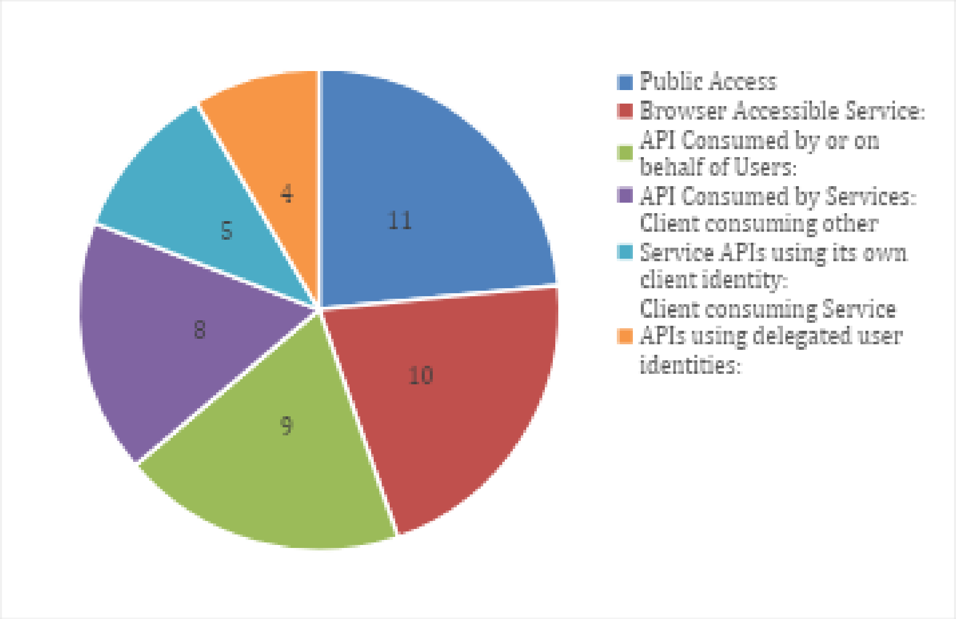

Figure 5: Overview of different service types provided by RIs

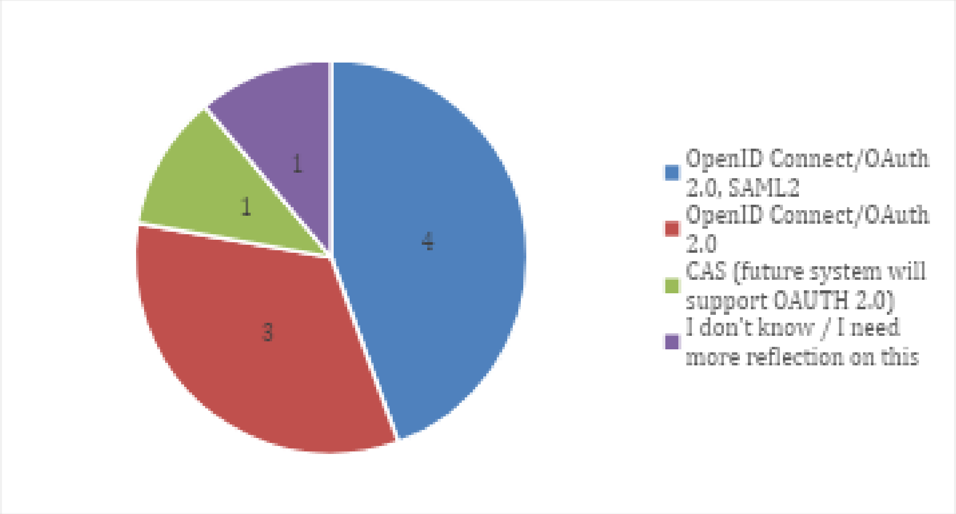

Figure 6: Authentication & Authorisation protocols supported by AAI within ENVRI-Hub NEXT RIs

1.4. Current Data and Metadata Policies and Legal Boundary Conditions

The analysis of the current landscape reveals significant insights into the data and metadata policies adopted by RIs and the legal constraints shaping these practices. Key observations are as follows:

1.4.1. Data Licensing Practices

- Diverse licensing models: RIs exhibit a variety of licensing practices, with most opting for standard licenses like Creative Commons Attribution 4.0 International (CC-BY-4.0) and Creative Commons Zero v1.0 Universal (CC0-1.0);
- Open access focus: A strong trend toward promoting open access through licensing aligns with the FAIR data principles:
- Flexibility for stakeholders: Use of multiple licenses allows differentiation between open access for the academic community and more restrictive options for industrial stakeholders.

1.4.2. Metadata Licensing Practices

- Alignment with FAIR principles: Metadata licensing predominantly supports FAIR principles, ensuring discoverability and accessibility;
- Licensing options: Both CC-BY-4.0 and CC0-1.0 are popular choices, with the latter enabling unrestricted use, particularly in citizen science;
- Non-licensed metadata: Some RIs opt not to license metadata but make it openly accessible through standard endpoints, reflecting an alternative approach to openness.

1.4.3. Legal Boundary Conditions

- GDPR compliance: Data privacy and protection measures are mandatory under EU regulations, necessitating a GDPR Data Controller for RIs;
- Cross-border challenges: Legal interoperability and data transfer restrictions across jurisdictions require careful policy design;
- Differentiated accessibility: Some RIs impose legal constraints limiting certain user groups' access, e.g., non-EU users or unaffiliated institutions.

1.4.4. AAI

- Implementation variance: Nine RIs have implemented AAI systems, leveraging role-based and attribute-based access controls to manage user permissions;
- Standards adoption: Most RIs use federated identity providers like eduGAIN and ORCID for authentication, ensuring compliance with institutional and international standards;
- Interoperability: Federated access between RIs is increasingly common, fostering cross-institutional collaboration while addressing security and compliance concerns.

1.5. Recommendations for Standardised Data and Metadata Policies

To address the diverse approaches and legal boundary conditions, the following recommendations aim to standardise data and metadata policies across RIs:

1.5.1. Data Licensing

- Adopt standard licenses: Encourage universal adoption of CC-BY-4.0 and CC0-1.0 to ensure consistency, openness, and attribution across RIs;
- Multiple licensing framework: Standardise the use of multiple licenses for different datasets, accommodating both unrestricted access for open science and selective access for industry stakeholders;
- Transparent policy publication: Require all RIs to publish data licensing policies openly to enhance transparency and user trust.

1.5.2. Metadata Licensing

- Promote open metadata: Standardise CC-BY-4.0 for research metadata to ensure attribution and CC0-1.0 for projects emphasizing maximum openness;
- Metadata discoverability: Make metadata machine-readable and compliant with European standards to improve accessibility and FAIR compliance;
- Unified metadata policies: Provide clear guidelines for licensing or ensuring open access to metadata, especially through API endpoints.

1.5.3. Legal and Compliance Framework

- Harmonise legal practices: Develop a shared legal framework ensuring GDPR compliance while enabling data sharing across borders through standardized agreements;
- User identity assurance: Incorporate strong identity assurance mechanisms (e.g., ORCID) to ensure unique user identification while adhering to privacy laws;
- Policy harmonisation: Align metadata and data policies with national and international legal frameworks to support seamless cross-border data sharing.

1.5.4. AAI System Recommendations

- Federated identity standards: Mandate the use of federated identity systems, such as eduGAIN, to allow users to access resources with their existing academic credentials, minimising the need for creating additional accounts. In addition, leverage widely adopted protocols like OpenID Connect and OAuth 2.0 to ensure standardised and interoperable authentication and authorisation processes;
- Interoperability mechanisms: Enhance collaboration by ensuring AAI systems support federation and reciprocal access agreements;
- Security enhancements: Align AAI systems with well-established security frameworks such as SIRTFI (https://refeds.org/sirtfi) standards and maintain designated security and GDPR contacts within RIs.

1.5.5. Standardized Accessibility Policies

- Inclusivity: Develop policies to accommodate non-EU and non-federated users through local account provisions or alternative identity systems;
- Cross-platform integration: Facilitate interoperability through modular architectures like the AARC Blueprint Architecture (BPA).

1.5.6. FAIR Data and Metadata Policies

- Enforce FAIR compliance: Ensure all RIs adopt policies embedding Findable, Accessible, Interoperable, and Reusable principles;
- Monitor and update policies: Establish a mechanism for continuous monitoring and updating of data and metadata policies to adapt to evolving legal and technical landscapes.

1.6. Legal Framework for Cross-Domain AAI for ENVRI-Hub NEXT

This legal framework establishes the principles, policies, and procedures for implementing a cross-domain AAI system within the ENVRI-Hub NEXT project. It aligns with GDPR, FAIR principles, and the recommendations outlined in the AARC Policy Development Kit (**https://aarc-community.org/policy/policy-development-kit/**), Privacy Policy, and Acceptable Use Policy templates.

1.6.1. Scope and Objectives

The legal framework aims to:

- Facilitate secure and seamless access to data and services across RIs within the ENVRI-Hub NEXT ecosystem through cross-domain AAI;
- Ensure compliance with legal and regulatory requirements, including GDPR;
- Promote open science by adhering to FAIR data principles;
- Foster interoperability between RIs through standardized policies and practices.

1.6.2. Key Legal and Policy Components

1.6.2.1. Privacy Policy

The Privacy Policy will ensure compliance with GDPR and outline the data processing, storage, and protection mechanisms for users' personal data. The privacy policy will be developed using the template and guidance provided by the **AARC Policy Development Kit (PDK)**.

Key elements include:

- Purpose of data collection: Clearly define the purposes for which user data will be collected, such as authentication, authorisation, and resource management;

- Data minimization: Limit the collection of personal data to only what is necessary for AAI operations (e.g., name, email, affiliation, ORCID ID, organisation information, user identifiers released by the user's authenticating identity provider, level of assurance information, groups and roles, and resource capabilities);
- Consent and rights: Obtain explicit consent for data processing and inform users of their rights, including data access, correction, and erasure;
- Data sharing and transfers: Define conditions under which personal data may be shared across RIs, ensuring compliance with GDPR's cross-border data transfer rules.
- Retention policy: Specify data retention periods and procedures for secure deletion of data.

1.6.2.2. Acceptable Use Policy (Terms/Conditions of Use)

The Acceptable Use Policy (AUP) will govern user behaviour and outline their responsibilities when accessing ENVRI-Hub NEXT services. To ensure alignment with established best practices, the AUP should be developed using the template and guidance provided by the AARC Policy Development Kit (PDK). Key provisions include:

- Authorized use: Limit access to authorized individuals and ensure users utilize resources solely for legitimate research and collaboration purposes;
- Prohibited activities: Clearly state restrictions, such as data misuse, security breaches, or infringement of intellectual property rights;
- User responsibilities: Require users to maintain the confidentiality of their credentials and report security incidents promptly;
- Compliance with policies: Mandate adherence to licensing agreements, data policies, and ethical standards.

This Acceptable Use Policy and Conditions of Use (“AUP”) defines the rules and conditions that govern your access to, and use (including transmission, processing, and storage of data) of the resources and services (“Services”) as granted by the ENVRI Community, a European network of environmental research infrastructures, projects, and stakeholders for the purpose advancing environmental research.

- You shall only use the Services in a manner consistent with the purposes and limitations described above; you shall show consideration towards other users including by not causing harm to the Services; you have an obligation to collaborate in the resolution of issues arising from your use of the Services;
- You shall only use the Services for lawful purposes and not breach, attempt to breach, nor circumvent administrative or security controls;
- You shall respect intellectual property and confidentiality agreements;
- You shall protect your access credentials (e.g. passwords, private keys or multi-factor tokens); no intentional sharing is permitted;
- You shall keep your registered information correct and up to date;

- You shall promptly report known or suspected security breaches, credential compromise, or misuse to the security contact stated below; and report any compromised credentials to the relevant issuing authorities;
- Reliance on the Services shall only be to the extent specified by any applicable service level agreements listed below. Use without such agreements is at your own risk;
- Your personal data will be processed in accordance with the privacy statements referenced below;
- Your use of the Services may be restricted or suspended, for administrative, operational, or security reasons, without prior notice and without compensation;
- If you violate these rules, you may be liable for the consequences, which may include your account being suspended and a report being made to your home organisation or to law enforcement.

The administrative contact for this AUP is: envri-hub-next-po@mailman.egi.eu

The security contact for this AUP is: envri-hub-next-po@mailman.egi.eu

The privacy statements (e.g. Privacy Notices) are located at:

https://login.staging.envri.eu/federation

1.6.2.3. Data and Metadata Licensing

To ensure harmonised access and reuse of data and metadata:

- Standard licenses: Use of licenses with compatible family such as CC-BY-4.0 for data requiring attribution and CC0-1.0 for open data with minimal restrictions;
- Public access: Encourage open access to data and metadata, aligning with FAIR principles and fostering scientific collaboration;
- Tailored licensing: Allow for multiple licensing models to cater to diverse stakeholders, such as industry partners or citizen scientists.

1.6.2.4. Authentication and Authorisation Policy

Standardised AAI policies will ensure secure and interoperable access to RI resources:

- Federated identity management: Utilise eduGAIN, ORCID, and other federated identity systems for authentication;
- Role-based access control (RBAC): Define roles and permissions based on users' affiliations and requirements;
- Attribute-based access control (ABAC): Supplement RBAC with attribute-based controls for fine-grained access management.

1.6.3. Governance Structure

1.6.3.1. Data Protection and Security

- GDPR compliance: Appoint a GDPR Data Controller within each RI to oversee data privacy and processing practices;
- Security contact: Designate a security officer to address incidents and maintain compliance with security standards like SIRTFI;
- Data breach response: Develop procedures for handling and reporting data breaches in compliance with GDPR.

1.6.3.2. Policy Implementation and Monitoring

- Policy alignment: Require all participating RIs to adopt the standardised policies outlined in this framework;
- Monitoring and audits: Conduct periodic reviews to ensure compliance with the legal framework and update policies as necessary;
- Stakeholder engagement: Facilitate workshops and consultations to address emerging legal, technical, and operational challenges.

1.6.4. Interoperability and Collaboration

- Federated AAI systems: Encourage reciprocal access agreements between RIs, leveraging federated identity providers to ensure seamless access across platforms;
- Data sharing agreements: Develop standardised data sharing agreements to regulate cross-domain data access, ensuring legal interoperability and compliance with EU and international regulations;
- Alignment with AARC blueprint architecture: Adopt the AARC Blueprint Architecture (BPA) for modular and interoperable AAI implementations across RIs.

1.6.5. User Accessibility

1.6.5.1. Inclusivity

- Provide mechanisms for non-EU users and non-federated institutions to access AAI systems through alternative authentication methods;
- Offer local accounts or social login options to ensure universal accessibility.

1.6.5.2. Offline Access

Enable authenticated offline access to critical datasets where necessary, ensuring continuity for specific research needs. Offline access should be limited to authorised users, tied to specific scopes or permissions, and adhere to data security policies. Mechanisms for token expiration, revocation, and auditing should be implemented to maintain security and accountability

1.6.6. Policy Development and Evolution

- Adaptability: Regularly update the legal framework to reflect technological advancements, regulatory changes, and user feedback;
- Open consultation: Engage stakeholders, including RIs, legal experts, and end-users, in the policy development process;
- Harmonisation with EU directives: Align all policies with European and international standards to foster interoperability;

2. Conclusion

The deliverable "Legal Framework for Cross-Domain AAI Systems" within the ENVRI-Hub NEXT project provides a comprehensive analysis and recommendations for standardising data and metadata policies, aligning them with the requirements of a cross-domain AAI system. By addressing diverse licensing practices, legal compliance, and technical interoperability challenges, this deliverable lays the foundation for a unified and secure ecosystem that supports interdisciplinary research and open science.

The analysis highlights the widespread adoption of Creative Commons licenses (CC-BY-4.0 and CC0-1.0), reflecting a strong commitment to FAIR data principles. However, variations in licensing practices and legal constraints underscore the need for harmonized policies. Recommendations include adopting standard licensing frameworks, enhancing metadata discoverability through machine-readable formats, and fostering interoperability with platforms like the European Open Science Cloud (EOSC). The deliverable also emphasizes the importance of GDPR compliance and robust privacy measures to build trust and ensure ethical data sharing.

The implementation of a cross-domain AAI system supported by federated identity protocols and secure access controls is pivotal for enabling seamless collaboration among RIs. The integration of such systems with legal frameworks ensures equitable access, particularly for diverse user groups, including non-EU and non-federated institutions.

In conclusion, this deliverable provides a pathway for ENVRI-Hub NEXT to advance its mission of delivering secure, interoperable, and FAIR-compliant data services. It serves as a critical step toward fostering global collaboration, enhancing data accessibility, and supporting the evolving needs of interdisciplinary research communities.

To ensure the practical adoption of these recommendations, the ENVRI-Hub NEXT project will implement and pilot the recommended policy framework, including documents such as the Privacy Policy and Acceptable Use Policy, specifically designed for the cross-domain AAI. These policies will be tested in collaboration with participating Research Infrastructures (RIs) to validate their applicability and identify areas for refinement.

Additionally, the project will provide an AARC BPA-compatible cross-domain AAI solution, catering to users and services of RIs that do not have their own AAI service. This solution will facilitate secure and interoperable access to resources, supporting seamless collaboration across RIs. Priority will be given to collaboration with EOSC and other research infrastructures to test interoperability and further refine the cross-domain AAI solution.