National Czech Programme | eosc

National Repository Platform Project Report

Document Information

Title	Initial AAI architecture for sensitive data in NRP
Project Registration Number	CZ.02.01.01/00/23_014/0008787
Key Activity Name and Number	4.3 Data access control
Outcome Name and Number	15 – Initial identity management integration for NRP
Responsible Person	Peter Lényi
Document License	

Version History

Version	Date	Comment
1.0	31. 12. 2025	Initial version.

Co-funded by the European Union

MINISTRY OF EDUCATION, YOUTH AND SPORTS

National Repository Platform for Research Data (NRP)
CESNET, zájmové sdružení právnických osob
Mailing Address: Generála Píky 430/26, 160 00 Prague
info@eosc.cz; www.eosc.cz

Authors

Institution	Name & Surname	ORCID	Email
MUNI	Peter Lényi	0009-0009-7840-5458	lenyi@ics.muni.cz
CESNET	Martin Kuba	0000-0002-0305-7446	makub@cesnet.cz
CESNET	Pavel Vyskočil	0000-0001-8376-2761	vyskocil@cesnet.cz
CESNET	Pavel Zlámal	0009-0006-1628-0440	zlamal@cesnet.cz

Contributors

Institution	Name & Surname	ORCID	Email
CESNET	Zdenka Dudová	0000-0002-7615-1396	dudova@cesnet.cz

List of Abbreviations

AAI	Authentication and Authorisation Infrastructure
AAL	Authentication Assurance Level
ABAC	Attribute-Based Access Control
ACM	Access Control Mechanism
API	Application Programming Interface
DAC	Data Access Committee
DAP	Data Access Policy
DAR	Data Access Request
DLA	Deposit License Agreement
DMP	Data Management Plan
DPA	Data Processing Agreement
DS	Discovery Service
DSP	Data Sharing Policy
DTA	Data Transfer Agreement
DUA	Data Use Agreement
EHDS	European Health Data Space
EOSC	European Open Science Cloud
FAIR	Findable, Accessible, Interoperable, Reusable
FAL	Federation Assurance Level
FOSS	Free and Open-Source Software
GBAC	Group-Based Access Control
GDPR	General Data Protection Regulation
GUI	Graphical User Interface
IAL	Identity Assurance Level
IAM	Identity and Access Management
IdP	Identity Provider
ISMS	Information Security Management System
K8s	Kubernetes

MFA	Multi-Factor Authentication
NRP	National Repository Platform for Research Data
OIDC	OpenID Connect
OP	OpenID Provider
PAP	Policy Administration Point
PBAC	Policy-Based Access Control
PDP	Policy Decision Point
PEP	Policy Enforcement Point
PI	Principal Investigator
PID	Permanent Identifier
PII	Personally Identifiable Information
PIP	Policy Information Point
RBAC	Role-Based Access Control
RP	Relying party
R/S	Repositories and Services
S3	Simple Storage Service
SAML	Security Assertion Markup Language
SD	Sensitive Data
SLA	Service-Level Agreement
SLO	Single Logout
SP	Service Provider
SPE	Secure Processing Environment
SSO	Single Sign-On
SWS	Scientific Workflow System
ToS	Terms of Service
TRE	Trusted Research Environment
TRL	Technology Readiness Level
VO	Virtual Organisation
VRE	Virtual Research Environment
WAYF	Where Are You From

WFMS	Workflow Management System

Table of Contents

Executive Summary

The National Repository Platform for Research Data is a key component of the National Data Infrastructure aiming to consolidate the fragmented research data environment by providing scientists with tools for data management in line with FAIR principles, reliable repositories with sufficient storage capacity, and technical, methodological, and educational support.

The Authentication and Authorisation Infrastructure is a core service of the Platform enabling single sign-on and facilitating user access control across service providers.

This document supersedes the *Initial AAI architecture for NRP*, restating the basic principles, reflecting the changing landscape and guidelines, and extending the recommendations for access control to solve the requirements of storing and processing sensitive data in NRP. It will be complemented by technical guidelines containing interface descriptions and default configurations. Target audience are the administrators and developers of repository systems and related services in NRP; technical, methodological, and educational support personnel in NRP will find it informative.

It delivers an access control blueprint for service providers, building on the AARC community guidelines, and combining role- and attribute-based access control approaches to utilize their strong points and offset the weak ones. This blueprint describes three components forming a centralised AAI: proxy identity provider, identity and access management, and policy engine. ProxyIdP integrates service providers using the OIDC protocol to enable single sign-on and facilitate system-to-system access using token exchange. IAM holds user attributes and their group membership to govern and synchronise user roles across services. And policy engine evaluates applicable machine-readable authorisation policies to allow or deny access for a combination of user, action, resource and environment attributes.

The main benefit is increased automation of user workflows and access control, resulting in fewer user errors and associated support costs. Other benefits are that service providers can delegate parts of access control implementation and operation to AAI further decreasing their costs, and we shift AAI focus towards identities and attributes pre-aligning it with upcoming ecosystem of digital identity wallets and federating initiatives like EOSC or Data Spaces.

The architecture will evolve in the upcoming years as we continue to explore fine-grained access control, integrate with international environment, and learn from the integration of pilot repositories and related services.

1 Introduction

This technical document is a deliverable no. 15 of the project[1] *National Repository Platform for Research Data*, delivered as a of the Key Activity 4.3 *Data access control* at the end of the project's second year. It follows the implementation of pilot repositories and supersedes the previous[2] *Initial AAI architecture for NRP*. It is also a prerequisite for the upcoming implementation of sensitive data (SD) repositories. Its target audience are the administrators and developers of repository systems and related services (R/S) in NRP; technical, methodological, and educational support personnel in NRP will find it informative.

The main aim is to describe an architecture of authentication and authorisation infrastructure (AAI), which reflects recent changes in the landscape, guidelines, and best practices, as well as extends the current solution with elements of attribute-based access control to reflect the needs of FAIR[3] sensitive data storage and processing use cases in NRP.

The document begins with this introduction and description of the methodology in the second chapter. The third chapter lists collected use cases and the fourth lists derived requirements. The solution architecture and core concepts are explained in the fifth chapter. Chapter six discusses the architecture, its limitations. and future work. Summary can be found in chapter seven. The document ends with a list of references; the appendices contain an exhaustive glossary and an illustrative example.

[1] https://www.eosc.cz/en/projects/national-repository-platform-for-research-data-nrp/

[2] https://www.eosc.cz/en/projects/outcomes-of-the-eosc-cz-initiative#AAI_NRP

[3] https://doi.org/10.1038/sdata.2016.18

2 Methodology

Since the publishing of *Initial AAI architecture for NRP*, we have been learning and gathering information through various collaborations. First, we have been integrating pilot repositories and related services in NRP and assisting them with their access control setup. Second, we have been talking to people working with sensitive data such as *WG Sensitive Data*[4], ESFRI research infrastructures *BBMRI-ERIC*[5] and *ELIXIR*[6], and members of the *EOSC-ENTRUST*[7] project about management of sensitive data in academic research. Finally, we have been discussing with KA2.2: *Invenio development & operation* and KA5.4: *Computational workflow integration* about access control automation for workflow-repository interactions.

We used all this information to produce several model use cases and to derive table-stake requirements for the access control mechanism in NRP. We discussed these use cases and requirements in a series of AAI specialist focus groups and devised a concept architecture, including ideas and criteria for its future implementation. We chose to base the concept on AARC BPA[8] "designed to implement access management solutions for international research collaborations" and used by many initiatives and institutions[9]. We must note that focus group members' experience with operating production AAI systems had major influence on the concept.

In the end, we worked the concept out into a proper architecture presented in this technical document, which we shall validate by implementing a pilot solution and deploying it for use by sensitive data repositories coming to NRP with the OS II[10] project to the extent they can.

---

[4] https://www.eosc.cz/en/working-groups/sensitive-data
[5] https://www.bbmri-eric.eu/
[6] https://elixir-europe.org/
[7] https://eosc-entrust.eu/
[8] https://aarc-community.org/architecture/
[9] https://wiki.geant.org/spaces/AARC/pages/738885722/WP5+Compendium+Recommendations
[10] https://www.eosc.cz/en/projects/open-science-ii

3 Use Cases

The model use cases in this chapter describe various uses of NRP with distinct requirements for access control. Based on the information provided by people working with sensitive data, we selected a representative set of use cases presented here.

The FAIR data use cases (FD-) have already been addressed in the previous version of the architecture but are restated here for a full overview. The sensitive data use cases (SD-) are the core focus of this document, complemented by data use (DU-) and data collection (DC-) use cases for both sensitive and non-sensitive data that reveal the real value of integrating with AAI for access control. Finally, there is a use case highlighting the role of administrators in NRP (SU-).

FD-1 Restricted FAIR dataset deposition

A data producer designated by a data rights holder deposits a FAIR dataset into a repository which does not allow unauthorised users to deposit data.

FD-2 Auditable FAIR dataset access

A data user accesses someone else's FAIR dataset in a repository which does not allow data to be accessed anonymously.

FD-3 FAIR dataset sharing embargo

A data producer designated by a data rights holder deposits a FAIR dataset into a repository. At the same time, the data producent sets up an embargo on access to the dataset for a set amount of time. This embargo does not apply to the data producer and specific authorised collaborators whom the data producer designates.

SD-1 Sensitive dataset deposition

A data producer designated by a data rights holder deposits a FAIR sensitive dataset into a sensitive data repository as the (authoritative) dissemination copy. The data producer may temporarily keep the dataset locked until DAC or the institutional rules permit its release for secondary use.

SD-2 Sensitive dataset sharing

A data producer designated by a data rights holder, or the data access committee (DAC) – depending on the repository's data sharing policy – sets up access to the dataset for specific authorised data users based on a data access request they made. A repository administrator or data curator may be involved in the technical process of dataset release.

DU-1 Sensitive dataset processing in VRE

A researcher, who was previously granted the permission to access a sensitive dataset per data use agreement (DUA), has it moved to a trusted research environment (TRE), inside of

which the researcher processes the dataset interactively with a virtual research environment (VRE). If DUA allows it, the researcher may invite specific authorised collaborators to VRE.

A data producer might manipulate the dataset as a pre-processing step before its deposition, while a data user would do this as part of their research work.

DU-2 Sensitive dataset processing with SWS

A researcher, who was previously granted the permission to run a specific data analysis of a specific sensitive dataset per DUA, has a scientific workflow system (SWS) move the dataset to a TRE and run the workflow non-interactively inside of the TRE. The researcher can never access the dataset directly, only the non-sensitive (e.g. anonymised) workflow output after an output control procedure. If DUA permits it, the researcher may share non-sensitive output with specific authorised collaborators within or outside of TRE.

The workflow output might also be sensitive. In such case, the researcher may not access it but may process it in a follow-up workflow or have it moved to a sensitive data repository.

DC-1 Automated FAIR dataset collection

A data producer, who produced a dataset with an automated instrument, has it automatically transferred into a repository which does not allow unauthorised users to deposit data. There is an assumption this was previously permitted by the data rights holder.

SU-1 Incident handling

AAI administrator responds to an incident and audits authentication and authorisation activity to determine the root cause and resolve the issue, as well as to notify the impacted parties.

4 Requirements

The requirements in this chapter are derived from the model use cases. They are split in two broad categories: functional and non-functional; and cover all representative cases that are essential to cover the solutions presented in the next chapter.

All requirements apply to sensitive and non-sensitive FAIR data alike, unless the requirement description explicitly states the sensitivity.

Note the key words "must", "must not", "required", "shall", "shall not", "should", "should not", "recommended", "may", and "optional" are to be interpreted as described in RFC2119[11]*.*

4.1 Functional

ID	Description
FR-1	Data user *may* authenticate to access open (publicly available) datasets.
FR-2	Data user, producer, etc. *must* authenticate to perform controlled operations.
FR-3	Data user, producer, etc. *should* be able to authenticate with an academic identity.
FR-4	Data user, producer, etc. *should* be able to authenticate with a social identity.
FR-5	Data user, producer, etc. *must* be able to use single sign-on capability.
FR-6	Data user, producer, etc. *should* only use one digital identity.
FR-7	Data user, producer, etc. *must* be able to link their digital identities.
FR-8	System administrator must authenticate to perform admin-level operations. It is recommended an independent authentication mechanism is used when system administrator logs into sensitive data repositories and services.
FR-9	A repository *must* control which users can access or manage which resources based on decisions made by relevant actors (e.g. data producers or DAC).
FR-10	A repository *must* allow data producers to access and manage own resources.
FR-11	A repository *must* allow relevant actors (e.g. data producers or DAC) to control access to the resources they are responsible for, i.e. allow one or a group of specific users to access a dataset.
FR-12	A repository must allow relevant actors (e.g. data producers or users) to access (e.g. view, download, move) the resources they have permission for.
FR-13	Data user, producer, etc. *should* authenticate with at least two factors to perform controlled operations with resources in repositories.

[11] https://doi.org/10.5281/zenodo.3672785

FR-14	Data user, producer, etc. *must* authenticate with at least two factors to perform controlled operations with resources in sensitive data repositories.
FR-15	Data user, producer, etc. *may* have their identity strongly vetted to perform controlled operations with resources in repositories.
FR-16	Data producer *should* have their identity strongly vetted to perform controlled operations with own resources in sensitive data repositories.
FR-17	Data user, curator, etc. *must* have their identity strongly vetted to perform controlled operations with someone else's data in sensitive data repositories.
FR-18	A repository *must* support embargo control for sensitive & non-sensitive datasets and *must* support combining it with other access control mechanisms.
FR-19	Data user, curator etc. *should* be able to transfer sensitive datasets to and from a TRE using a cross-system access control mechanism.
FR-20	It is *recommended* SWS or VRE (inside or outside of a TRE) employs a cross-system access control mechanism to enforce applicable policies and agreements (e.g. DUA) when a data user processes data (interactively or non-interactively).
FR-21	An AAI administrator *must* monitor the AAI for anomalies.
FR-22	An administrator *must* audit the AAI during incident investigation.

Note we list the functional requirements above in no particular order. Also note, they are all restricted to the scope of NRP.

4.2 Non-functional

ID	Name	Description
		Justification
NFR-1	Security	*must* minimize the number of known system vulnerabilities
		maintaining trust in the access control mechanism
NFR-2	Reliability	*must* maximize the mean time between system failures
		maintaining trust in the access control mechanism
NFR-3	Availability	*must* minimize planned and unplanned system down time
		maintaining trust in the access control mechanism
NFR-4	Compliance	*must* maximize system conformity to applicable rules
		satisfying legal and other applicable obligations
NFR-5	Interoperability	*must* maximize system compatibility with other systems
		interacting with other repository platforms or external systems
NFR-6	Maintainability	*must* minimize the mean time to correct system failures
		maintaining trust in the access control mechanism
NFR-7	Extensibility	*must* minimize the mean time to add system functionality
		adapting to NRP development and user feedback

Note we list the non-functional requirements above in the order of descending importance. Also note, these system qualities are scalar, not binary, so e.g. resolving a major compliance issue would still take precedence over resolving a negligible security issue.

5 Architecture

The architecture described in this chapter is based on the model use cases and requirements from the previous chapters, and common AAI standards. It is described in four functionality-oriented parts: infrastructure, authentication, authorisation, and auditing.

We evaluate the architecture, compare it to the previous version, and outline its limitations and future work needed in the next chapter. Technical details such as interface descriptions and default configurations will be published in complementary technical guidelines.

5.1 Infrastructure

Starting point for the solution is *AARC Blueprint Architecture*[12], state of the art in European academic research, complemented by AARC guidelines[13] which help implement and operate AAI in more effective and interoperable ways, and chosen REFEDS standards[14] that address the needs of academic identity federations worldwide.

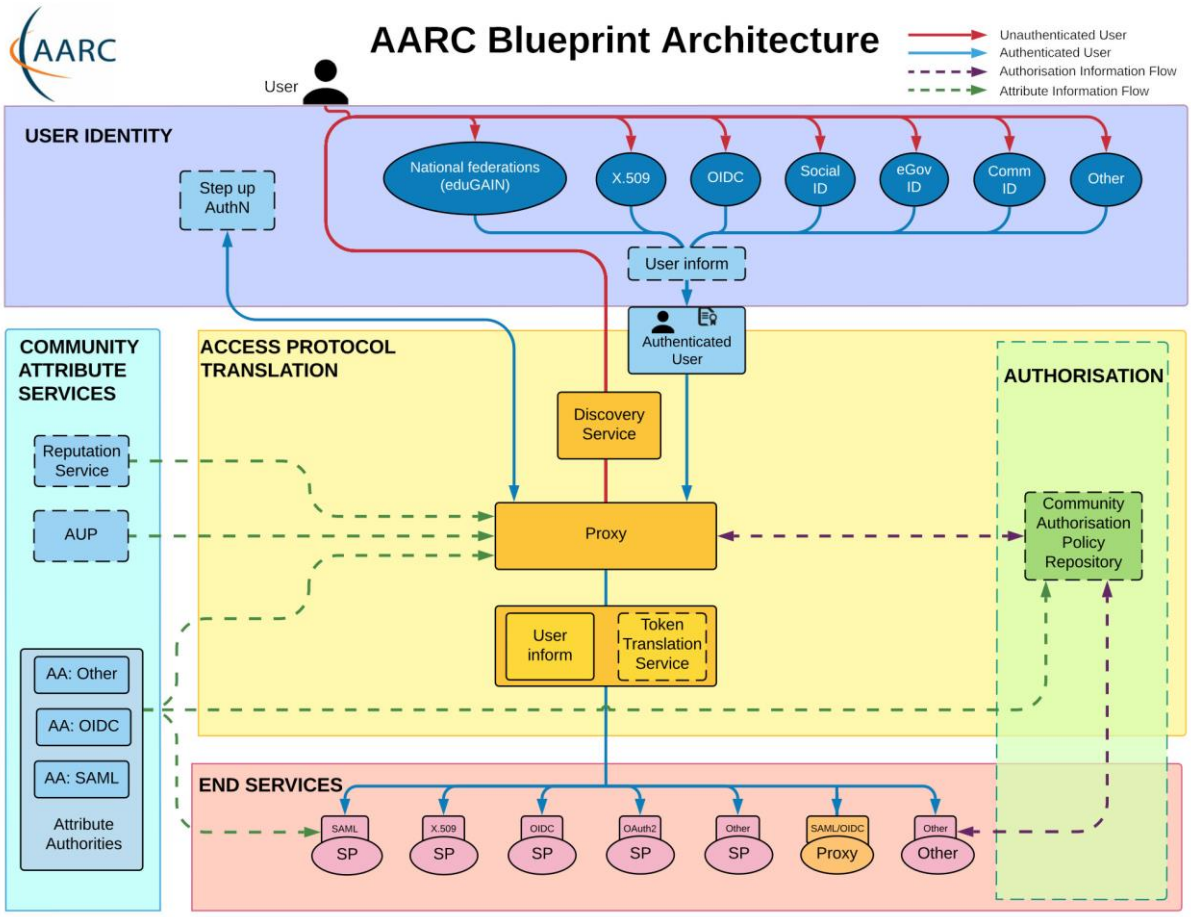

Figure 1: Component layers of the AARC Blueprint Architecture (AARC-BPA-2019) [15]

[12] https://doi.org/10.5281/zenodo.3672785
[13] https://aarc-community.org/guidelines/
[14] https://wiki.refeds.org/
[15] https://aarc-community.org/architecture/

In other words, our solution is an authentication and authorisation infrastructure (AAI) which enables federated identity and access management (IAM) for NRP. This AAI integrates with academic[16], national[17] and social identity providers (IdP) on one side, and with repositories and related services (R/S) acting as service providers (SP) on the other side which enables basic authentication and authorisation workflows.

For the implementation, we choose Perun AAI[18]; more specifically, its e-INFRA CZ AAI[19] instance operated by CESNET[20] for the national research e-infrastructure e-INFRA CZ[21] and to fulfil the needs of implementing EOSC in Czech Republic[22] and European Union[23].

Additional capabilities are delivered through add-on components: discovery service (WAYF/DS), identity consolidator, multi-factor authenticator (MFA), notice and consent presentation, user registrar, policy engine, and more.

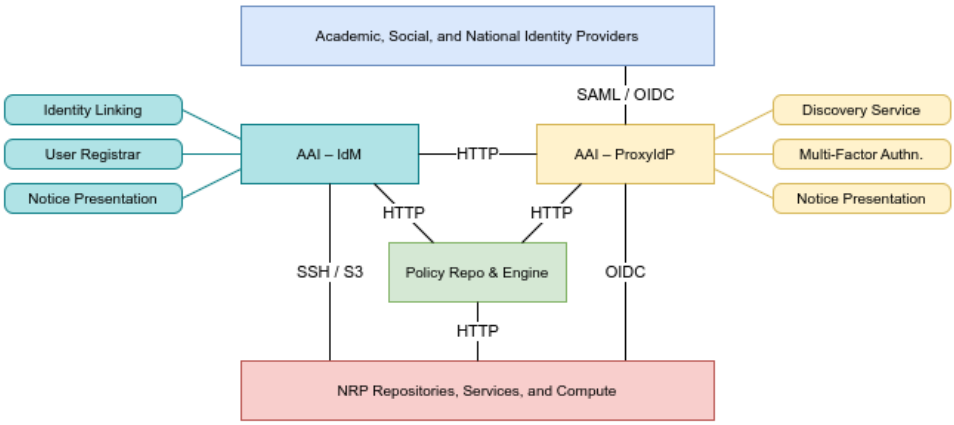

Figure 2: AAI components and their connections to IdPs and SPs; colour-coded to match AARC BPA above

AAI and its components are connected to logging, monitoring, and backup systems, and they comply with applicable regulations, guidelines, policies, and best practices. CESNET, which operates the solution, implements ISMS and ISO/IEC 27k standards on information security, ensures CSIRT oversight, and specialized L0-L3 support to end users and service providers.

Note we require repositories and related services in NRP to integrate with our AAI, but we do not forbid them to also *integrate with additional AAIs or implement their own. However, in such cases the services are fully responsible for the compliance of used alternatives with the requirements in chapter 4.*

5.2 Authentication

Authentication begins in a R/S which requests authentication from AAI using OIDC protocol – usually OIDC authorisation code flow with PKCE in some cases. AAI's proxy delegates the

[16] https://edugain.org/
[17] https://www.identita.gov.cz/
[18] https://perun-aai.org/
[19] https://perun.e-infra.cz/
[20] https://www.cesnet.cz/
[21] https://www.e-infra.cz/
[22] https://www.eosc.cz/
[23] https://eosc.eu/

request to an IdP user chooses (hinting[24] is supported) and forwards normalised response back to the R/S while establishing a single sign-on (SSO) session.

Additionally, AAI registers new users, presents applicable notices, gathers required consents, ensures fallback MFA, and informs R/S about effective identity and authentication assurance levels (IAL, AAL), and freshness of attributes.

5.3 Authorisation

The access control mechanism (ACM) varies based on the specific R/S requirements and capabilities. Most use cases are still solved using the role-based access control approach, **but the more powerful attribute-based access control approach is now available, too**.

5.3.1 RBAC

Role-based access control[25] is an access control variant in which the authorisation policy is made of *subject–role–operation–object* rules and an access decision is based on subject's roles. This approach is common thanks to its flexibility and simplicity.

NRP employs AAI-managed role-based ACM for access to R/S, to infrastructure if provided by NRP, and to R/S tenants and roles during cross-system workflows. NRP further employs AAI-managed user access pre-/de-/provisioning for R/S with API access. Specifically:

- Users are authorised to access R/S in NRP based on respective virtual organisation (VO) membership managed in AAI – representing the “user” role. AAI checks for this role during user login before redirecting the user to the R/S.
- Users are authorised to access elements of the infrastructure (S3 or K8s) provided by NRP based on respective groups membership managed in AAI – representing “infra admin” role. AAI provisions this data to the infrastructure via a JSON connector.
- Users can use a service (like a VRE or SWS) to perform operations (like file transfer) in a repository on their behalf. As a prerequisite, a user must be authorised themself based on group membership managed in AAI. Then the service can obtain an access token from AAI based in exchange for the user’s access token and use it for machine-to-machine access to the repository.

5.3.2 ABAC

Attribute-based access control[26] is an access control variant in which the authorisation policy is made of *subject–operation–object–environment* rules and an access decision is based on entities' attributes. ABAC is rare due to its complexity but is more powerful than RBAC.

Technically, an attribute-based access control mechanism consists of four functional points: a policy administration point manages authorisation policies, a policy information point provides relevant attributes, a policy decision point makes access control decisions, and a policy enforcement point enforces the decisions. A context handler orchestrates the access control process. [27]

---

[24] https://doi.org/10.5281/zenodo.4596667
[25] https://en.wikipedia.org/wiki/Role-based_access_control
[26] https://doi.org/10.6028/NIST.SP.800-162
[27] https://doi.org/10.6028/NIST.SP.800-162

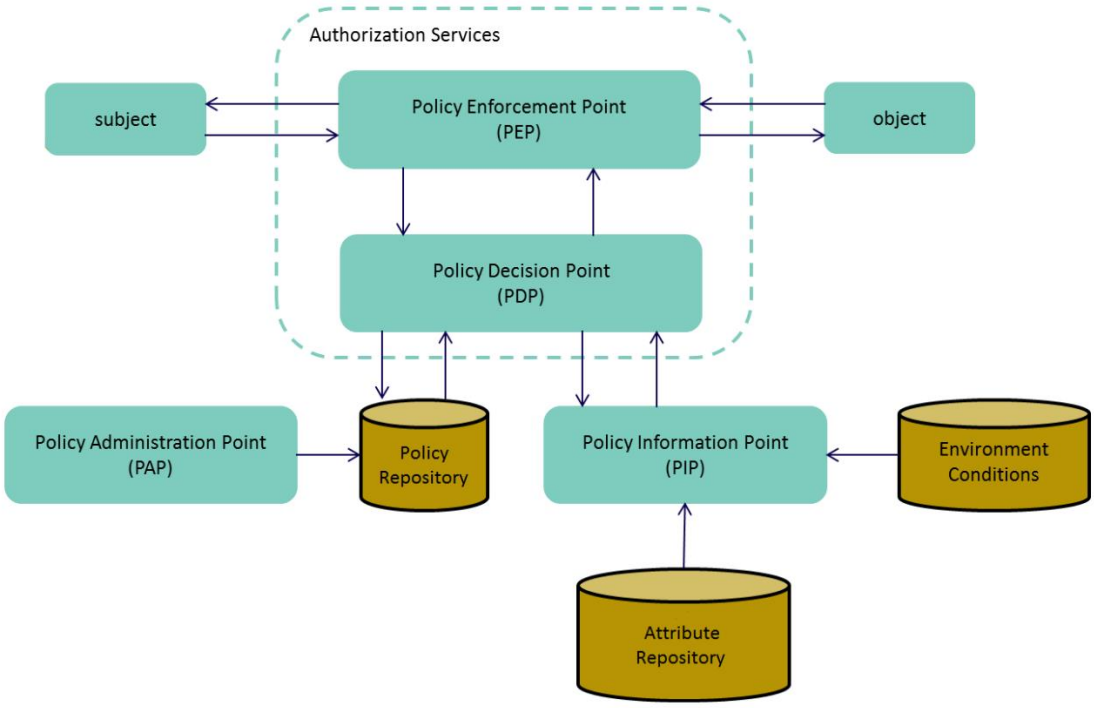

Figure 3: An Example of ACM Functional Points [28]

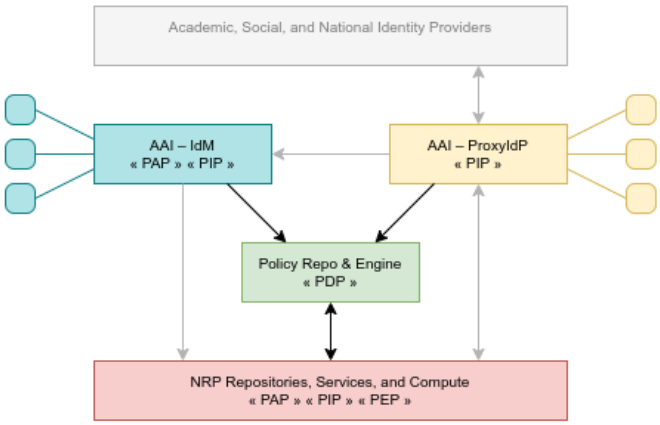

Figure 4: Mapping the ACM functional points onto NRP. AAI components IdM and ProxyIdP act as PIP providing user attributes. ProxyIdP additionally provides some environmental attributes (e.g. session info). IdM additionally acts as PAP, providing "general" access policy (e.g. always deny banned users). Repositories act as PIP providing dataset and action attributes, as PAP providing policies (e.g. license), and as PEP enforcing the access decision. Other services might be involved; for example, a tool for submitting and approving DARs would act as another PAP providing agreements (i.e. DUA).

NRP employs AAI-assisted attribute-based ACM for access to repository datasets and other similar medium-grained resources, and for automated enforcement of machine-readable data -related documents, policies, and agreements. For example:

- Users are authorised to access a dataset based on predefined data policies. A policy engine loads the data sharing and access policies, collects user, dataset, operation, and environment attributes from AAI, repository, DAR, and other sources, and makes

[28] ibid.

an access decision. Policy is written in ODRL or Rego language, and attributes are supplied in JSON files.

- Users are authorised to perform data analysis following the rules set by DUA. A policy engine loads the DUA, collects user, dataset, operation, and environment attributes from AAI, repository, VRE/SWS, and other sources, and makes an access decision. Policies are written in ODRL or Rego language, and attributes are supplied in JSON files.

Note we are leaning towards deploying policy engines close to R/S for two reasons: trust and efficiency. If the component is running as part of R/S, there is less fear of "giving up" access decisions to AAI, and it is best practice to have PDP (policy engine) close to PEP (R/S). But having a central "catch-all" policy engine deployed as part of AAI to save R/S effort remains an open possibility, too, pending pilot feedback.

5.4 Auditing

Auditing of authentication and authorisation activity in NRP, triggered by compliance checks or anomaly detection, is distributed across systems, and followed by consolidation to get the full picture. It requires cooperation of respective system administrators.

6 Discussion

This chapter has four sections: fulfilment of requirements by the architecture, changes since the previous version, outstanding limitations and their mitigation, and future work to be done.

6.1 Evaluation

Twenty-two functional and seven non-functional requirements have been initially identified:

- FR-1 to FR-7 are solved with a centralized AAI that integrates IdPs and R/S in NRP, lets any NRP user register, and assigns a unique, permanent, and non-re-assignable identifier to each of them.
- FR-6 and FR-7 are solved with AAI's identity consolidator, the use of which is optional to protect privacy, but is recommended to avoid issues with identity switching.
- FR-2, FR-5, and FR-8 are solved by integrating all R/S with AAI.
- FR-3 and FR-4 are solved by integrating AAI with eduGAIN and common social IdPs, and by integrating all R/S with AAI.
- FR-9 to FR-12 are primarily the responsibility of R/S; AAI contributes to automation and economy of the ACM by letting R/S delegate authentication, user management, and user role or attribute management to the AAI. A choice between role-based and attribute-based ACM is offered to find the right balance between automation and user experience.
- FR-13 and FR-14 are primarily the responsibility of IdPs to perform MFA and R/S to check AAL when making an access control decision. AAI contributes by collecting the requests for minimum AAL, providing fallback MFA if necessary, and informing R/S about the effective AAL.
- FR-15 to FR-17 are primarily the responsibility of IdPs to perform identity vetting and R/S to check IAL when making an access control decision. AAI assists by collecting the requests for minimum IAL, facilitating identity vetting by involving IdPs capable of it, and informing R/S about the effective IAL.
- FR-18 is solely the responsibility of data repositories. However, AAI might contribute if a repository implements an attribute-based ACM and relies on the AAI to implement any functional point.
- FR-19 and FR-20 are shared responsibility of R/S and AAI. The former implements the ACM; the latter guarantees technical and semantic interoperability and facilitates the communication between the R/S.
- FR-21 and FR-22 are solved by integrating AAI with logging, monitoring, and backup facilities.
- Non-functional requirements are addressed by using FOSS, complying with relevant regulations, guidelines, policies, and best practices, implementing ISMS and ISO/IEC 27k standards, having CSIRT oversight, and integrating AAI with logging, monitoring, and backup facilities.

In conclusion, the architecture proposed in this document fulfils all requirements, and thus all use cases, too.

6.2 Comparison

The previous version of AAI architecture for NRP[29] set out to solve two use cases: enabling SSO across NRP and restricting access to SPs (that were like FD-1 and FD-2 in chapter 3). It provided a glossary, defined core technical and procedural principles, and described how the authentication, authorisation and auxiliary functions are to be implemented. The solution it introduced was an AAI whose architecture, implementation, and interfaces match the one we describe in sections 5.1, 5.2, 5.3.1 and 5.4 of this document.

This document introduces several new model use cases and numerous derived requirements connected to sensitive data storage and processing, as well as scientific workflows to NRP. The solution to these is new AAI capabilities that we describe in sections 5.3.1 and 5.3.2 of this document, i.e. ABAC and token exchange.

The new architecture does not change or remove any requirements of the old architecture.

6.3 Limitations

There are several known sources of bias which have influenced the solution architecture:

- *AARC Blueprint Architecture*[30] is state of the art in European academic research, so alternatives were not explored. We believe this is fine since the BPA is also used by the EOSC AAI Architecture[31] of the EOSC Federation[32].
- New *AARC Blueprint Architecture 2025* will be published in early 2026, but this will be backwards compatible. We expect it will have no impact on this architecture.
- e-INFRA CZ AAI[33] was selected as a pre-existing solution supporting implementation of EOSC in the Czech Republic; it meets all requirements and has TRL 9.
- Use case collection involved a limited number of people working with sensitive data, only model use cases were ever discussed, and national use cases were prioritised. We find this acceptable for an initial architecture that will continue to evolve.
- Focus groups for concept architecture engaged just a few AAI specialists. Again, we deem this acceptable for an initial architecture.

There is also a significant gap in functionality as there is currently no generally working identity vetting solution. Academic SPs or Proxies cannot be legally connected to eID in the Czech Republic and known commercial solutions are too expensive. Unless the Czech legal landscape changes, the only efficient option may be to integrate with the identity layer of the EOSC AAI Federation.

6.4 Future Work

This document describes the AAI architecture and references available standards. However, an instruction manual and user documentation must be produced so that the administrators and developers of R/S in NRP can integrate and configure their systems, and the technical,

---

[29] https://www.eosc.cz/en/projects/outcomes-of-the-eosc-cz-initiative#AAI_NRP
[30] https://aarc-community.org/architecture/
[31] https://doi.org/10.5281/zenodo.15388270
[32] https://doi.org/10.5281/zenodo.14999577
[33] https://perun.e-infra.cz/

methodological and educational support personnel in NRP have the necessary information to aid repository administrators and end users with (not just) SD access control use cases.

The architecture also needs to be validated using real-life use cases; an opportunity will arise with the introduction of SD repositories coming with the OS II[34] project in Czech Republic.

The landscape surrounding both the AAI and SD keeps evolving. The EHDS regulation[35] and EOSC Federation[36] interoperability guidelines will have the greatest influence on our work in NRP. Knowledge and experience transfer from international projects focusing on SD will continue as well. EUDIW[37] might trigger a paradigm shift, but the current consensus is that actual adoption is still years away, likely beyond the lifetime of the NRP project.

Finally, there are three more planned deliverables of the project key activity 4.3: *Data access control*, no. 16-18. These will further explore fine-grained access control and AAI integration with the international environment.

---

[34] https://www.eosc.cz/en/projects/open-science-ii

[35] https://eur-lex.europa.eu/eli/reg/2025/327/oj/eng

[36] https://eosc.eu/building-the-eosc-federation

[37] https://ec.europa.eu/digital-building-blocks/sites/spaces/EUDIGITALIDENTITYWALLET/pages/694487738/EU+Digital+Identity+Wallet+Home

7 Conclusion

The primary goal of this project deliverable was to describe an AAI architecture that fulfils NRP requirements ahead of project's third year and reflects relevant changes outside of the project's scope.

Most of the new requirements relate to sensitive data storage and processing in NRP, but the needs of computational workflows were considered as well. Both categories increase the complexity significantly; therefore, we introduce attribute-based access control as one of AAI's new capabilities, on top of the previous version of the architecture, offering new options to repositories and related services in NRP for implementing data access control. Moreover, enforcing increased identity and authentication assurance levels is now a requirement rather than recommendation.

These new requirements also impact repository systems, repositories, and related services, which must implement new capabilities and integrate with new interfaces if they interact with sensitive data or computational workflows. Necessary implementation details like interface descriptions and default configurations will be in accompanying technical guidelines.

Finally, more work is needed to deal with truly fine-grained access control requirements and to integrate with international environment, which we will deal with in the future version of AAI architecture for NRP.

References

- LÉNYI, Peter, Pavel VYSKOČIL, and Lukáš VOJÁČEK. *Iniciální architektura AAI pro NRP* [online]. 2025 [Accessed 31. 12. 2025]. Available at: https://www.eosc.cz/media/3854788/inicialni-architektura-aai-pro-nrp.pdf (PDF)
- WILKINSON, Mark D. et al. *The FAIR Guiding Principles for scientific data management and stewardship* [online]. Sci Data 3, 160018, 15. 3. 2016. Available at: https://doi.org/10.1038/sdata.2016.18
- AARC COMMUNITY and APPINT. *AARC Blueprint Architecture 2019 (AARC-G045)* [online]. AARC-G045. 6. 11. 2019. Available at: https://doi.org/10.5281/zenodo.3672785
- BRADNER, Scott. *Key words for use in RFCs to Indicate Requirement Levels* [online]. RFC 2119, March 1997. Available at: https://doi.org/10.17487/RFC2119
- AARC COMMUNITY and APPINT. *A specification for IdP hinting* [online]. AARC-G061, 10. 3. 2021. Available at: https://doi.org/10.5281/zenodo.4596667
- HU, Vincent C. et al. *Guide to Attribute Based Access Control (ABAC) Definition and Considerations* [online]. NIST, January 2014. Available at: https://doi.org/10.6028/NIST.SP.800-162
- KANELLOPOULOS, Christos et al. *EOSC AAI Architecture 2025*. 12. 5. 2025. Available at: https://doi.org/10.5281/zenodo.15388270
- EOSC ASSOCIATION. *EOSC Federation Handbook* [online]. Version 1. 27. 3. 2025. Available at: https://doi.org/10.5281/zenodo.14999577
- *Regulation (EU) 2025/327 of the European Parliament and of the Council on the European Health Data Space and amending Directive 2011/24/EU and Regulation (EU) 2024/2847* [online]. OJ L, 2025/327, 5. 3. 2025. Available at: https://eur-lex.europa.eu/eli/reg/2025/327/oj/eng
- MATYSKA, L. et al. *Conditions for Creating New and Modifying Existing Domain Repositories in the National Repository Platform* [online]. Version 3.3. 12/25. Available at: https://www.eosc.cz/en/projects/outcomes-of-the-eosc-cz-initiative#repositories_conditions
- *Regulation (EU) No 910/2014 of the European Parliament and of the Council of 23 July 2014 on electronic identification and trust services for electronic transactions in the internal market and repealing Directive 1999/93/EC* [online]. OJ L 257, 28. 8. 2014, pp. 73–114. Available at: https://eur-lex.europa.eu/eli/reg/2014/910/oj/eng
- *Regulation (EU) 2024/1183 of the European Parliament and of the Council of 11 April 2024 amending Regulation (EU) No 910/2014 as regards establishing the European Digital Identity Framework* [online]. OJ L, 2024/1183, 30. 4. 2024. Available at: https://eur-lex.europa.eu/eli/reg/2024/1183/oj/eng
- AARC COMMUNITY. *Account linking and LoA elevation use cases and common practices for international research collaboration* [online]. AARC-G009, 13. 6. 2017 [Accessed 31. 12. 2025]. Available at: https://aarc-community.org/wp-content/uploads/2017/03/AARC-JRA1.4H.pdf (PDF)
- TEMOSHOK, David et al. *Digital Identity Guidelines* [online]. NIST, July 2025. Available at: https://doi.org/10.6028/NIST.SP.800-63-4 (PDF)
- *Commission Implementing Regulation (EU) 2015/1502 on setting out minimum technical specifications and procedures for assurance levels for electronic identification means pursuant to Article 8(3) of Regulation (EU) No 910/2014 of the European Parliament and of the Council on electronic identification and trust services for electronic transactions in the internal market* [online]. OJ L 235, 9.9.2015, pp. 7–20. Available at: https://eur-lex.europa.eu/eli/reg_impl/2015/1502/oj/eng
- REFEDS. *REFEDS Assurance Framework* [online]. 5. 12. 2023. Available at: https://doi.org/10.5281/zenodo.10277233

- GRASSI, Paul A. et al. *Attribute Metadata* [online]. NIST, January 2018. Available at: https://doi.org/10.6028/NIST.IR.8112
- REFEDS. *REFEDS Multi-Factor Authentication Profile v1.2* [online]. 15. 11. 2023. Available at: https://doi.org/10.5281/zenodo.10135577
- HUSGAFVEL, Miska, Martin KUBA, Nicolas LIAMPOTIS, and Luděk MATYSKA. *EOSC-ENTRUST D16.1 The AAAI for TREs Blueprint, first version* [online]. 11. 3. 2025. Available at: https://doi.org/10.5281/zenodo.15006945
- *Glossary for the OASIS Security Assertion Markup Language (SAML) V2.0* [online]. OASIS Standard, 15. 3. 2005 [Accessed 31. 12. 2025]. Available at: https://docs.oasis-open.org/security/saml/v2.0/saml-glossary-2.0-os.pdf (PDF)
- SAKIMURA, N. et al. *OpenID Connect Core 1.0 incorporating errata set 2* [online]. 15. 12. 2023 [Accessed 31. 12. 2025]. Available at: https://openid.net/specs/openid-connect-core-1_0.html
- HARDT, D., ed. *The OAuth 2.0 Authorization Framework* [online]. RFC 6749, October 2012. Available at: https://doi.org/10.17487/RFC6749
- FOSTER, Ian, Carl KESSELMAN, and Steven TUECKE. *The Anatomy of the Grid - Enabling Scalable Virtual Organizations* [online]. 29. 3. 2001. Available at: https://doi.org/10.48550/arXiv.cs/0103025
- *Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act)* [online]. OJ L 152, 3.6.2022, pp. 1–44. Available at: https://eur-lex.europa.eu/eli/reg/2022/868/oj/eng
- *Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast)* [online]. OJ L 172, 26.6.2019, pp. 56–83. Available at: https://eur-lex.europa.eu/eli/dir/2019/1024/oj/eng
- WG SENSI. *Charter of Working Group Sensitive Data* [online]. 22. 1. 2024 [Accessed 31. 12. 2025]. Available at: https://www.eosc.cz/media/3680820/charta_ps_sensi.pdf (PDF)
- *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)* [online]. OJ L 119, 4.5.2016, pp. 1–88. Available at: https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng
- ARDC. *Data sharing policy development guidelines* [online]. 20. 1. 2023. Available at: https://doi.org/10.5281/zenodo.7553182
- SÆTROM, Pål et al. *EOSC-ENTRUST D13.4 Year one version of EOSC-ENTRUST Blueprint & Interoperability Framework* [online]. 10. 12. 2024. Available at: https://doi.org/10.5281/zenodo.14362388
- DARE UK. *DARE UK Federated Architecture Blueprint* [online]. Version 2.2. 20. 11. 2024. Available at: https://doi.org/10.5281/zenodo.14192786

Appendix A: Glossary

NRP

National Repository Platform for Research Data[38] project responds to the need to support the implementation of the EOSC initiative in the Czech Republic[39] by creating and piloting a key component of the National Data Infrastructure[40] in order to consolidate the fragmented research data environment, to provide scientists with the necessary tools and facilities for data management according to the FAIR principles[41], and to ensure sufficient storage capacity and reliable repositories. The project also includes technical and methodological support and creating coordinated educational support.

Repository[42] is the technical, personal, and procedural provision of a long-term storage for the preservation and publication of citable digital objects.

Repository system[43] is a software package for operating a repository. In the context of NRP, there are three basic repository systems with first-class support: CESNET Invenio[44], CLARIN-DSpace[45] and ASEP/ARL[46]. However, it is also possible to operate repositories using alternative repository systems such as Digitalia-Islandora[47].

Basic and advanced services[48] (in the context of NRP) is a collection of tools and services developed and integrated to enable management of metadata profiles, working with licenses, data access control, FAIR-ification of data, data management planning, data and metadata collection automation, and integration of computational workflows.

---

[38] https://www.eosc.cz/en/projects/national-repository-platform-for-research-data-nrp/
[39] https://www.eosc.cz/en/about-eosc-cz/initiative-eosc-cz
[40] ibid.
[41] https://doi.org/10.1038/sdata.2016.18
[42] https://www.eosc.cz/en/projects/outcomes-of-the-eosc-cz-initiative#repositories_conditions
[43] ibid.
[44] https://nrp-cz.github.io/docs/
[45] https://github.com/ufal/clarin-dspace?tab=readme-ov-file
[46] https://github.com/LIBCAS/ASEP-ARL?tab=readme-ov-file#en
[47] https://github.com/ArtsFacultyMU/digitalia/wiki
[48] https://www.eosc.cz/en/projects/national-repository-platform-for-research-data-nrp/

IAM

Identity

User is a natural person using a software system (e.g. AAI, R/S). We assume a user has one legal identity for the purpose of entering into a legally binding contact or participating in legal proceedings.

Digital identity is a set of data in a digital system identifying a user, which often contains personal information. A user typically has multiple digital identities of various kinds: **national identity** is provided by states, **academic identity** is provided by academic institutions, and **social identity** (aka social login) is provided by social networking services.

User account is a set of data in a software system associated with a digital identity, which usually contains personal information, system settings and other data required to access the system's functionality. A user typically has numerous user accounts across many systems.

Service account is a representation of machine access to a software system. It has its own digital identity independent from the digital identities of users who control it.

Identity linking[49] (aka account linking) is a process of connecting the user's digital identities for consistent user identification or representation, accounting of resource access and usage, and traceability and security incident response.

Identity vetting[50] is a process to verify that a digital identity represents a particular user; often used in a way that implies elevated level of identity assurance.

Identity assurance[51,52] is a degree of trust in the link between a digital identity and a user. The trust has three elements: how good is the identity vetting process (identity assurance level), how safe is the authentication process (authentication assurance level), and how reliable is the identity federation (federation assurance level).

Identity freshness[53] is a degree to which the digital identity and associated data is up to date. It is usually expressed as the maximum allowed latency between user information changing and digital identity receiving an update (e.g. within a day, within a month).

Identity provenance[54] is information about the entities, activities, and agents used to record how digital identity data was derived to assess its quality, reliability, and trustworthiness.

---

[49] https://aarc-community.org/wp-content/uploads/2017/03/AARC-JRA1.4H.pdf
[50] https://doi.org/10.6028/NIST.SP.800-63-4
[51] ibid.
[52] https://eur-lex.europa.eu/eli/reg_impl/2015/1502/oj/eng
[53] https://doi.org/10.5281/zenodo.10277233
[54] https://doi.org/10.6028/NIST.IR.8112

Authentication

Authentication[55] is the process of verifying that a user owns a digital identity; often in privacy-preserving manner (contrast with identity vetting).

Multi-factor authentication[56,57] is a process of verifying that a user owns a digital identity using two or more distinct authentication factors: something the user knows (e.g. password, PIN), has (e.g. smartphone, key) and is (e.g. fingerprint).

User login is a process that consists of user authentication immediately followed by access to a user account in a software system.

Single sign-on[58] is a mechanism that allows a user to log into multiple software systems while authenticating just once for the session duration. **Single logout** is an analogous mechanism allows a user to log out of multiple software systems with one action.

---

[55] https://en.wikipedia.org/wiki/Authentication
[56] https://en.wikipedia.org/wiki/Multi-factor_authentication
[57] https://doi.org/10.5281/zenodo.10135577
[58] https://en.wikipedia.org/wiki/Single_sign-on

Authorisation

Authorisation[59] is the process of establishing an authorisation policy and specifying whether a user has the right to access a resource.

Access control[60] is the process of deciding whether a user has the right to access a resource. In a broad sense, the term also encompasses authentication and authorisation.

Role-based access control[61] is an access control variant in which the authorisation policy is made of *subject–role–operation–object* rules and an access decision is based on subject's roles. This approach is common thanks to its flexibility and simplicity. **Group-based access control** replaces roles with groups but is equivalent.

Attribute-based access control[62] is an access control variant in which the authorisation policy is made of *subject–operation–object–environment* rules and an access decision is based on entities' attributes. This approach is rare due to its complexity but is more powerful than RBAC. **Policy-based access control** focuses to policies but is equivalent.

Access control mechanism[63] is an implementation of access control process. It consists of four functional points: a policy administration point manages authorisation policies, a policy information point provides relevant attributes, a policy decision point makes access control decisions, and a policy enforcement point enforces the decisions; and a context handler that orchestrates the process. In a complex or distributed environment like NRP, these functional points are multiplied and distributed across several software systems.

Note: although the term access control mechanism and the terms for its functional points are defined in the context of ABAC, we consider the concepts they describe to be general and so we use these terms in the context of RBAC, too.

---

[59] https://en.wikipedia.org/wiki/Authorization
[60] https://en.wikipedia.org/wiki/Computer_access_control
[61] https://en.wikipedia.org/wiki/Role-based_access_control
[62] https://doi.org/10.6028/NIST.SP.800-162
[63] ibid.

Auditing

Accounting[64] is the process of collecting data to enable Auditing. This data should be able to answer the questions of who, what, where, when, why and to whom.

Auditing[65] is the process of reviewing data collected in accounting to ensure compliance or to investigate anomalous behaviour.

Logging[66] is the process of keeping a log of events that occur in a computer system, such as problems, errors, or broad information on current operations to monitor and understand the operation of the system, to debug problems, or during an audit.

Monitoring[67] is usually understood as the technical process of collecting data to determine system availability and performance, be it network, hardware, or software. More generally, it is the process of continuously observing system properties and events for a certain purpose.

---

[64] https://doi.org/10.5281/zenodo.15006945

[65] ibid.

[66] https://en.wikipedia.org/wiki/Logging_(computing)

[67] https://en.wikipedia.org/wiki/Application_performance_management

Federation

Identity provider[68] is a software service that authenticates a user and releases information about their digital identity to service providers using a specific protocol. It is also responsible for managing sessions to provide single sign-on and single logout capabilities.

Note we prefer this term even in cases when the underlying protocol is OIDC and the more appropriate term would be OpenID Provider.

Service provider[69] is a software service that provides resource access to users authenticated by an identity provider using a specific protocol. A service provider is also usually responsible for deciding whether a user has access to a resource and enforcing that decision. In the context of NRP, data repositories and related services are service providers; although these might not necessarily map one-to-one.

Note we prefer this term even in cases when the underlying protocol is OIDC and the more appropriate term would be Relying Party.

OIDC[70] is an authentication protocol built on top of OAuth 2.0[71]. It provides a standardized way to perform user authentication and authorization, while also providing additional features such as user profile information and session management to enable single sign-on.

SAML[72] is an XML-based open standard enabling single sign-on and facilitating exchange of SAML assertions containing information about a user's identity, authentication status, and authorized access rights.

eduGAIN[73] is an inter-federation service connecting academic identity federations around the world. End-users authenticate at identity providers and gain access to service providers. **eduID.cz**[74] is the Czech academic identity federation.

---

[68] https://docs.oasis-open.org/security/saml/v2.0/saml-glossary-2.0-os.pdf
[69] ibid.
[70] https://openid.net/specs/openid-connect-core-1_0.html
[71] https://doi.org/10.17487/RFC6749
[72] https://www.oasis-open.org/standard/saml/
[73] https://edugain.org/
[74] https://www.eduid.cz/wiki/en/

AAI

Authentication and authorisation infrastructure is a set of technical components and related processes used to verify user's digital identity so that permissions can be assigned to the user so that the digital resources they access can be both shared and protected.

Perun AAI[75] is a complex open-source software solution, a mix of adapted OSS components and own development, representing a technological foundation for an EOSC-compliant AAI based on AARC BPA[76] in the worldwide federated environment of academic institutions and research infrastructures.

Perun IdM is one of two main Perun AAI components. It provides user identity and access management: managing user's digital identities, managing user's memberships in VOs and groups, assigning resources to groups, and provisioning[77] information to services providers for local access control.

Perun ProxyIdP is one of two main Perun AAI components. It facilitates user authentication and may control access to services: exchanging authentication requests and responses between identity and service providers, enriching the responses with user attributes from Perun IdM, and managing sessions to provide single sign-on and single logout capabilities.

e-INFRA CZ AAI[78] is an AAI instance, powered by Perun AAI[79] and operated by CESNET[80], used to control user access to and within the Czech national research e-infrastructure e-INFRA CZ[81] and to fulfil the needs of implementing EOSC in Czech Republic[82] and EU[83].

Virtual organisation[84] is a dynamic collection of individuals, institutions, and resources which engage in flexible, secure, and coordinated resource sharing.

---

[75] https://perun-aai.org/
[76] https://aarc-community.org/architecture/
[77] https://en.wikipedia.org/wiki/Provisioning_(technology)#User_provisioning
[78] https://perun.e-infra.cz/
[79] https://perun-aai.org/
[80] https://www.cesnet.cz/
[81] https://www.e-infra.cz/
[82] https://www.eosc.cz/
[83] https://eosc.eu/
[84] https://doi.org/10.48550/arXiv.cs/0103025

Data

Data[85] (under DGA) means any digital representation of acts, facts or information and any compilation of such acts, facts, or information, including in the form of sound, visual, or audiovisual recording.

Research data[86] (in PSI) means documents in a digital form, other than scientific publications, which are collected or produced in the course of scientific research activities and are used as evidence in the research process, or are commonly accepted in the research community as necessary to validate research findings and results.

Sensitive data (in NDI) refers to data subject to the following restrictions: strictly intended for the internal use of a precisely defined group of individuals; requires special regulation due to its nature, typically explicitly protected by law or based on trade secrets, contracts, licenses, etc; access by unauthorized individuals outside the specified group is likely to cause significant or damaging consequences. The categorization of data sensitivity is determined by the data owner unless otherwise specified by law. The categorization of data sensitivity is determined by the data owner unless otherwise specified by law.

Personal data[87] (under GDPR) mean any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person. **Personal electronic health data**[88] (under EHDS) means data concerning health and genetic data, processed in an electronic form.

[85] https://eur-lex.europa.eu/eli/reg/2022/868/oj/eng
[86] https://eur-lex.europa.eu/eli/dir/2019/1024/oj/eng
[87] https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng
[88] https://eur-lex.europa.eu/eli/reg/2025/327/oj/eng

Actors

Note legal person actors are usually represented by a natural person in system interactions.

Data subject[89] (under GDPR) means an identified or identifiable natural person.

Data controller[90] (under GDPR) means the natural or legal person, public authority, agency, or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data.

Data processor[91] (under GDPR) means a natural or legal person, public authority, agency, or other body which processes personal data on behalf of the controller.

Data recipient[92] (under GDPR) means a natural or legal person, public authority, agency, or another body, to which the personal data are disclosed, whether a third party or not.

Data rights holder[93] (per DSSC) means a natural or legal person with legitimate interests to exercise rights under Union law affecting the use of data or imposing obligations on other parties in relation to the data. Colloquially referred to as **data owner**.

Data producer means a natural person, who is not a data subject with respect to the specific data in question, or group thereof that produces research data and then deposits it into data repositories. A data producer may be the data rights holder or have a contract with one.

Data holder[94] (under DGA) means a legal person, including public sector bodies and international organisations, or a natural person who is not a data subject with respect to the specific data in question, which, in accordance with applicable Union or national law, has the right to grant access to or to share certain personal data or non-personal data. **Health data holder**[95] (under EHDS) means any natural or legal person, public authority, agency or other body in the healthcare or the care sectors or [...] that has either the right or obligation to process personal electronic health data for the provision of healthcare or care or [...] or the ability to make available non-personal electronic health data [...]

Data user[96] (under DGA) means a natural or legal person who has lawful access to certain personal or non-personal data and has the right to use that data for commercial or non-commercial purposes. **Health data user**[97] (under EHDS) means a natural or legal person, including Union institutions, bodies, offices, or agencies, which has been granted lawful access to electronic health data for secondary use pursuant to a data permit or a health data request approval or [...]

Data access committee[98] means one or more natural persons who review data access requests and make decisions about who can access sensitive data in line with applicable

---

[89] https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng
[90] ibid.
[91] ibid.
[92] ibid.
[93] https://dssc.eu/space/BVE2/1071252161/Alphabetical+List+of+All+Defined+Terms+in+Blueprint+v2.0
[94] https://eur-lex.europa.eu/eli/reg/2022/868/oj/eng
[95] https://eur-lex.europa.eu/eli/reg/2025/327/oj/eng
[96] https://eur-lex.europa.eu/eli/reg/2022/868/oj/eng
[97] https://eur-lex.europa.eu/eli/reg/2025/327/oj/eng
[98] https://ega-archive.org/access/data-access-committee/what-is-dac/

policies. **Health data access body**[99] (under EHDS) means a legal person to which health data holders must make electronic health data available, and which assesses the information provided by health data applicants, based on which it should be able to issue a data permit for the processing of personal electronic health data; it has many more other responsibilities.

[100] means a natural person that is the holder of an independent grant and the lead researcher for the grant project.

Activities

Data processing[101] (under GDPR) means any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction

Data sharing[102] (under DGA) means the provision of data by a data subject or a data holder to a data user for the purpose of the joint or individual use of such data, based on voluntary agreements or Union or national law, directly or through an intermediary, for example under open or commercial licences subject to a fee or free of charge.

Secondary use[103] (under EHDS) means the processing of electronic health data for the purposes other than the initial purposes for which they were collected or produced. Scientific research is one of the permitted purposes.

Documents

Data access request[104] is submitted by a principal investigator to a data access committee for the purpose of obtaining access to sensitive research data.

Health data request[105] (under EHDS) is submitted by a health data applicant to a health data access body for the permitted purposes with the aim of obtaining a response only in an anonymised statistical format.

Health data access application[106] (under EHDS) is submitted by a health data applicant to a health data access body the purpose of public interest with the aim of obtaining access to electronic health data in a pseudonymised or anonymised format.

Health data permit[107] (under EHDS) is issued by a health data access body following for the purposes of granting access to electronic health data.

---

[99] https://eur-lex.europa.eu/eli/reg/2025/327/oj/eng
[100] https://en.wikipedia.org/wiki/Principal_investigator
[101] https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng
[102] https://eur-lex.europa.eu/eli/reg/2022/868/oj/eng
[103] https://eur-lex.europa.eu/eli/reg/2025/327/oj/eng
[104] https://ega-archive.org/access/request-data/how-to-request-data/
[105] https://eur-lex.europa.eu/eli/reg/2025/327/oj/eng
[106] ibid.
[107] ibid.

Policies

Note the policies may be defined on national, institutional, repository, collection, dataset, etc. level or a combination thereof.

Data management plan[108] means a structured document outlining how research data will be collected, organized, stored, shared, and preserved throughout and after a research project.

Data sharing policy[109] means a structured document outlining the types of data that is in custody, the principles and strategy of sharing it, and related governance and management procedures. Usually says *whether* data should be shared.

Data access policy means a structured document outlining who can access the data and under what conditions such as the purpose, training, technology, etc. Usually says *how* data can be shared if it is.

Agreements

Terms of Service[110] means a legally binding contact between a service provider and a service user to which the latter must abide to use the offered service.

Service level agreement[111] means a legally binding contract between a service provider and a service user defining the aspects of providing the subject service such as quality, availability, responsibilities, etc.

Deposit license agreement[112] means a legally binding contact between a data submitter and the repository administrator, in which the submitter declares they are authorised by the data rights holder and grant a licence to the administrator to use the dataset for the sole purpose of permanent storage and making the dataset available in the repository.

Data use agreement means a legally binding contract between a data holder and data user specifying the purposes and conditions for accessing, storing, analysing, redistributing, etc. the subject data.

Data transfer agreement is a legally binding contract between a data holder and data user specifying the conditions for transferring the subject data from the data holder's storage to the data user's: technologies, safeguards, etc.

Data processing agreement[113] (under GDPR) is a legally binding contract or other legal act under Union or Member State law, that is binding on the processor with regard to the controller and that sets out the subject-matter and duration of the processing, the nature and purpose of the processing, the type of personal data and categories of data subjects and the obligations and rights of the controller.

---

[108] https://en.wikipedia.org/wiki/Data_management#Data_management_in_research
[109] https://doi.org/10.5281/zenodo.7553182
[110] https://en.wikipedia.org/wiki/Terms_of_service
[111] https://en.wikipedia.org/wiki/Service-level_agreement
[112] https://www.eosc.cz/media/4024968/legal_licensing.pdf
[113] https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng

Research

Project[114] (in research) is an approved time-bound research activity with a clear purpose and goals led by a principal investigator, usually involving other project members, that requires use of research data.

Virtual research environment[115] is an online system helping researchers collaborate, usually including collaboration support, document hosting, and some discipline-specific tools, such as data analysis, visualisation, or simulation management.

Workflow management system[116] provides an infrastructure for the set-up, performance, and monitoring of a defined sequence of tasks arranged as a workflow application.

Scientific workflow system[117] is a specialized form of a workflow management system for composing and executing a series of computational or data manipulation steps, or workflow, in a scientific application. Each system typically provides a visual front-end, allowing the user to build and modify complex applications with little or no programming expertise. There are two SWS platforms available in NRP: Galaxy[118] and Lexis[119].

Trusted research environment[120] is a highly secure computing environment providing remote access to sensitive data for approved research. A "maximal" TRE features a research analytics zone, a secure data zone, and a query management zone. It interfaces with the outside world via a security server. [121]

Secure processing environment[122] (under both DGA and EHDS) is the physical or virtual environment and organisational means to ensure compliance with Union law, in particular with regard to data subjects' rights, intellectual property rights, and commercial and statistical confidentiality, integrity and accessibility, as well as with applicable national law, and to allow the entity providing the secure processing environment to determine and supervise all data processing actions, including the display, storage, download and export of data and the calculation of derivative data through computational algorithms.

---

[114] https://doi.org/10.5281/zenodo.14362388
[115] https://en.wikipedia.org/wiki/Virtual_research_environment
[116] https://en.wikipedia.org/wiki/Workflow_management_system
[117] https://en.wikipedia.org/wiki/Scientific_workflow_system
[118] https://usegalaxy.cz/
[119] https://docs.lexis.tech/
[120] https://www.hdruk.ac.uk/access-to-health-data/trusted-research-environments/
[121] https://doi.org/10.5281/zenodo.14192786
[122] https://eur-lex.europa.eu/eli/reg/2022/868/oj/eng

Appendix B: Example

7.1 Alice's Story: Depositing Sensitive Data

Alice is a researcher in academia. She has been working on a research project the result of which is sensitive research data. Both the project funder and the journal in which she is going to publish her research require Alice to deposit that dataset in a recognized repository in line with FAIR data principles, so she devises and follows a data management plan.

Alice consults the National Repository Catalogue and identifies a suitable thematic repository in the National Data Infrastructure. She navigates to its homepage, and reviews its terms of service, data sharing policy, data access policy, and metadata model to understand her rights and responsibilities.

The Repository administrator Grace has restricted data uploads to authorised users since the National Repository Platform storage capacity is a limited resource. Thus, Alice authenticates via her home institution, registers with AAI which collects information about her (like a bona-fide researcher status), and reviews the notices that AAI presents to her (like NDI ToU and privacy policy). The she can access the Repository.

Once in, Alice obtains the submitter role from Grace, deposits the dataset, and provides its metadata, and accepts dataset's depository license agreement, but keeps the dataset locked due to its sensitive nature in line with DMP. Finally, the Repository automatically propagates the metadata to the National Metadata Directory (NMD).

7.2 Bob's Story: Processing Sensitive Data

Bob is a researcher in academia. He is currently working on a research project for which he requires specific data. Fortunately, he discovers similar dataset has already been produced by Alice and published in NDI, so he need not produce it himself. He uses NMD to identify the Repository in which the dataset is stored.

The access to the dataset is restricted, therefore Bob authenticates via his home institution, registers with AAI which collects information about him (such as his academic affiliation), and reviews the notices that AAI presents to him. Then he can access the Repository.

Once in, Bob submits a data access request for the dataset. Grace notifies the data access committee members (the data owner, a scientific advisor, and legal and ethics specialists). If the negotiation is successful, DAC approves the request. Bob signs data transfer and use agreements, which are converted to Rego policies and coupled with the dataset, giving him conditional access to the dataset in the Repository.

In the next phase, Bob wants to use the dataset. However, he cannot simply download it as the terms of DUA and DTA dictate, he must use a TRE and ensure a secure data transfer. So, Bob negotiates use of a TRE operated by Ted in NDI, who must sign a data processing agreement with Grace as a prerequisite for any data transfer. Next, Ted creates an isolated project space in the TRE for Bob, giving him the PI role, and attaching DPA to the project.

Afterwards, Bob authenticates via his home institution. As the TRE requires elevated identity and authentication assurance levels, Bob must perform step-up authentication and undergo identity vetting in the AAI, which then lets him access the TRE and releases the assurances.

Bob’s PI role gives him access to TRE capabilities like such as file transfer. He requests data download based on the dataset’s PID he found in NMD.

TRE's file transfer tool goes to AAI to exchange its access token (NDI scope, TRE audience) for another one that would enable it to access the dataset on Bob’s behalf (download scope, repository audience). The tool uses the new token to authenticate with the Repository and requests the dataset. A context handler is triggered by the Repository to orchestrate the access decision process. The handler obtains a general authorisation policy from the AAI and pushes it to the Repository policy engine together with the DUA and DTA policies coupled with the dataset. Then the handler collects the user attributes from AAI, the dataset attributes from the Repository, the file transfer operation properties from the TRE, and the environment attributes from all systems. In the end, it triggers the policy engine to evaluate the policies with the provided attributes, and it converts the decision to a format that the Repository can understand and enforce. Finally, the tool is authorised to execute the transfer, and Bob can enjoy the benefits of FAIR data.