EOSC-IF

ARIA Data Access Management (ADAM) Interoperability Guideline

EOSC-IF / ARIA Data Access Management (ADAM) Interoperability Guideline

Lead by **CESSDA ERIC**
Authored by Kostas Papagiannopoulos (CESSDA/EKKE), Joshua Tetteh Ocansey (CESSDA), Carsten Thiel (CESSDA), John Shepherdson (CESSDA), Marcus Povey (Instruct-ERIC), Alec Mathews (Instruct-ERIC), Marcus Lowndes (Instruct-ERIC)
Reviewed by Jorik van Kemenade (SURF), Gavin Farrell (ELIXIR Europe)

Dissemination Level of the Document

Public

Abstract

These Interoperability Guidelines describe the ARIA Data Access Management (ADAM) solution developed by Instruct-ERIC and CESSDA ERIC as part of the EOSC Future project. The solution enables data repositories to use ARIA for processing of access request applications from researchers. This can be integrated into Social Science and Humanities study level metadata catalogues, so that user access becomes seamless.

Version History

Version	Date	Authors/Contributors	Description
V0.1	27.02.2023	Kostas Papagiannopoulos (CESSDA/EKKE), Joshua Tetteh Ocansey (CESSDA), Carsten Thiel (CESSDA), Marcus Povey (Instruct-ERIC), Alec Mathews (Instruct-ERIC), Marcus Lowndes (Instruct-ERIC)	First draft
V0.2	02.03.2023	Anke Friedrich (GEANT)	Formatting of first draft
V0.3	30.03.2023	Joshua Tetteh Ocansey (CESSDA), John Shepherdson (CESSDA), Marcus Povey (Instruct)	Updates based on reviewer's comments
V0.4	04.04.2023	John Shepherdson (CESSDA)	Further updates, based on additional review comments
V1.0	11.05.2023	Kostas Papagiannopoulos (CESSDA/EKKE), Joshua Tetteh Ocansey (CESSDA), Carsten Thiel (CESSDA), Marcus Povey (Instruct-ERIC), Alec Mathews (Instruct-ERIC), Marcus Lowndes (Instruct-ERIC), Ron Dekker (TGB), Mike Chatzopoulos (ATHENA)	Final Version

Copyright Notice

Table of Contents

Table of Tables

Table of Figures

1.   Glossary

EOSC Future project Glossary is incorporated by reference: https://wiki.eoscfuture.eu/x/JQCK

2. List of Abbreviations

Acronym	Definition
AAI	Authentication and Authorization Infrastructure
ADAM	ARIA Data Access Management
ARIA	Access to Research Infrastructure Administration. A collection of cloud services provided by Instruct to research infrastructures, facilities and user communities within a range of scientific fields
CDC	CESSDA Data Catalogue
CESSDA	Consortium of Social Sciences Data Archives
DDI	Data Documentation Initiative
EOSC	European Open Science Cloud
Instruct	Integrated Structural Biology/Instruct-ERIC
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
RI	Research Infrastructure
So.Da.Net	The Greek Research Infrastructure for Social Sciences
SSH	Social Science and Humanities

1 Intended Audience

This interoperability guideline is intended for adoption by data repositories that maintain catalogues of Social Science and Humanities (SSH) study level metadata and want to make use of the Access to Research Infrastructure Administration (ARIA) Access Management Platform for managing user requests for access to any restricted data described by the metadata. The guideline describes how repository maintainers register their metadata in ARIA [6], then model the corresponding access request forms and processes in ARIA and finally add the correct references to their original metadata holdings for the users to find.

2 Description and main features

The ARIA Data Access Management Platform provided by the Integrated Structural Biology (Instruct) ERIC offers institutions the ability to manage all access requests to their physical facilities and data. Institutions can import their holdings from their existing Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) endpoint, map their access processes and workflows to ARIA and allow them to model the entire access application process with ARIA. Making use of ARIA's workflow modelling engine, researchers can be presented with tailored application forms, specific to the institution's use case.

By modelling the process in ARIA, researchers get a consistent experience for access applications at any participating institution. The institution does not need any additional workflow management software, (e.g. a ticketing system), as the entire process is managed within ARIA.

Supported models include:

- Verification as researcher
- Verification of successful completion of required training courses or certification
- Verification of successful completion of required documents, such as e.g.
    - Terms-of-Use
    - Licence Agreement
    - Non-Disclosure Agreements
- Submission of a full research proposal for formal review with fields required by the data owner or data processor

3 Response to Community Need

Managing user access requests can be particularly challenging for growing organisations that do not (yet) have solutions in place to manage such requests. At the same time, existing solutions might not be easily adaptable to changing needs and requirements of the institution. In contrast to adoption of generic workflow solutions, such as ticketing systems, the use of the ARIA platform enables repository providers to outsource the work with all of its parts into a single platform, where the interactions with researchers and reviewers are all realised in a single place. ARIA has been built from the ground up to manage access to scientific resources (initially physical, but adapted through this project to manage virtual resources as well). It has an integrated review process, reporting and feedback collection, etc. ARIA is "generic" in as far as it can be customised for specific applications - e.g. in the Consortium of Social Sciences Data Archives (CESSDA) use case a white labelled version could be created, with appropriate branding and catalogue, but utilising the same platform under the hood.

An additional – optional – feature is the use of central user management through the European Open Science Cloud (EOSC) Authentication and Authorization Infrastructure (AAI), in particular LifeScience ID [8], for managing user identities directly in ARIA.

4 High-level Service Architecture

To use the ARIA Data Access Management Platform, metadata must be available from a repository that has an OAI-PMH endpoint. This endpoint must be registered in ARIA, so that metadata records can be harvested into ARIA. Then, the repository maintainer can create the access application workflows in ARIA.

Once the workflows are defined, the respective identifiers must be added to the original metadata at the repository by the maintainer. Then the link to the application form can be displayed in the repository and any other aggregating metadata catalogue, such as the CESSDA Data Catalogue (CDC) [1] or the OpenAIRE Explore catalogue [4].

Figure 1 is a high level representation of the overall workflow of the beta implementation. As a precursor, metadata (describing restricted data) from the So.Da.Net Data Catalogue is registered in ARIA, the access application workflows are defined in ARIA and the application links for users are included in the original catalogue and then harvested into CDC.
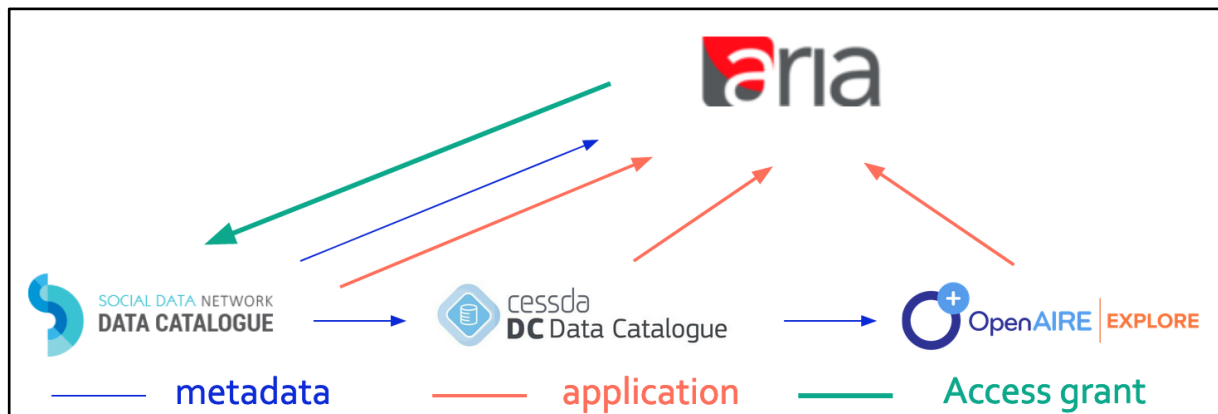
Figure 1: Pilot implementation in EOSC Future

The benefits from the User's perspective are twofold: a) a paper-based access application process is replaced with an easy to follow workflow in ARIA (see figure 2); b) all the user has to do is click on a button in the study level metadata labelled 'apply for access to data' or similar; the access process is the same, whether initiated from the So.DaNet Data Catalogue or CDC.
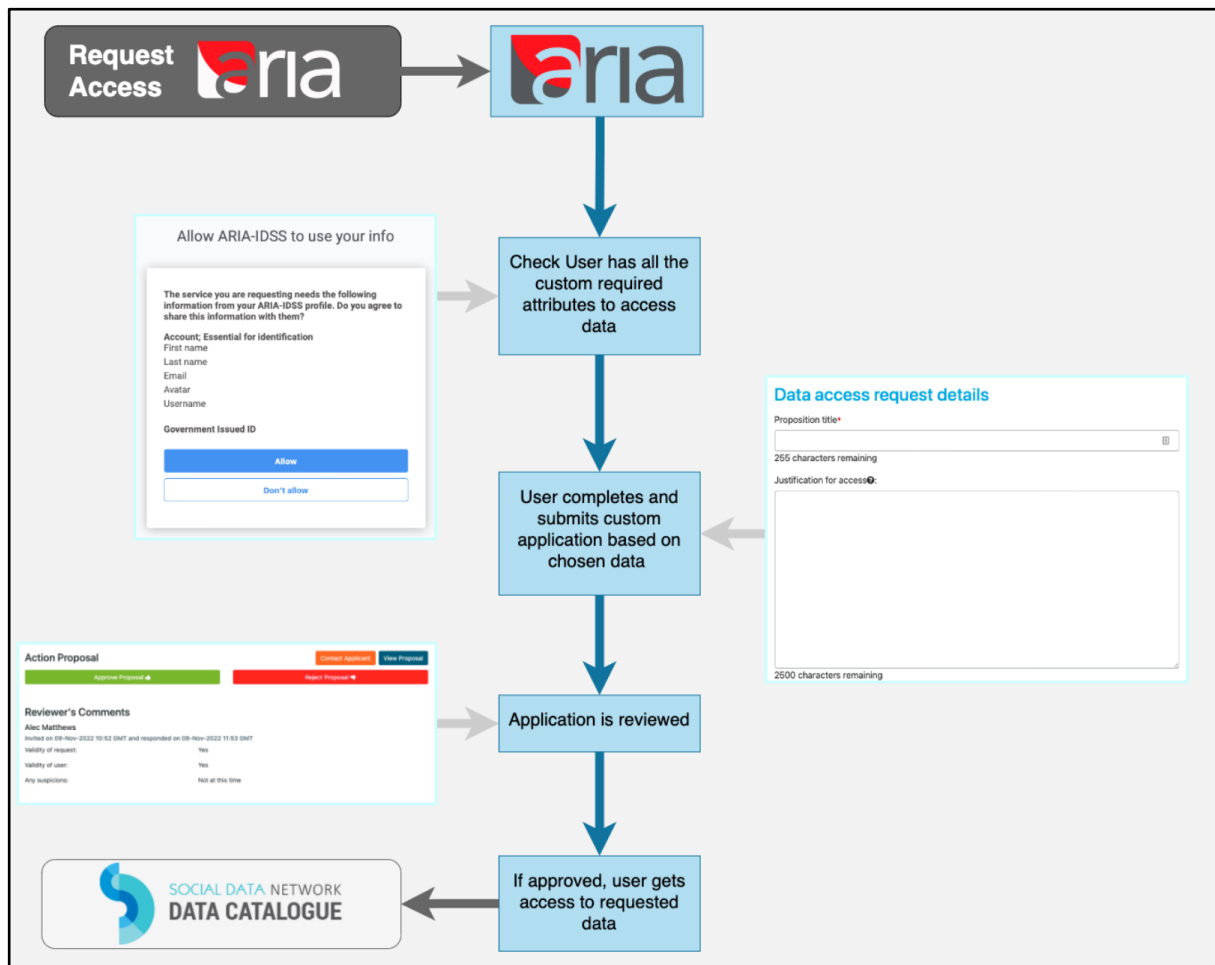
Figure 2: User workflow realised in ARIA

5 Definitions

The terms Institution, Data Provider, Catalogue Provider, Repository Provider and Organisation mean the same, i.e. the originator of the study level metadata.

6 Licensing Information

Instruct-ERIC offers ARIA to scientific projects and Research Infrastructures (subject to fees), but this is done as an agreement between the individual Research Infrastructure/Project and Instruct-ERIC (Research Infrastructures are facilities that provide resources and services for research communities to conduct research and foster innovation, see [13] for a list of RIs). This is subject to timelines and Instruct's capacity at the time. ARIA isn't a commercial platform that anyone can sign up to with a credit card. See [9] for more information about accessing ARIA.

7 Related Guidelines

Table 7-1: Related Guidelines

Resource Type	Title	Short Description	relatedIdentifier
Service	OpenAIRE Explore	A comprehensive and open dataset of research information	https://explore.openaire.eu/
Service	ARIA	ARIA is a cloud platform for Access and Facility	https://aria.services/

| | | management build by Instruct-ERIC | |

8 Adopted Standards

Table 8-1: Adopted Standards

Resource Type	Title	Short Description	relatedIdentifier
Standard	OAI-PMH – Open Archives Initiative Protocol for Metadata Harvesting (version 2.0)	Protocol used by data catalogues to describe their holdings	https://www.openarchives.org/OAI/openarchivesprotocol.html
Standard	JSON – JavaScript Object Notation & GraphQL (version ISO/IEC 21778:2017)	Used in the ARIA API for returning dataset applications	https://www.iso.org/standard/71616.html
Standard	Dublin Core Metadata standard (version 1.1)	Most common metadata standard in general use	https://www.dublincore.org/specifications/dublin-core/dcmi-terms
Standard	REST API - Representational state transfer (no version information)	A Web API conforming to the REST architectural style	https://restfulapi.net/

9 Integration Options

EOSC AAI [10] - for central user management, in particular LifeScience ID can be used for managing user identities directly in ARIA.

EOSC Helpdesk [11] - add ticket creation form in ARIA User Interface to create issues in a dedicated queue.

EOSC Monitoring [12] - can be used to periodically check availability of ARIA platform and provide uptime reports. Dependency on onboarding ARIA to EOSC Catalogue and Marketplace.

10 Interoperability Guidelines

In order to integrate with the ADAM System, Catalogues with OAI-PMH endpoints that contain restricted datasets must be available for harvesting and registered by ARIA. ARIA generates an access link that can be used to establish a link on the Catalogue page, which will drive users to the ARIA platform to complete the application process.

Repository providers can outsource their access mechanism to the ARIA platform as a single platform. It gives data repositories a seamless flow to allow access to their restricted dataset.

In order to integrate, the Catalogue provider has to provide the following set of parameters:

Attribute Name	Definition	Type	Required
Catalogue Name	Name of the catalogue resource	String	Mandatory
Catalogue URL	Web page of the resource	String	Optional
Providers Name	Name of organisation providing the resource	String	Mandatory
Catalogue OAI-PMH	OAI-PMH link to the resource	String	Mandatory
Contact Name	Contact person of the providers organisation	String	Mandatory
Contact Email	Email address of the contact person	String	Mandatory

Datasets of data catalogues that are to be integrated with the ARIA Data Access Management (ADAM) service are required to have restricted access elements, i.e., access to data should be provided after the application is approved by the supervisors of the data. Initially, an OAI–PMH endpoint, structured in DublinCore metadata format, must be provided by the data catalogue. Optionally, a *setspec* element can be provided to gather all the studies with restricted elements under one URL endpoint. This endpoint should provide ARIA with the required information of which the mapping process will take place, linking data information with the form creation that will take place.

Supervisors of the data catalogues should provide a form with required and optional fields that need to be completed when an individual user has requested access to restricted resources, see Section 11.2 for an example implementation. Once the mapping takes place, the ARIA Service returns the application ID of the form via Rest API and JSON technology, the retrieval of which will have to form the URL that leads to the request access form in the ARIA website. This URL needs to be mapped in the study metadata schema, specifically held by the Data Documentation Initiative (DDI) field *specPerm*. If this information exists in a study, the button that holds this link will be implemented programmatically in the study page of the data catalogue repository and to every other harvest location that requires this service. This process will lead to an integrated service of which the management of requesting accesses by the users and the supervisors that will grant or decline access will take place in a single web application, provided by ARIA.

Whenever a user completes the application form, an email is sent to the user, confirming the request, the repository maintainer then reviews the application in ARIA. If the review is positively approved, the user gets an email message with a link from ARIA which directs them to the original repository that contains the data in question. If the review is negative, an email message is sent to the user with an explanation of why the request failed. The time taken to review such requests could be subject to a Service Level Agreement.

11 Examples of solutions implementing this specification

11.1   11.1 CESSDA DDI based approach

These are instructions for the use of the implementation that is being developed in the ARIA Science Project of EOSC Future, and the example is specific to the SSH domain.

Within CESSDA, SSH study-level metadata is harvested from the national Service Providers into CDC via OAI-PMH, using the DDI 2.5 Codebook format [2]. This includes the contents of the So.Da.Net Data Catalogue [5].

The metadata fields to use are defined in the CMM [3], Section 1.4.6, specifically the XML XPATH:

```
/codeBook/stdyDscr/dataAccs/useStmt/specPerm
```

Note that a human-readable summary of the restriction, for display in catalogues such as the CDC, is also provided, see CMM [3], Section 1.4.2, specifically the XML XPATH:

```
/codeBook/stdyDscr/dataAccs/useStmt/restrctn
```

CDC uses the same Access Rights vocabulary as OpenAIRE [7]. The name of the vocabulary is i*nfo:eu-repo-Access-Terms vocabulary*. Only the codes *restrictedAccess* and *openAccess*" are used by CESSDA. All data with access rights that fall within the description "*Free download. May require registration to the system and/or accepting the terms of use online to gain access to the data. No restriction on the type of use"* are mapped to *openAccess*. Any other restrictions on data that are not within that scope should be mapped to *restrictedAccess*. In DDI2.5 this is defined as the value of this XML XPATH:

```
/codeBook/stdyDscr/dataAccs/useStmt/conditions
```

and the vocabulary name is given by the value of this XML XPATH:

```
/codeBook/stdyDscr/dataAccs/useStmt/conditions/@elementVersion
```

11.2   11.2 Prototype implementation in EOSC Future

A prototype of the solution for the SSH domain has been implemented that connects the So.Da.Net Data Catalogue and CESSDA Data Catalogue to the ARIA Access Management solution. Figures 3 and 4 show the paper-based form that a researcher currently has to complete to apply for access to restricted resources. Figure 5 shows the prototype workflow in ARIA that will replace the paper-based approach.

Figure 3: Paper form in use by EKKE as of 2022, Page 1
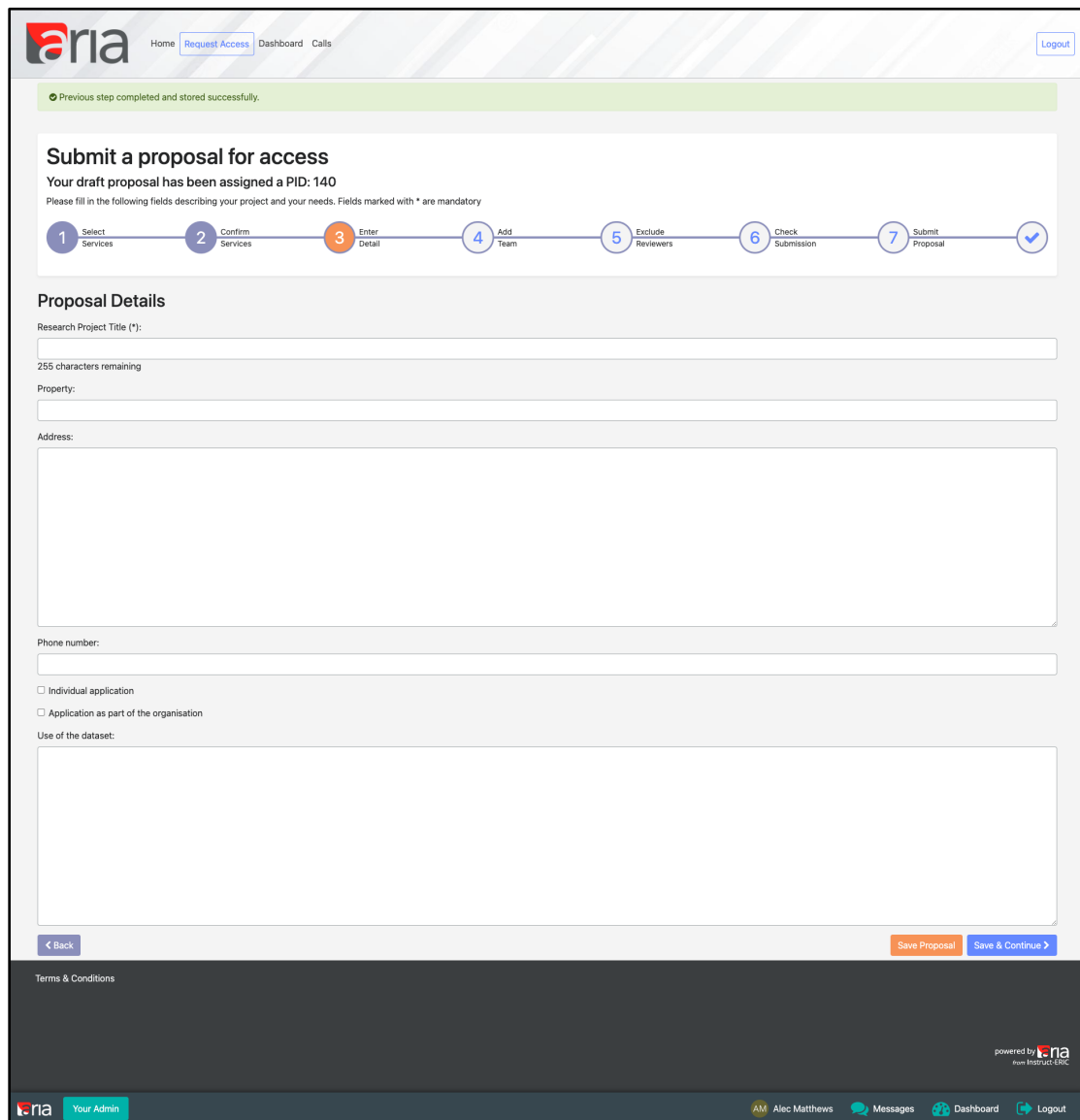
Figure 4: Paper form in use by EKKE as of 2022, Page 2

Figure 5: Realisation of corresponding application form in ARIA

11.3   11.3 Constructing a URL for ARIA

The following instructions describe how to use the current running implementation of ARIA, and the example is specific to the structural biology domain.

ARIA proposal creation supports GET and will be supporting POST requests in the near future. In order to start an application on a specific machine and services a user must:
- specify an access route
- specify a service/service type the machine belongs to
- send the machine variable as an associative array of *(platform_uri => array(machine_uri))* therefore the array, if provided in a GET URL, must be serialised.

When specifying the above, URI, external reference (for example DOI) and internal ARIA ID are all supported. Full documentation will be released in the near future.

Here is an example: I would like to automatically start a proposal on the 1) Instruct-ERIC site for a proposal funded by the access route 2) "Research Infrastructure Alpha" (acid = 91). This proposal will include a visit for the 3) "chromatography" service and a visit for 4) "electron microscopy" service, specifying sessions within this

ARIA Data Access Management (ADAM) Interoperability Guideline

electron microscopy visit on machines 5) "Fragment screening High-RES" and 6) "Fragment screening Low-RES".

3. References

[1] CESSDA Data Catalogue. 2023 [online] Available at https://datacatalogue.cessda.eu/

[2] CESSDA Data Catalogue: Providing Metadata. 2023 [online] Available at https://datacatalogue.cessda.eu/documentation/providing-oai-pmh.html

[3] Akdeniz, Esra & Moilanen, Katja. 2023. CMM CESSDA Metadata Model (3.0). [Zenodo] https://doi.org/10.5281/zenodo.7528240

[4] OpenAIRE Explore. 2023 [online] Available at https://explore.openaire.eu/

[5] Social Data Network Data Catalogue (SoDaNet). 2023 [online] Available at https://datacatalogue.sodanet.gr/

[6] ARIA. 2023 [online] https://aria.services

[7] Access Level. OpenAIRE Guidelines v3. 2023 [online] Available at https://guidelines.openaire.eu/en/latest/literature/field_accesslevel.html

[8] Life Science Login. 2023 [online] Available at https://lifescience-ri.eu/ls-login.html

[9] What is ARIA?. 2023 [online] Available at https://instruct-eric.org/help/about-aria

[10] EOSC AAI. 2023 [online] Available at TBC

[11] EOSC Helpdesk. 2023 [online] Available at https://wiki.eoscfuture.eu/display/PUBLIC/EOSC+Helpdesk+-+integration+factsheet

[12] EOSC Monitoring. 2023 [online] Available at https://wiki.eoscfuture.eu/display/PUBLIC/EOSC+Monitoring+-+integration+factsheet

[13] ESFRI Research Infrastructure table. 2021 [online] Available at https://roadmap2021.esfri.eu/projects-and-landmarks/view-the-table/