D3.2 The EUreka3D AAI architecture

Due date: M22

Dissemination level: Public

Authors:
Ignacio Lamata Martinez (EGI Foundation)

HISTORY OF CHANGES			
Version	Date	Author	Comments
0.1	15/10/2024	Ignacio Lamata Martinez (EGI)	First draft
0.2	24/10/2024	Jolan Wuyts (EF), Valentina Bachi (Photoconsortium)	Internal review
0.3	28/10/2024	Ignacio Lamata Martinez (EGI)	Completed version
0.4	07/11/2024	Antonella Fresa (Photoconsortium)	Final review
1.0	08/11/2024	Valentina Bachi (Photoconsortium)	Submitted version

TABLE OF CONTENTS

TABLE OF CONTENTS 2
EXECUTIVE SUMMARY 3
1. INTRODUCTION 4
1.1 ROLE OF THIS DELIVERABLE IN THE PROJECT 4
1.2 RELATIONSHIP TO OTHER DELIVERABLES 4
1.3 LIST OF ACRONYMS 4
1.4 STRUCTURE OF THE DOCUMENT 5
2. BASIC CONCEPTS ON AAI 6
2.1 AUTHENTICATION 6
2.2 AUTHORISATION 7
2.3 SINGLE SIGN-ON 7
2.4 PROTOCOLS FOR AAI 8
3. EGI CHECK-IN 9
3.1 IDENTITY MANAGEMENT IN CHECK-IN 9
3.2 ACCESS MANAGEMENT IN CHECK-IN 10
3.2.1 Authentication in Check-in 10
3.2.2 Authorisation in Check-in 11
3.2.3 Putting Authentication and Authorisation together 13
3.3 CHECK-IN ENVIRONMENTS 14
3.4 EOSC AAI AND CHECK-IN 15
4. AAI IN EUREKA3D 16
4.1 AUTHENTICATION OF USERS 17
4.2 AUTHORISATION OF USERS 17
4.2.1 Organisation of the EUreka3D community 17
4.2.1 Joining the EUreka3D community 19
4.3 ACCESS IN DATAHUB 20
4.4 ACCESS TO THE CLOUD COMPUTE 21
5. CONCLUSIONS 23
REFERENCES 24
ANNEX A: USER ONBOARDING PROCEDURE 25

EXECUTIVE SUMMARY

EUreka3D brings numerous advances for the cultural heritage sector, such as a modernised digital workflow for cultural heritage assets, the transition to the cloud, a well-established infrastructure and services, an enriching training and dissemination programme, and a range of high quality 3D digital collections working towards the concept of "Memory Twin" in cultural heritage, that validate the technical approach. The mortar that protects all these bricks, and glues them together, is *security*, an essential part of the system working intrinsically, but as transparent as possible, at every stage.

What value would cultural heritage assets have if they could be manipulated and changed, maliciously or by mistake, by the wrong hands? Data must be protected, and the data owner should be enabled to decide whom the data can be shared with, for example to work cooperatively on shared projects with partner organisations. While the creation of the EUreka3D infrastructure is precious, the data coming from the Content Providers is the outcome that gives meaning to the whole system. It is the most valuable asset to protect and therefore an Authentication and Authorisation Infrastructure (AAI) is an essential element of EUreka3D.

This document explains the measures taken in EUreka3D to protect the assets uploaded by Content Providers (including raw data, metadata and paradata) from unauthorised users. It also describes how users are authenticated to identify them, and how the authorisation rules are created to assess what they can do in the system. Content Providers need to be in control of the data they produce and share publicly. As EUreka3D serves the Cultural Heritage Institutions (CHIs), the digital implementation of its infrastructure should reflect the requirements coming from the cultural heritage sector, by organising the EUreka3D community in different groups or sets of users.

EUreka3D uses a service from EOSC European Open Science Cloud called EGI Check-in to protect its assets, and the EUreka3D community is organised through a Virtual Organisation, that divides users into different groups according to their responsibilities and access capabilities. The system, described in this document, uses some of the most widespread Web standards, such as OIDC and SAML, and supports single sign-on and digital identity management for users. It also facilitates the management of users' personal data in accordance with GDPR regulations.

1. Introduction

1.1 Role of this deliverable in the project

This deliverable documents the AAI configuration implemented in the EUreka3D project, which in essence deals with two big categories:

- The protected access to the EUreka3D infrastructure. This includes access to the EGI DataHub used in EUreka3D.
- The protection of the stored data, uploaded by the Content Providers.

1.2 Relationship to other deliverables

This document is closely related to the following deliverables:

- **Deliverable 3.1** *"Report on the EUreka3D services and resource hub: design and implementation"* (Project Month 4, April 2023) that documents the initial technical design of the EUreka3D services and resource hub.
- **Deliverable 3.3** *"Final report on the EUreka3D services and resource hub"* (Project Month 22, October 2024), which is the update of D3.1 at the end of the project.

There is also an evident link between the intermediary technical progress reports (D1.3, D1.4, D1.5, D1.6) and the integration reports (D1.2 and D1.7), where updates on the progress of WP3 tasks, including the connection and interoperability with Europeana, are provided.

1.3 List of acronyms

Acronym	Description
AAI	Authentication and Authorisation Infrastructure
AARC	Authentication and Authorisation for Research and Collaboration
AARC BPA	AARC Blueprint Architecture
API	Application Programming Interface
AUP	Acceptable Use Policy
CH	Cultural Heritage
CHI	Cultural Heritage Institutions
EOSC	European Open Science Cloud

GDPR	General Data Protection Regulation
GUI	Graphical User Interface
HTTPS	Hypertext Transfer Protocol Secure
IT	Information Technology
OAuth	Open Authorisation
OIDC	OpenID Connect
OS	Operating System
PKCE	Proof Key for Code Exchange
QoS	Quality of Service
SAML	Security Assertion Markup Language
SSO	Single Sign-On
TLS	Transport Layer Security
VM	Virtual Machine
VO	Virtual Organisation

1.4 STRUCTURE OF THE DOCUMENT

The rest of this document is organised as follows:

- **Section 2** briefly explains some basic yet important concepts to understand AAI.
- **Section 3** describes EGI Check-in, the AAI service used in EUreka3D.
- **Section 4** explains the use of AAI in the EUreka3D project.
- Finally, **Section 5** provides some conclusions
- In addition, the Annex A is the EUreka3D User Onboarding Procedure that must be followed to create a group in the EUreka3D Virtual Organization in EGI Check-in, and to allow a registered user to access the EUreka3D Data Hub.

2. Basic concepts on AAI

Authentication and Authorisation Infrastructure (AAI) refers to the set of software tools, protocols, hardware devices and principles that enable controlled access to protected resources, such as data, servers and applications. In a federated environment, these resources are distributed across distributed sites.

The controlled access to the resources is enforced and implemented through two mechanisms: authentication and authorisation.

2.1 Authentication

Authentication (sometimes written as *AuthN*) is the mechanism to prove the identity of the user, in other words, identifying **who the user is**. This is typically implemented with the knowledge of a password ("*something the user knows*"), the possession of a hardware key ("*something the user owns*''), such as the one shown in Figure 1, the possession of a code delivered by an application installed in a smartphone or via biometrics ("*something the user is*", such as a fingerprint).

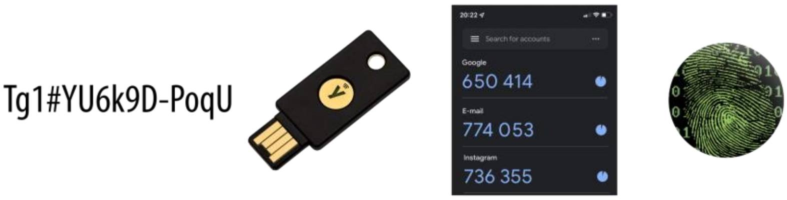
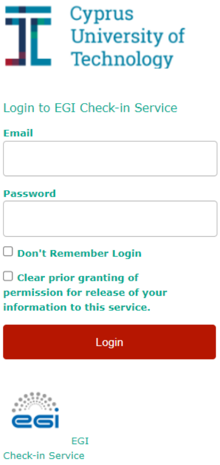


Figure 1: Example of different authentication methods

Recently, it has become common to implement a combination of these different types of authentication in order to improve security, a practice called **Two-Factor Authentication** (2FA) in which *e.g.* the user is asked for both a password and a code they receive on their phone. That way, a potential attacker will be unable to access a resource if only a password is compromised. Figure 2 shows the University of Oxford login process, based on a user name (an email address) and password, to prove the identity of the user.



Cyprus University of Technology

Login to EGI Check-in Service

Email

Password

□ Don't Remember Login

□ Clear prior granting of permission for release of your information to this service.

Login

EGI Check-in Service

Figure 2: Example of the authentication process done by Cyprus University of Technology

2.2 Authorisation

Authorisation (sometimes written as *AuthZ*) is the mechanism used to determine **what the user is allowed to do in a system**. Commonly, the authorisation process follows a previous authentication process, since a system must first identify the user to determine what permissions apply to them. Authorisation rules are typically in the form of attributes or roles previously assigned to the user, and the verification process is normally applied by the different applications.

An example of this process can be found in Figure 3, which shows a user trying to access some protected data. First, the system needs to identify who the user is, so it presents a login form where the user can input his/her credentials. Once these are validated, the authentication process is completed. The user's attributes are retrieved from a local database or external source and stored in memory, typically in some data structure called a *session* that will exist for as long as the user remains authenticated. Next, the system needs to verify that the action the user is attempting (reading some data) is accepted, so it verifies that the user's attributes contain the required permission. This is the authorisation process which, if completed successfully, will cause the system to deliver the requested data to the user.

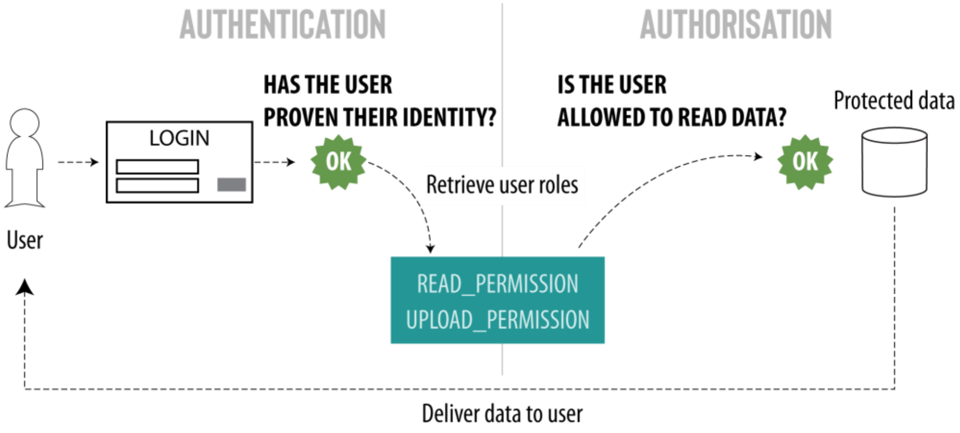


Figure 3: Flow showing a user attempt to access some protected data

2.3 Single Sign-On

Another interesting concept is *Single Sign-On* or *SSO*. In the above discussion, it was implicit that the user was accessing a single system. However, if users need to access different yet related systems, it is very tedious for them to be forced to authenticate separately on each of the systems they access. Moreover, this is inconvenient for users and becomes a problem if the resources the users access are federated, such as those from EGI.

SSO addresses this problem by allowing the user to authenticate once to access resources that are located in disparate systems, without the need to enter user credentials for each application or service. One of the ways to implement this is through a trusted third party that authenticates the user and can share this state with a set of trusted applications protecting the resources. EGI Check-in provides Single Sign-On in the EGI Federation, as explained in more detail in Section 3.

2.4 Protocols for AAI

Multiple protocols and standards can be used for authentication and authorisation. The most relevant industry-standard mechanisms for authentication and authorisation include OIDC, SAML and certificates (X.509):

- **OIDC** (OpenID Connect[1]) is an authentication protocol, which securely verifies the user connected to a system through, for example, a Web browser or mobile app. It works by adding an identity layer on top of the **OAuth 2.0 framework**[2], which offers *authorisation delegation*. OIDC works with Web protocols and has become the *de facto* standard for modern Web authentication.
- **SAML** (Security Assertion Markup Language[3]) is an open standard published by OASIS for exchanging authentication and authorisation data between different parties. Its syntax is based on XML and allows Web applications to transfer information between the system authenticating the user and the system providing access to a resource.
- **X.509** is a standard that defines the format of public key certificates. Public key cryptography is based on the notion of asymmetric cryptography, in which different but mathematically bound keys are used for encryption and decryption. The identity of the certificate owner is bound to the certificate using a digital signature, which requires an implicit or explicit trust in the entity that signs the certificate. X.509 certificates are widely used on the Web to protect communications (via the TLS and HTTPS protocols) and ensure secure access to information resources. Through the use of public key cryptography, it is possible to authenticate users.

These protocols are supported by EGI Check-in, as explained in the following section.

[1] https://openid.net/developers/how-connect-works

[2] https://oauth.net/2/

[3] https://www.oasis-open.org/standard/saml/

3. EGI Check-in

The EGI Check-in service[4] provides **identity** and **access management** components that facilitate users to access community services and resources. Check-in acts as an intermediate system connecting users, authentication servers and services, offering users authenticated access to services and enabling single sign-on. It supports widely adopted standards and open technologies, including OIDC/OAuth, SAML and X.509, which facilitates interoperability and integration with existing AAI services (responsible for managing the Authentication and Authorisation) of other Research Infrastructures and Research Communities.

As a summary, Check-in has the following features:

- It provides Single Sign-On.
- It accepts multiple federated authentication sources using different technologies.
- It is federated in eduGAIN as a service provider, publishing REFEDS RnS[5] and Sirtfi[6] compliance.
- It provides a GUI for user registration and management, which allows identity unification through account linking.
- It provides an API for programmatic user management.
- It can combine user attributes originating from various authoritative sources and deliver them to the connected service providers in a transparent way. This process contributes to GDPR compliance.
- It is connected with EOSC services.

Some of these aspects are discussed in the following sections.

3.1 Identity management in Check-in

The digital identity of a user refers to the digital information that exists about the user in the digital world. It is a *simplified* representation of the person in the real world. This set of data is made of attributes that define the user at different levels. Commonly, these attributes are spread around different sites, as represented in Figure 4. For example, a State can have digital information about the name, date of birth and similar personal details about a user, whereas a social network like Facebook can store information about the hobbies of the user. They all form different digital “identities” of the same user.

[4] https://www.egi.eu/service/check-in/
[5] https://refeds.org/
[6] See: https://refeds.org/sirtfi and https://aarc-project.eu/policies/sirtfi/

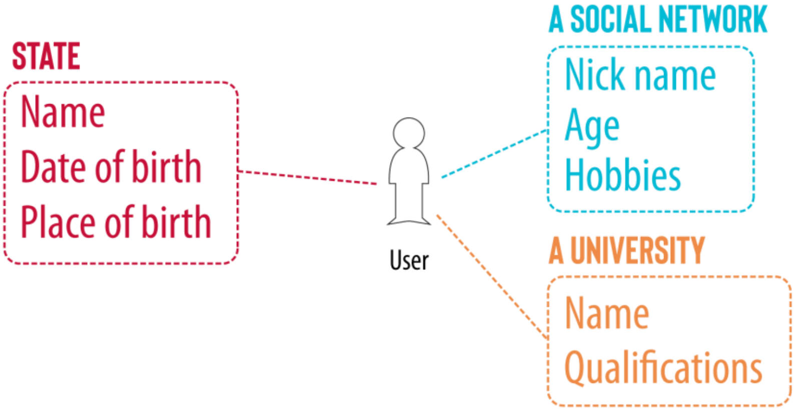


Figure 4: Digital identity of a user

Often, different entities are interested in different information about a user. Thus, the Government may be interested in someone's date and place of birth, while a recruiter may only need to know the person's age and qualifications. However, although these different "identities" refer to the same person, they are unrelated in the digital world. Check-in, as an identity management system, can **integrate user attributes from different sources**, so that these different "identities" are unified into one.

The different digital identities of a user are usually supported by accounts. Herein, a user can have one identity on GitHub, containing their development projects, another on LinkedIn, containing employment-related information, and another on ORCID, containing relevant research information about the user. Check-in provides mechanisms for the user to indicate that these three accounts correspond to the same person, forming a unified user identity to access applications and systems.

3.2 Access management in Check-in

Access management is controlled through the two mechanisms explained in Section 2: authentication, as described in Section 3.2.1, and authorisation, as described in Section 3.2.2.

3.2.1 Authentication in Check-in

Check-in works as an *authentication proxy*, meaning that it acts as an intermediary system that connects the user with an authentication server, so users are redirected to a third party to carry out the actual authentication process. This authentication server receives the name of **identity provider** (idP). Herein, Check-in does not authenticate users by itself, it is the idP the entity that verifies the user's identity. This has multiple benefits: (i) Users do not need to create extra credentials (such as a password) for Check-in, they just use their existing credentials. (ii) Users do not have to store or share their passwords with Check-in, and (iii) Users can use their home organisation to authenticate, an entity they trust, and through the login process they are accustomed to. Check-in offers **a wide range of idPs** to users from which to authenticate. These include the large list of research institutions in eduGAIN[7] (which includes many Universities and other organisations in the academic and research world), and companies that publicly offer accounts, such as Google, LinkedIn and ORCID. This provides users with a high flexibility to choose their authentication method.

[7] https://edugain.org/

Additionally, Check-in provides its own idP, called EGI SSO[8], for users that do not have an account in any idP. Authentication through Home organisations in eduGAIN has a higher *level of assurance/trust*, provided that these organisations perform physical verifications of user's identities. For example, a university normally carries out some identification process for its students (ID cards, etc) whereas Google does not require any identification step from a real person to create an account. This level of assurance is exposed to applications to make authorisation decisions.

The authentication process is simple for the user and is depicted in Figure 5. It starts when the user accesses an application, such as EGI DataHub. As access to this application is protected, the user has to be identified first. The application renders a login button that the user must click to start the login process. If the user is already logged in to Check-in, they are granted immediate access. Otherwise, the user is redirected to the **Discovery page** of Check-in to select the idP he or she wants to use. Once the user selects his/her identity provider, he/she is redirected to the corresponding login page (in the figure, the login page of the University of Oxford) where the user can enter its credentials. After successful authentication, the user's information will be transferred and he/she will be identified to the application. It is time for the application to verify if the identified user is allowed to access the application, and under what role or permissions. This is done through the authorisation process that will be discussed in the following section.

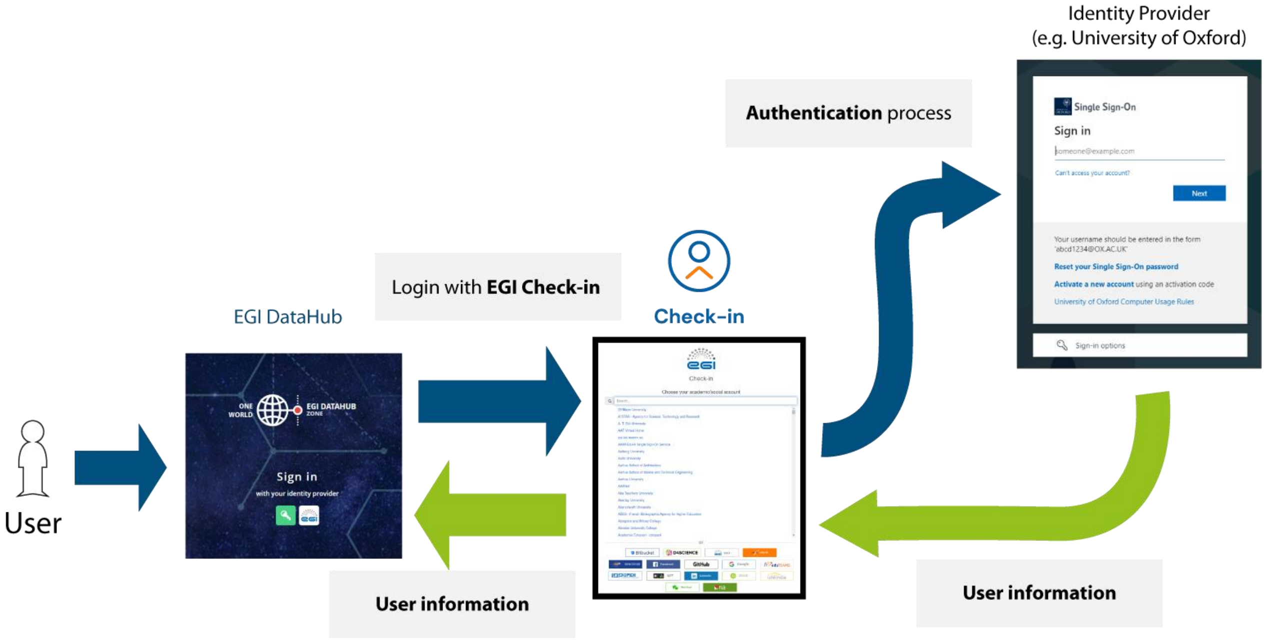

Identity Provider
(e.g. University of Oxford)
Authentication process
Single Sign-On
Sign in
someone@example.com
Can't access your account?
Next
Your username should be entered in the form 'abcd1234@OX.AC.UK'
Reset your Single Sign-On password
Activate a new account using an activation code
University of Oxford Computer Usage Rules
Sign-in options
Login with EGI Check-in
EGI DataHub
ONE WORLD
EGI DATAHUB ZONE
Sign in
with your identity provider
Check-in
User
User information
User information

Figure 5: A user accesses an application and authenticates via Check-in

3.2.2 Authorisation in Check-in

One of the main objectives of Check-in is to provide protected access to resources. In an academic environment, it is common to find the situation depicted in Figure 6, where a set of resources, such as servers, data or applications, need to be accessed by different research communities. This access has to be controlled, so for example Community A must be able to deploy servers to an infrastructure provider but not perform

[8] https://sso-admin.egi.eu/admin/

any action on a data set. Community B must be able to modify some aspects of an application, while the general public is only allowed to visualise, and not edit, these aspects. These access controls have to be enforced during the authorisation process.

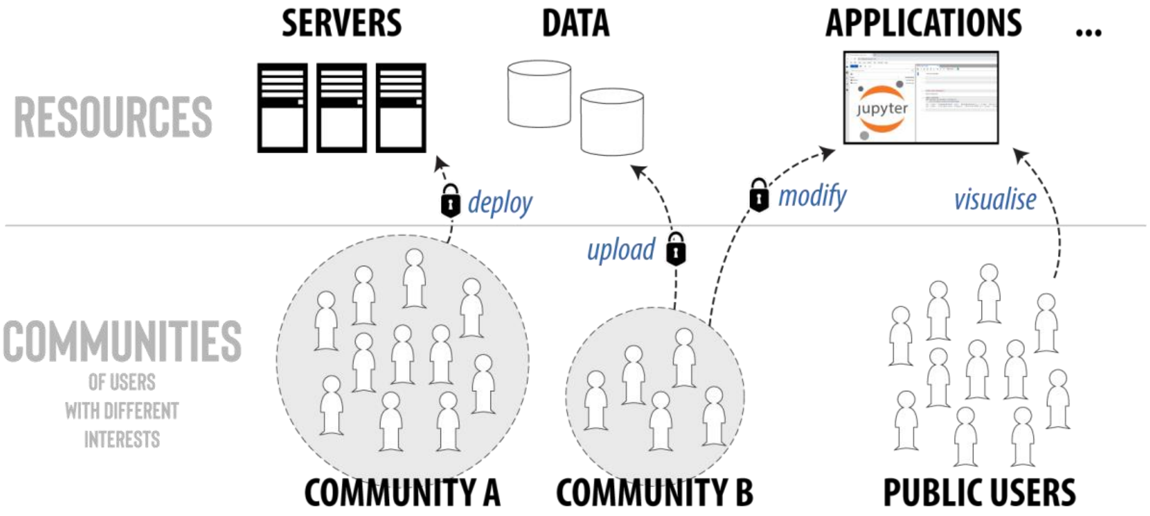


Figure 6: Authorised access to resources

Access to resources is typically implemented through an entity called a **Service Provider** (SP), which represents the services and applications that need to be accessed in a secure and trusted manner. SPs offer services to users, meaning that they provide a resource or run and manage an application. Authorisation rules are often, but not always, applied directly by SPs.

To facilitate the management of users and the organisation of the research communities (as in the example of Figure 6), Check-in uses the concept of Virtual Organisation. A **Virtual Organisation** (VO) is a group of users that represents a community with common research interests. It is also the main mechanism used by Check-in to grant access to resources and organise users. Syntactically it looks like the *hostname* part of a URL (but it has nothing to do with it), such as, for example: *culturalheritage.vo.egi.eu*.

VOs are managed autonomously, and they must have one or more administrators, called **VO Managers**, who are responsible for the operation of the VO. VO Managers are the contact point between EGI and a Community (i.e. the EUreka3D community in our case) and deal with all management tasks related to the VO, such as membership requests to accept new members.

VOs can have **groups** and **subgroups** to classify users. Applications can then make authorisation decisions based on the membership of these groups. Additionally, users can have **roles** assigned, which can also be used for authorisation decisions.

Although using VOs is a common mechanism to define authorisation in Check-in, other authorisation models based on identity assurance, capabilities and affiliation are also available.

3.2.3 Putting Authentication and Authorisation together

Once the general concepts have been explained separately, the diagram in Figure 7 can be easily understood. When a user accesses a Service Provider, the user first authenticates with an Identity Provider. For this, the user credentials are used, which can be a username and password or a client certificate, for example. Once authenticated, Check-in will combine the attributes provided by the idP together with other internal attributes. Since Check-in has to send some of the user attributes to the SP, it shows a **Consent Screen** to the user, where he/she can check what information will be transferred to the SP and the policies that govern his/her data use, including the information requested by the **GDPR**[9]. If accepted, the information will be transferred to the service, so it can be used by it to make authorisation decisions. These are often based on the user's membership of a VO, a group, and so on. For example, the application may deny access unless the user is a member of a specific VO, and may deny "write permissions" in the application unless the user is additionally a member of a specific group of that VO.

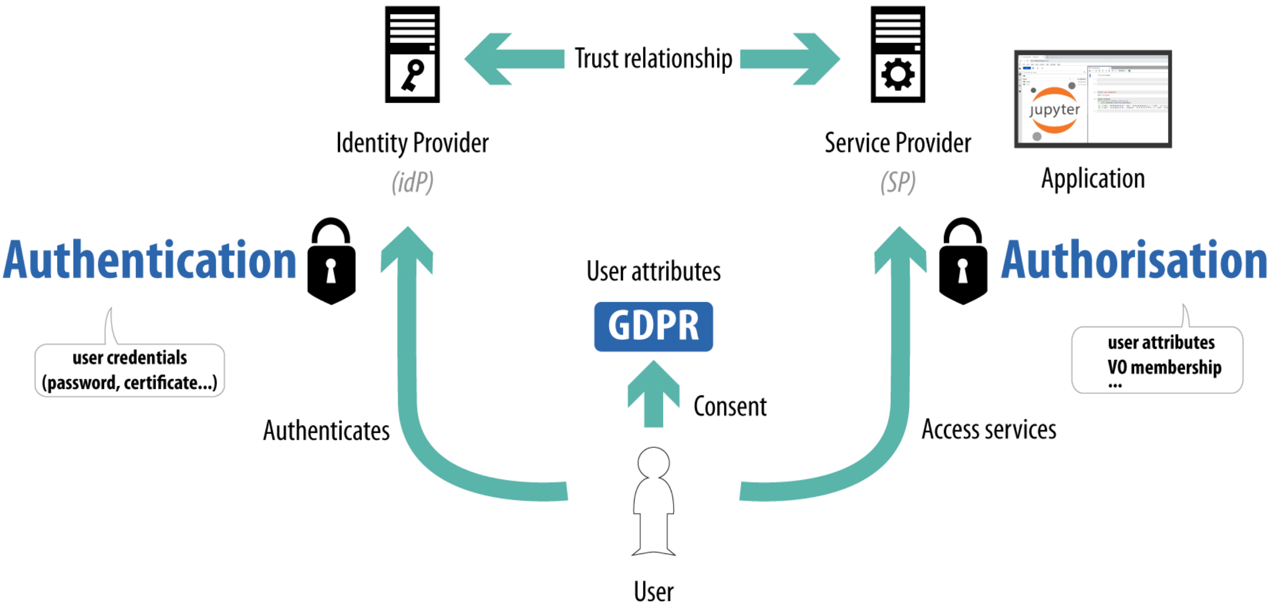


Figure 7: User accessing a Service Provider

It is important to highlight that idP and SP must maintain a relationship of trust. Indeed, services must be registered on a Check-in portal called the **Federation Registry**[10], where the SP configures technical and legal aspects: the protocols that will be used in the communication with the user and Check-in, the Privacy Policy of the service, the Acceptable Use Policy and Conditions of Use, etc.

[9] https://eur-lex.europa.eu/eli/reg/2016/679/oj

[10] https://aai.egi.eu/federation

3.3 Check-in environments

Check-in is run in three different instances that have different purposes:

- **Development**. This instance is mainly used for developing Check-in as a product. New and experimental features are tested in this environment before being moved to another environment. For this reason, this instance is considered *less stable*, as Check-in developers can restart it at any time without warning. The servers running Check-in in this environment have less computing power than in Production. Some identity providers, such as those from eduGAIN, are not available for authentication in this environment. Additionally, services can be registered and modified in the Federation Registry (see Section 3.2.3) without any technical verification or approval from the Check-in team.
- **Demo**. This instance is used for testing and piloting activities. It is a replica of the Production instance, both in terms of hardware and software used, so that developers can be sure that they test their applications in an environment as close as possible to the Production environment. Registration and Modification of services in the Federation Registry require technical verification from the Check-in team. Normally, research communities will test their applications in this environment and move to Production once everything proves to be working properly.
- **Production**. This instance is used to run production services that will be accessed by the end users. For this reason, it is the most stable and closely monitored environment. Registration and Modification of services in the Federation Registry not only requires a technical verification from the Check-in team, but also a more exhaustive inspection to ensure that all regulations are complied with and to check that everything will be correct for the end user. For example, the service policies (such as the Privacy Policy and the Acceptable Use Policy of the service and others) will be exposed by Check-in to the end user, so their contents are reviewed.

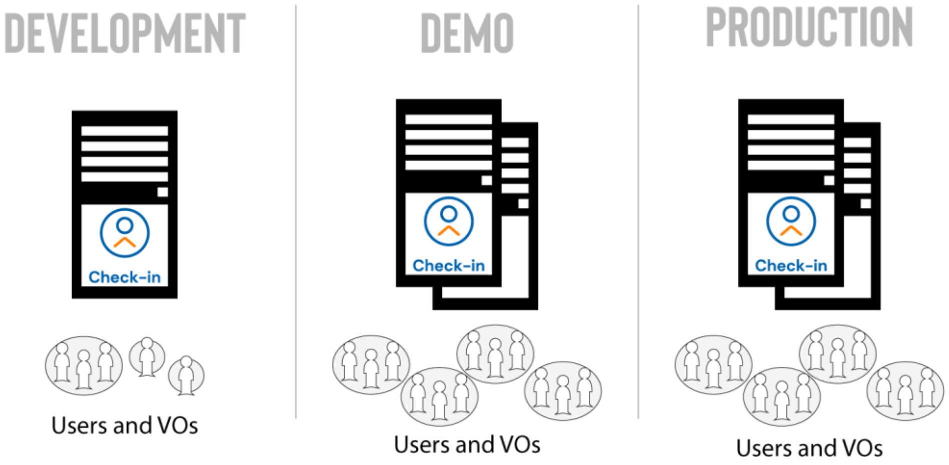


Figure 8: The three Check-in environments - Development, Demo and Production

As represented in Figure 8, the three environments are completely isolated from each other, so each one has its own VO configuration.

3.4 EOSC AAI and Check-in

EOSC runs on a Federated environment of systems and services. The purpose of the Authentication and Authorisation Infrastructure in EOSC is to support the FAIR principles for data and services while enabling high-trust collaborations to be established and maintained with little or no friction to the end user. AARC (Authentication and Authorisation for Research and Collaboration) is an initiative that addresses federated access through authentication and authorisation mechanisms in research and e-infrastructures, being the EOSC AAI architecture based on the **AARC Blueprint Architecture**. Figure 9 illustrates this architecture, showing the data flow from top to bottom: from the moment the user is authenticated by an Identity Provider, releasing user attributes for authentication and authorisation, to the access to the end services offered by an SP.

Implementing and configuring an AAI service compatible with EOSC AAI (and, herein, with AARC BPA) is complex. However, this is facilitated by EGI Check-in, which acts as an AARC Proxy service that connects federated Identity Providers with Service Providers. **EGI Check-in service is a registered member of the EOSC AAI Federation** and this means that all the services integrated with it will be accessible to all EOSC users in an interoperable and trusted manner: it will allow users to authenticate with a large variety of idPs, providing a simple, integrated method to ensure EOSC users the use of services according to their defined access policies.

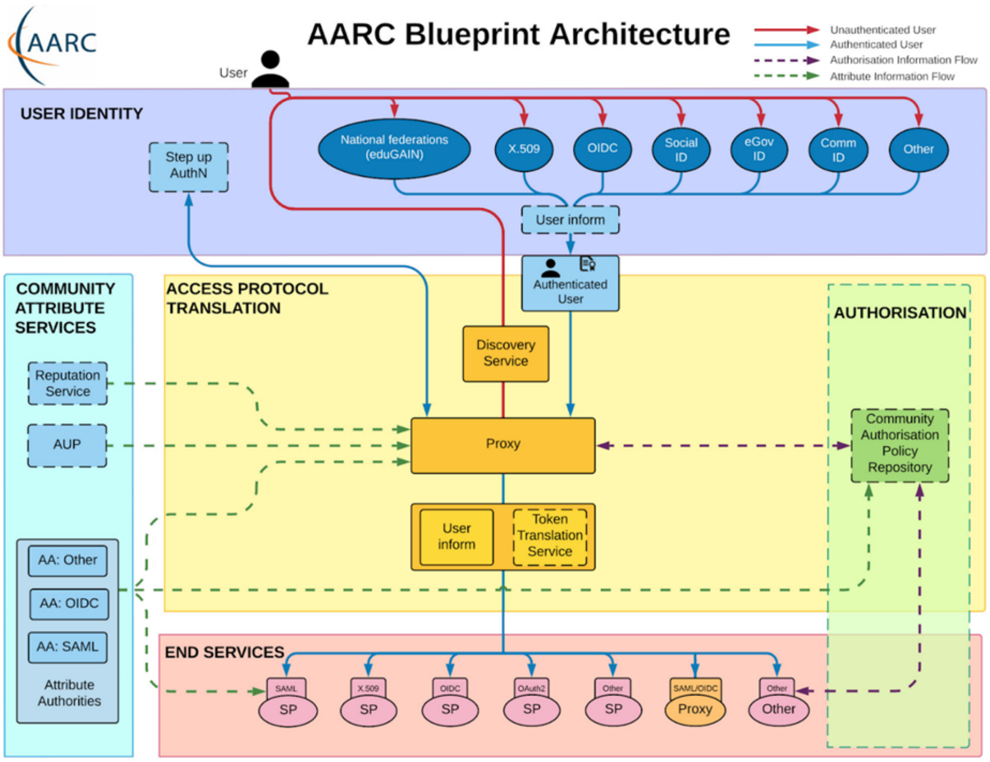


Figure 9: AARC Blueprint Architecture

4. AAI IN EUREKA3D

The infrastructure created in EUreka3D uses EGI Check-in for the implementation of its AAI, mainly on these two aspects:

- **Authentication of users** to access applications, such as Content Providers that upload and share 3D models in DataHub or the cloud administrators that create virtual servers for the project. This is discussed in Section 4.1.
- **Organisation of the EUreka3D community**, to assign permissions to the different groups of Content Providers and other users. This is discussed in Section 4.2.

The architecture of the EUreka3D infrastructure is represented in Figure 10, which shows a lock where Check-in acts and protects the access to the different components. The *Data Management* is normally accessed by Content Providers, which are the creators of the Cultural Heritage data, whereas the *Compute Power* is used by the cloud operators of the project that develop and install applications.

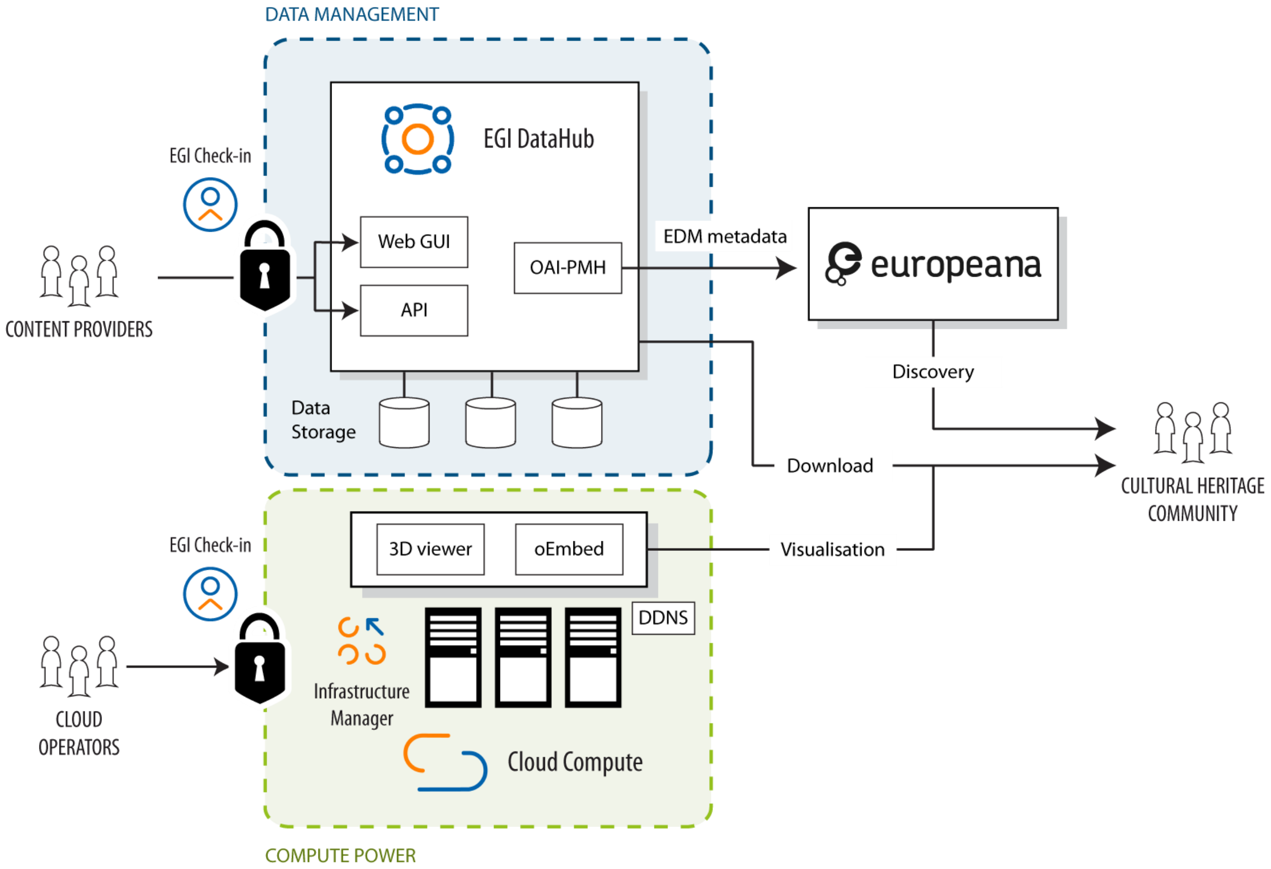


Figure 10: Overview of the architecture of the EUreka3D infrastructure

For simplicity, remote access to the servers is configured via SSH keys, which is a standard way to access remote servers. This is discussed in Section 4.3.

4.1 Authentication of users

Different users are involved in EUreka3D, which can be categorised into users that manage and develop EUreka3D, such as the cloud operators, and users that use EUreka3D services, typically Content Providers. From the authentication point of view, both categories are equal, and they can use any of the wide variety of idPs available in Check-in (see Section 3.2.1 for a description of idP). In general terms, the members of EUreka3D use these idPs:

- Their Home Organisation idP, if it is part of **eduGAIN**. This is the case of Cyprus University of Technology or the members of CNRS supporting Bibracte.
- **Social accounts**, such as ORCID, LinkedIn and Google, for the users whose organisations are not part of eduGAIN. This is common for small CHIs and museums.
- In less common cases, the **EGI idP** is also used when not covered by any of the previous cases.

4.2 Authorisation of users

Authorisation of users is based on the *entitlements* of a user (in simplified terms, the list of permissions that the user has), especially on the membership of the user in the EUreka3D Virtual Organisation and its groups.

4.2.1 Organisation of the EUreka3D community

EUreka3D uses the *culturalheritage.vo.egi.eu* VO to support its activities. Implicitly, the VO is organised into two broad categories:

- **Infrastructure**, for developers and other technical personnel working in operational tasks.
- **Working groups**, for the different Content Providers and special users.

This is depicted in Figure 11.

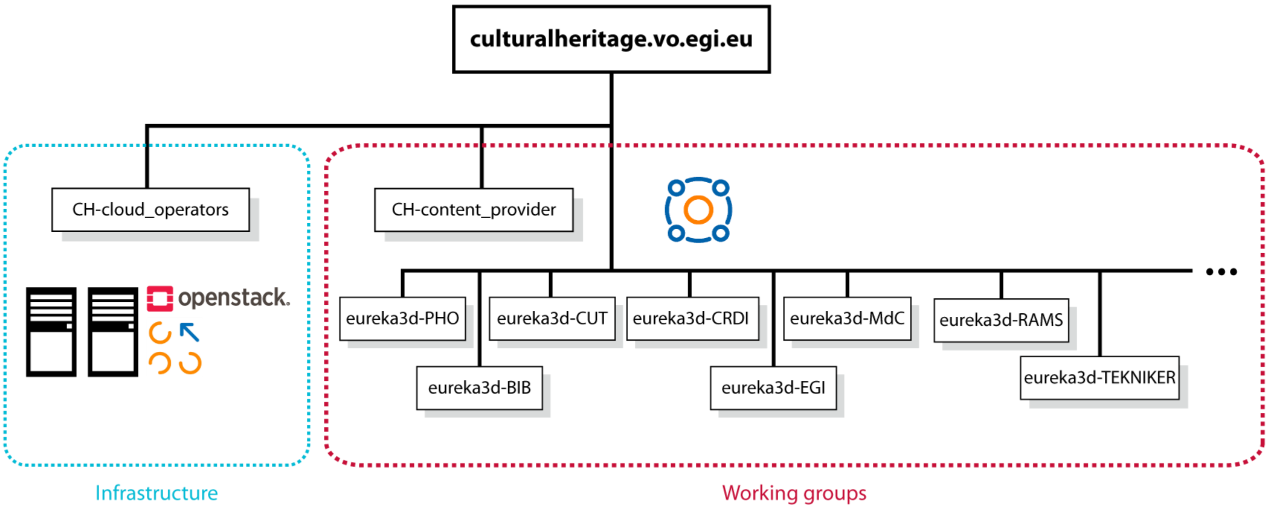


Figure 11: Structure of the EUreka3D Virtual Organisation

The Infrastructure category is represented by the group *CH-cloud_operators*, which are allowed to access OpenStack and the underlying cloud technology. The Working Groups category represents the groups that enable the side of the CH tasks. The first group is *CH-content_provider*, which represents the permissions that every Content Provider (the providers of CH data) should have. Every user that will upload content to the DataHub must be a member of this group. Next, there is a group in Check-in for each Content Provider institution. These groups are prefixed with "eureka3d-" and allow users to be grouped into their respective organisations. This grouping has two purposes:

1. It enables sharing of data files and directories between the EUreka3D community. Any Content Provider can share data with other Content Providers and allow them not only to visualise but to work cooperatively on some model.
2. It enables splitting into sets for data publishing (the *setSpec* in OAI-PMH), so that items from the same Content Provider are grouped together for selective harvesting. This is realised at the moment the user selects the account to retrieve a PID when publishing an object.

The way the Check-in groups (Figure 11) are related to the general architecture (Figure 10) can be seen in Figure 12.

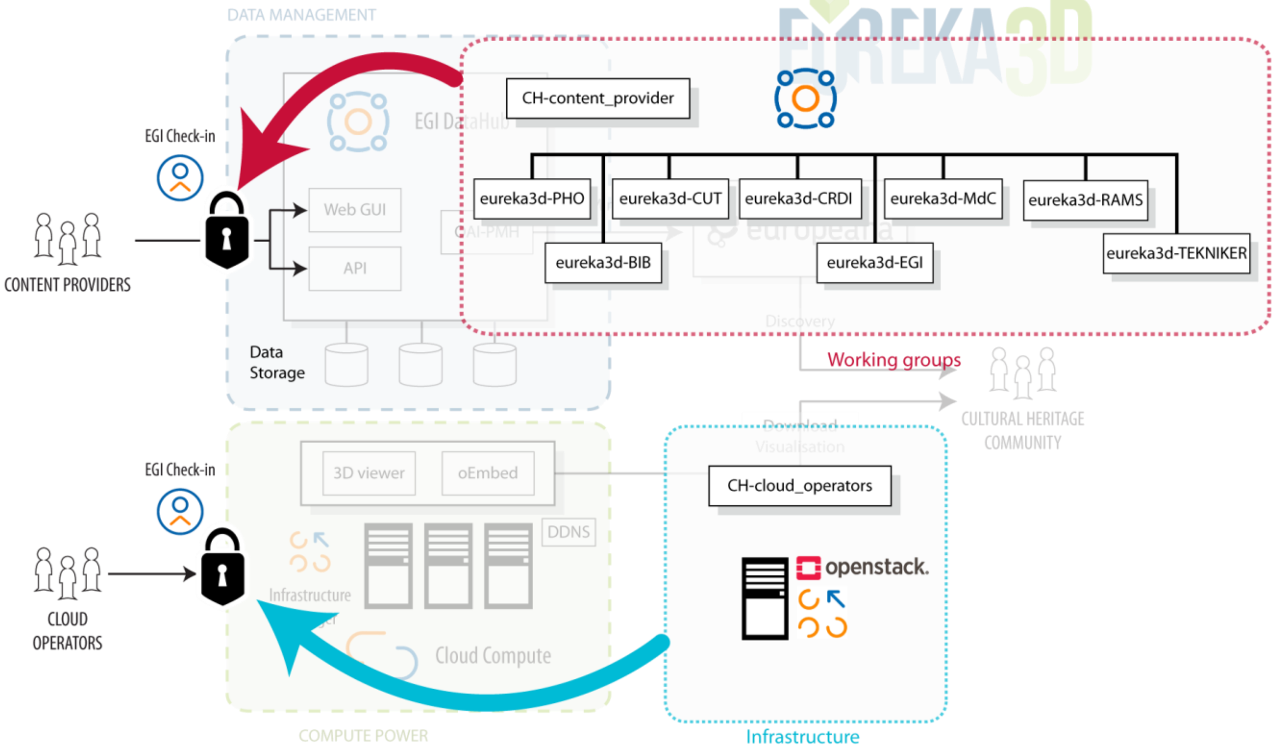


Figure 12: How the different Check-in groups are related to the general architecture

4.2.2 Joining the EUreka3D community

In order to join the EUreka3D community and be assigned a group, it is mandatory to be a member of the EUreka3D VO. The process is standard and documented in Check-in's documentation[11], consisting of two steps:

1. Registering a Check-in account (Figure 13)[12]. The process is fairly simple and the user only has to visit a URL (https://aai.egi.eu/signup) and proceed with the login process of the identity provider that wants to associate to the Check-in account. The example in the figure shows the University of Oxford as idP, but it is similar with other idPs. In case that the idP does not provide sufficient information to register the Check-in account, Check-in will ask the user for the missing information.

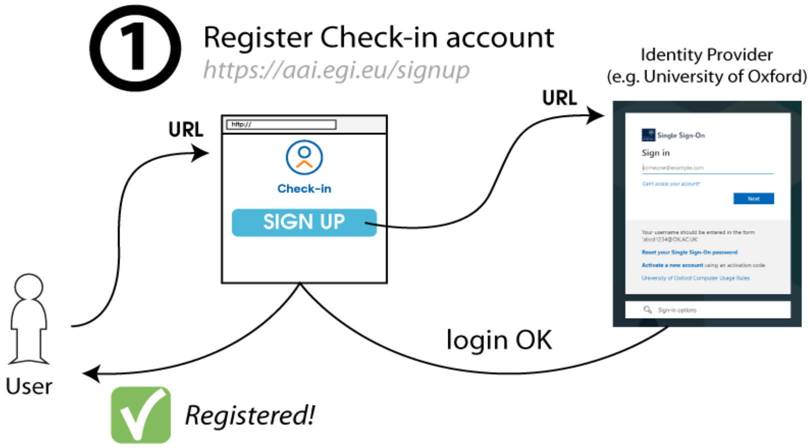

① Register Check-in account
https://aai.egi.eu/signup
Identity Provider
(e.g. University of Oxford)
URL
URL
http://
Check-in
SIGN UP
Single Sign-On
Sign in
someone@example.com
Can't access your account?
Next
Your username should be entered in the form 'abcd1234@OX.AC.UK'
Reset your Single Sign-On password
Activate a new account using an activation code
University of Oxford Computer Usage Rules
Sign-in options
User
login OK
Registered!

Figure 13: Flow showing a Check-in account registration

2. Joining the EUreka3D Virtual Organisation (Figure 14). Once the user has registered a Check-in account, he/she has to visit a URL to request membership in the EUreka3D VO (https://aai.egi.eu/registry/co_petitions/start/coef:632). The user is redirected to a Web page and, after acceptance of the Privacy Policy and the Acceptable Use Policy, the request is submitted. This will generate a notification email for the administrators of the VO (the VO Managers of EUreka3D), who can approve or reject the membership request. On approval, the user becomes a member of the EUreka3D VO and is notified by email. Once the user has been accepted in the VO, the right permissions have to be assigned by the VO Managers. Refer to *Annex A* for more information about this process. Policies and terms of use for Check-in and DataHub services are available for consultation online[13].

[11] https://docs.egi.eu/users/aai/check-in/joining-virtual-organisation/

[12] https://docs.egi.eu/users/aai/check-in/signup/

[13] [3] https://datahub.egi.eu/ozw/onezone/i#/public/privacy-policy, https://datahub.egi.eu/ozw/onezone/i#/public/terms-of-use, https://aai.egi.eu/proxy/module.php/themeegi/views.php?id=privacy

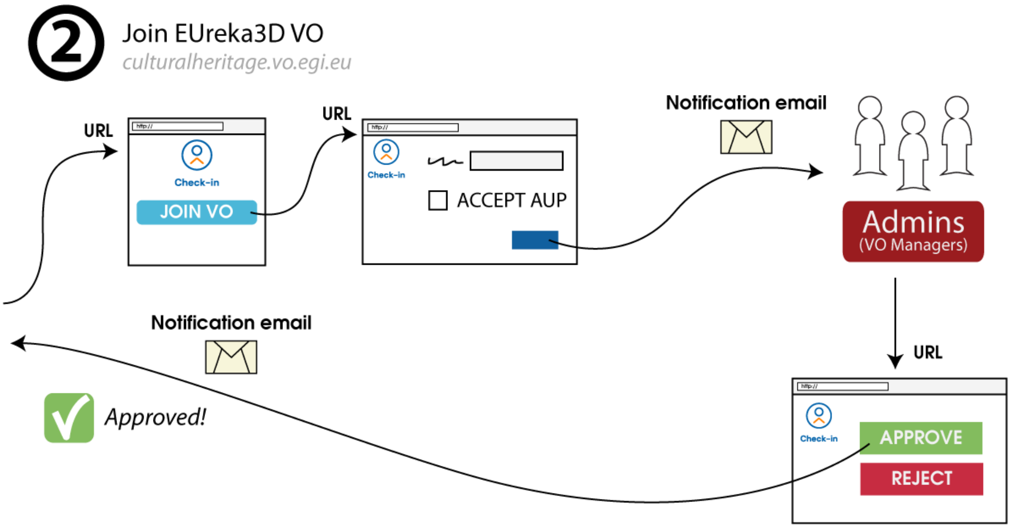
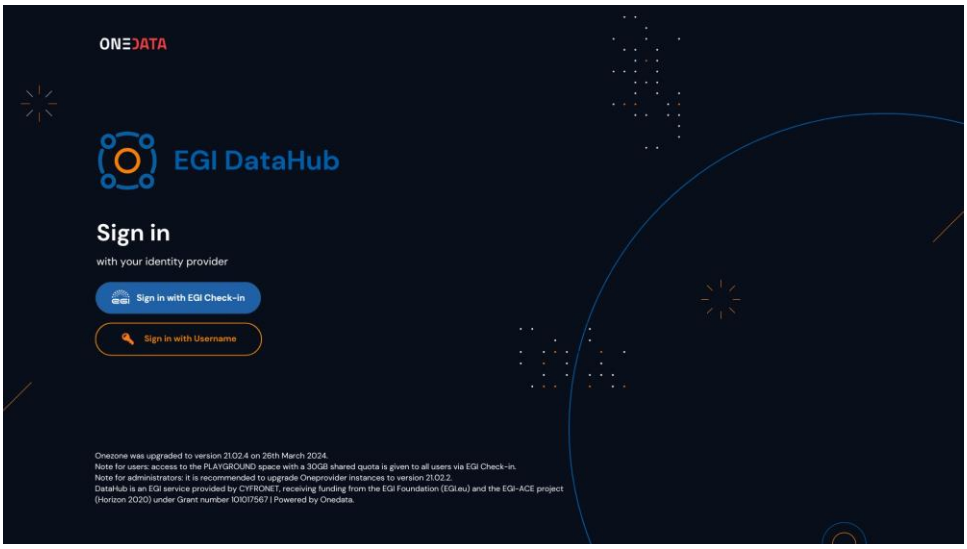


Figure 14: Flow showing the EUreka3D VO joining process

These two steps are also documented in the *Content Provider Handbook*, which is an informal document that has been made available in EUreka3D as a user manual for Content Providers.

4.3 Access in DataHub

DataHub is integrated with Check-in and is able to use it for both authentication and authorisation. When an unauthenticated user accesses DataHub, the login screen is presented, and the user can use Check-in to authenticate, as depicted in Figure 15.

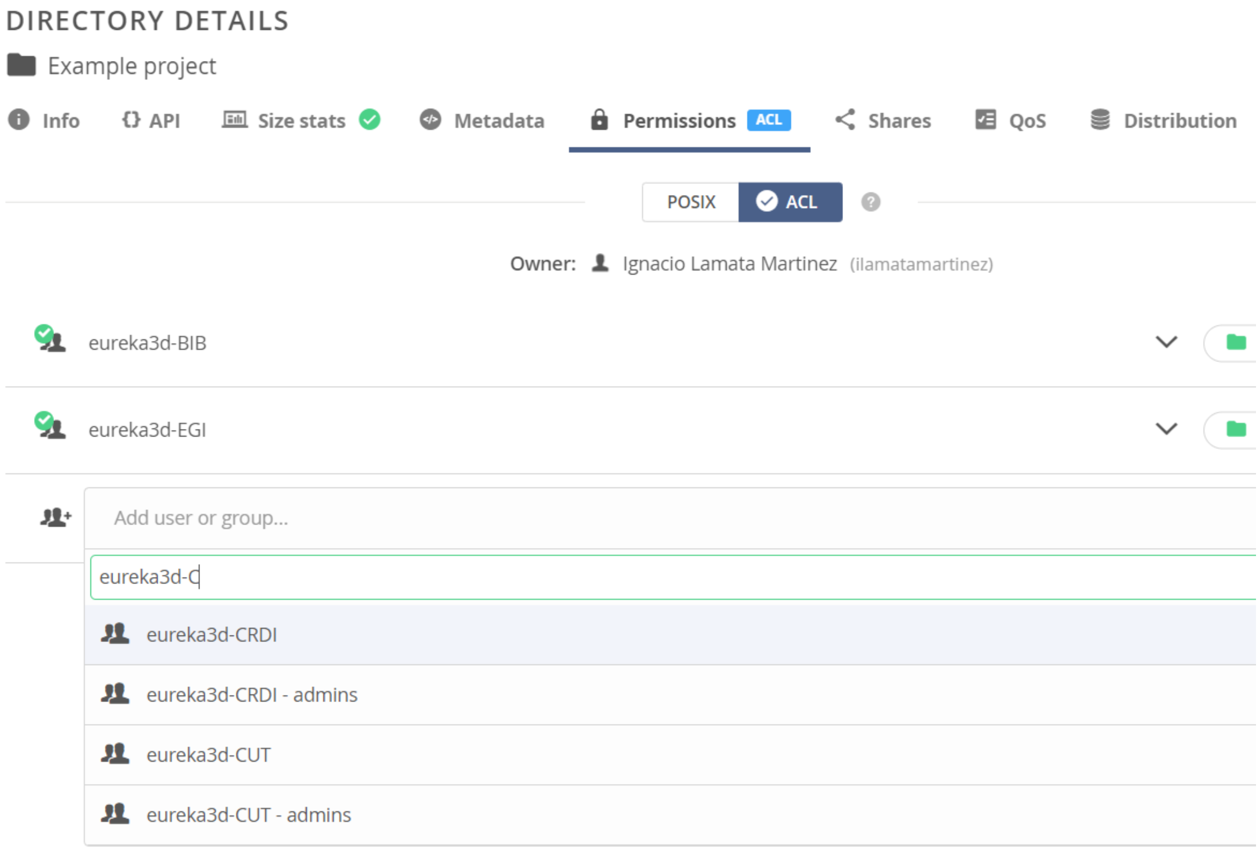


Onezone was upgraded to version 21.02.4 on 26th March 2024.
Note for users: access to the PLAYGROUND space with a 30GB shared quota is given to all users via EGI Check-in.
Note for administrators: it is recommended to upgrade Oneprovider instances to version 21.02.2.
DataHub is an EGI service provided by CYFRONET, receiving funding from the EGI Foundation (EGI.eu) and the EGI-ACE project (Horizon 2020) under Grant number 101017567 | Powered by Onedata.

Figure 15: EGI DataHub login screen, with Check-in as a login option

If the authentication has been completed successfully, DataHub obtains the user's attributes and permissions from Check-in. These include the user *entitlements*, which specify the groups to which the user belongs.

Files and directories in DataHub can be protected with ACLs[14], which specifies the list of groups that can have access to the data and under which conditions. The different Check-in groups depicted in Figure 11 are available in DataHub to be used in ACLs. Figure 16 shows an example of the creation of an ACL in which different EUreka3D groups can be added. The specific permissions that can be assigned to a particular group in the ACL include the possibility to add files and subdirectories, traverse directories, delete files, read and write metadata and attributes and manage the ACL itself.



DIRECTORY DETAILS

Example project

Owner: Ignacio Lamata Martinez (ilamatamartinez)

eureka3d-BIB

eureka3d-EGI

Add user or group...

eureka3d-C

eureka3d-CRDI

eureka3d-CRDI - admins

eureka3d-CUT

eureka3d-CUT - admins

Figure 16: Permission assignment to a specific group in DataHub

Additionally, DataHub provides a testing environment to test the next features that will be released in production. This testing environment uses the *Demo* environment of Check-in, as described in Section 3.3, which means that there is a specific VO with the same name that is used for testing. In this VO, only users that need to test the new features of DataHub have been added.

4.4 Access to the Cloud Compute

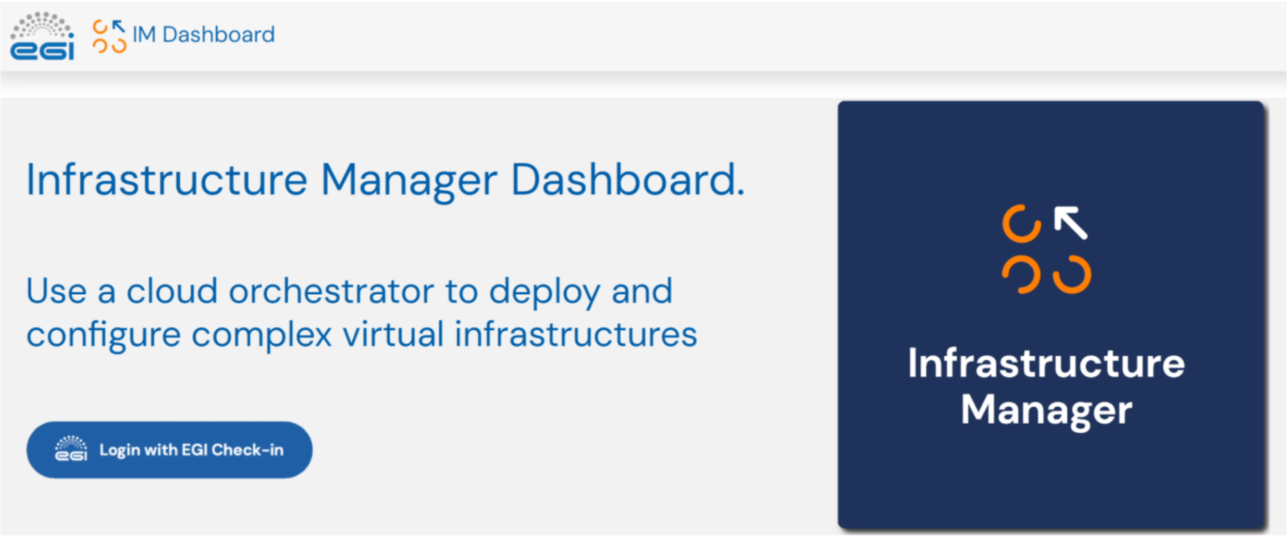
The management of virtual servers is done through *OpenStack*[15] (refer to Deliverable 3.1 and 3.3 for an explanation of cloud technology), and these systems in the EGI Federation are integrated with Check-in.

[14] Access Control List

[15] https://www.openstack.org/

Similarly to DataHub, OpenStack can use Check-in to authenticate and authorise users. As discussed before, users managing infrastructure must belong to the group “*CH-cloud_operators*” of the EUreka3D VO.

The Infrastructure Manager (see Figure 17) is another service that is used in EUreka3D to deploy virtual servers and install applications, and it is equally integrated with Check-in.



Infrastructure Manager Dashboard.

Use a cloud orchestrator to deploy and configure complex virtual infrastructures

Figure 17: Login screen of the Infrastructure Manager, showing the Check-in login button

The actual access to the virtual servers is not currently managed by Check-in, but done directly through SSH, which is a network protocol to access a remote machine securely. The access is configured by installing the public keys of the developers and operators in the servers. Without the private keys that are linked to the public keys inserted in the servers, access is denied.

However, there have been recent developments to provide SSH access via OIDC, and the use of SSH access via OIDC will be explored in the coming projects e.g. EUreka3D-XR. This means that SSH will still be used, but users will be able to use Check-in to access the virtual servers without needing to manage and share SSH keys explicitly. The feature is available through *motley clue*[16], which can be automatically installed by the Infrastructure Manager at the time the server is deployed in the cloud.

[16] Mapper Oidc To Local idEntitY with loCal User managEment. https://motley-cue.readthedocs.io/en/latest/

5. CONCLUSIONS

This document has described the Authentication and Authorisation Infrastructure (AAI) configured for the EUreka3D infrastructure. This infrastructure protects the Content Provider's data from both unauthorised access and manipulation, and allows for selective and flexible decisions on what data to share with whom. In this way, Content Providers can store and share their content securely (as is the case in the context of the EUreka3D project) and be prepared for collaborative actions with other partner institutions and communities (for example, co-creating narratives and XR experiences reusing 3D data, as will be the case in the follow-on EUreka3D-XR project starting in February 2025). The document first defined some basic concepts required to understand the terminology of the rest of the document, mainly the mechanisms of authentication (identifying a user), authorisation (verification of the user's permissions in a system) and Single-Sign On (which allows users to sign in once and remain authenticated in multiple systems). Additionally, some of the common protocols for authentication and authorisation have been briefly described.

The implementation of the AAI in EUreka3D has been done through EGI Check-in, so it has been necessary to provide details about this service. Check-in is an identity and access management system and is widely used in the EGI Federation, so it is prepared to operate in a federated environment and overcome the main challenges encountered in research and academia. A federated system involves some components, such as identity providers and service providers, which have been explained in the text. The concept of Virtual Organisation, one of the main mechanisms in the EGI Federation to organise users and enable authorisation, has also been explained. The description of Check-in has covered the necessary information in the scope of EUreka3D, so its relation within the EOSC AAI, the benefits for GDPR compliance and the different environments used for testing and production have been discussed.

Once the basic AAI concepts are clear and EGI Check-in has been explained, it is described how these concepts fit into the actual implementation of the AAI adopted in the EUreka3D project. The EUreka3D infrastructure interacts with two main different categories of users: producers and consumers of data. Check-in protects the data management system so that only Content Providers can upload data. It is equally important that the different Cultural Heritage Institutions (CHIs) can control their data, and share it with other CHIs in a secure manner. In this sense, the EUreka3D community has been modelled in different groups, one per each CHI using the system.

EUreka3D uses two main services for its operations, EGI DataHub and EGI Cloud Compute, and they are both integrated with EGI Check-in. The way they are used in terms of AAI has been described in the document.

REFERENCES

[1] European Commission, Directorate-General for Research and Innovation, Wierenga K., Johansson L., Kanellopoulos C., Groep D., Vaghetti D., Liampotis N. (2021) *EOSC Authentication and Authorization Infrastructure (AAI): report from the EOSC Executive Board Working Group (WG) Architecture AAI Task Force (TF)*, Publications Office. https://data.europa.eu/doi/10.2777/8702

[2] Brocke L. (2023) Certificate-based OpenSSH for Federated Identities. Master's Thesis. DOI: 10.5445/IR/1000165236. Accessible at https://publikationen.bibliothek.kit.edu/1000165236 (date last accessed: October 2024).

ANNEX A: USER ONBOARDING PROCEDURE

EUreka3D

User Onboarding procedure

Intended for VO Managers.

Follow these steps sequentially.

Only for new institutions that require a new group

1- Request the creation of the group in Check-in:

- Decide name in this style: eureka3d-ACRONYM (e.g. eureka3d-CUT)
- Send email to checkin-support@mailman.egi.eu, adding other VO Managers in CC. Template [1].

2- Once created, assign this group to ourselves in Check-in and logout + login in DataHub.

- Make sure that the group appears in the option“Groups” of the left menu. Otherwise, wait for a few minutes and logout+login again in **DataHub**.

3- Request the configuration of the group in DataHub:

- Send email to lukasz.opiola@cyfronet.pl, adding other VO Managers in CC. Template [2].

4- Once configured, remove the group membership from us in Check-in.

5- Share the Content Provider Handbook with the user if it has not been shared already. The user should **send a request to join the VO**, as explained in the Handbook.

6- **Approve** or reject the VO membership request in Check-in (for example, by following the link coming from the email).

7- **Assign** both the “CH-content_provider” and the corresponding "eureka3d-..." groups to the user in **Check-in**.

8- **Send email to the user**, adding other VO Managers in CC, to inform them that their access has been granted. Template [3].

DONE.

- -

[1]
Dear Check-in team,

For the VO "culturalheritage.vo.egi.eu"
Could you please create a group: eureka3d-_____

Many thanks,

- -

[2]
Please configure group: eureka3d-_____
In the EUreka3D space.

Thanks.

- -

[3]
Dear _______,

We are pleased to welcome you to the EUreka3D community!

Your access to the EUreka3D data hub has been successfully configured. You should be able to follow the steps in the Content Provider Handbook [1] to upload content, share it with the rest of the community and publish it in Europeana.

You are part of the “eureka3d-_____” group, which can be used to restrict access to data to members of your group only, and to share content with other groups in EUreka3D.

Should you have any questions, please contact us.

Best regards,

The EUreka3D team.

[1] https://go.egi.eu/eureka3d_handbook

- -