DELIVERABLE REPORT

WP8: JRA3 – Research on e-infrastructure for data and information management

D8.1
Design of Trusted authentication source for NFFA-EUROPE services

31 August 2016

M12

NFFA-Europe has received funding from the EU's H2020 framework programme for research and innovation under grant agreement n. 654360

nffa.eu

PROJECT DETAILS

PROJECT ACRONYM	PROJECT TITLE
NFFA-Europe	NANOSCIENCE FOUNDRIES AND FINE ANALYSIS - EUROPE
GRANT AGREEMENT NO:	**FUNDING SCHEME**
654360	RIA - Research and Innovation action
START DATE	
01/09/2015	

WP DETAILS

WORK PACKAGE ID	WORK PACKAGE TITLE
WP8	**WP8: JRA3 – Research on e-infrastructure for data and information management**

WORK PACKAGE LEADER

Stefano Cozzini (CNR/IOM)

DELIVERABLE DETAILS

DELIVERABLE ID	DELIVERABLE TITLE
D8.1	Design of Trusted authentication source for NFFA-EUROPE services

DELIVERABLE DESCRIPTION

Design of Trusted authentication source for NFFA-EUROPE services

EXPECTED DATE	ESTIMATED INDICATIVE PERSONMONTHS
M12 31/08/2016	6

AUTHOR(S)

Stefano Cozzini (CNR/IOM)

PERSON RESPONSIBLE FOR THE DELIVERABLE

Stefano Cozzini (CNR/IOM)

NATURE

Choose an item.

DISSEMINATION LEVEL

- ☒ P - Public
- ☐ PP - Restricted to other programme participants & EC: (Specify)
- ☐ RE - Restricted to a group (Specify)
- ☐ CO - Confidential, only for members of the consortium

REPORT DETAILS

ACTUAL SUBMISSION DATE
dd/mm/aaaa hh.mm AM

NUMBER OF PAGES
16 (right-click and select "update the field")

FOR MORE INFO PLEASE CONTACT
Stefano Cozzini (CNR/IOM) Tel. +39-040-3787508 Email: cozzini@iom.cnr.it

Version	Date	Author(s)	Description / Reason for modification	Status
0	09/06/2016	Stefano Cozzini	First draft	Draft
1	13/06/2016	S.Cozzini/R.Aversa	New revision	Draft
2	30/07/2016	S.Cozzini/T.Jejkal	Added use case section	Update
3	06/08/2016	S.Cozzini	Re-arrangements of section	Update
4	10/08/3016	S.Cozzini/R.Aversa	First Revision	Revision
5	30/08/2016	S.Cozzini/R.Aversa	Finalization	Final

Contents

1. Executive summary

This document presents the guideline of the design of Authentication and Authorization Infrastructure (AAI) for NFFA-EUROPE community. This AAI will allow NFFA-EUROPE scientific users to access NFFA services through a simplified approach. Such services include the NFFA portal where proposal are submitted, access to all the physically distributed repositories at different facilities and finally global access to the Information and Data Repository Platform (IDRP) that is presented in the twin Deliverable 8.2: "Design of the finalized repository architecture".

This report reviews requirements for shared authentication across the NFFA facilities and proposes a prototype authentication system in view of the needs of the NFFA user community

The document is organized as follows:

- Section 2 will introduce the basic terms and briefly discuss the general problem of AAI as it appears to date.

- Section 3 will details the NFFA uses cases, the requirements and needs in term of AAI for the NFFA user community and some in depth analysis of them.

- Section 4 discusses the proposed architecture and analyses how it can match the identified requirements.

Three short appendices illustrate some of the available tools to provide Authentication and Authorization mechanism on distributed infrastructures and the NFFA evaluation performed with respect to our specific needs.

2. Introduction

We give here a very short introduction to the fundamental concepts of Authorization and Aunthentication problem, we then focus on the NFFA requirements so far identified and we discuss the typical workflow of NFFA users which involve AAI. This introduction will then set the scene to present an discuss our solution.

2.1 Fundamental concepts

Authentication and authorization

We will give here the definition of authentication and authorization:

- AUTHENTICATION IS THE IDENTIFICATION OF A PERSON BY MEANS OF SPECIFIC CRITERIA (E.G. USERNAME, PASSWORD, OFFICIAL DOCUMENT, ID-PICTURE).
- AUTHORIZATION IS THE MEAN TO PROVIDE CERTAIN RIGHTS (E.G. ACCESSING A DATA FILE, A REMOTE COMPUTER) TO A PERSON WHO HAS BEEN AUTHENTICATED.

Authorization can also be given to groups (e.g. members of a research group) but in this document we consider it as a one-to one-relation between a person and access rights, i.e. an authenticated person as part of one research organisation has certain rights. Thus, authentication and authorization are differentiated. Nonetheless these terms are inter-related, as e.g. a security-critical authorization will, in general, require a more stringent authentication (e.g. inspection of a passport rather than a simple Google-type email handshake).

Identity provider and Service provider

Other important functional terms are identity provider (IdP) and service provider (SP). A SP offers a specific service, e.g. an application which requires a license for its use. If a specific user requests this service, then the SP will, by means of the IdP, determine if access to this service is authorized. Therefore, the SP sets the rights for its use based on the information retrieved from the IdP.

2.2 AAI approaches for distributed collaborations

Controlling access to research-related resources and collaborative tools is challenging, particularly when dealing with research communities that can be geographically dispersed across Europe. This is clearly the case of NFFA-EUROPE project.

In general what is needed is an AAI framework for authentication and authorization that will allow research communities to share information resources and services easily and effectively across e-Infrastructures in a secure, well-controlled fashion, while at the same time reducing operational burden.

We underline here that AAI plays a key role and has to be carefully designed since, by its nature, the elements of a NFFA research infrastructure are distributed with different national and legal structures.

NFFA-EUROPE recognizes that AAI in distributed systems is a technology that is not yet readily available. There are presently several European efforts and (not yet completed) solutions for implementing an authentication and authorization infrastructure for distributed and federated research community. We mention here the AARC project [1] that aims at "*enable the design and pilot of an integrated Federated Authentication and Authorization Architecture for e-Infrastructures*" (Deliverable DJRA1.1 of the AARC project [2])

NFFA-EUROPE is a distributed infrastructure where the network of NFFA centers is at the core and special federations with community centers will be created – sometimes temporarily. For all of these, there must be agreements that specify the expectations and services and they need to be made explicit and published.

NFFA-EUROPE project will work together with other project to learn, share, and hopefully improve the European AAI situation.

3. NFFA AAI: use cases/requirements and concepts

In this section we will discuss in detail why and when an AAI is needed for NFFA-EUROPE. We will start presenting the basic use case scenario for the standard NFFA user. Such a basic use case scenario is also described in Deliverable 8.2 (Design of the finalized repository architecture) and it is reported below. We will highlight here all the points where authorization and authentication mechanisms are needed but without giving any technical details about how to implement them.

We then analyse it from the technical points of view identifying clearly the needs and the AAI concepts required.

3.1 Typical use case scenario

This section describes an use case covered by a standard NFFA user. In this scenario scientist Bobby wants to measure a sample using two facilities F1 and F2.

Bobby registers at and logs in to the NFFA portal and creates a proposal involving F1 and F2. While creating the proposal, the scientist provides a basic set of metadata following the NFFA metadata model[3]. At this point the scientist is supported by vocabularies and intelligent filtering of possible inputs, e.g. after selecting F1 only instruments available at F1 should be selectable for this facility. This minimizes the error suspectibility as well as effort for the scientist. The information provided with the proposal finally delivers a basic set of metadata describing the experiment(s) as well as proposal- and facility-specific information, e.g. which are the coproposers getting immediate access to measurement results, which facilities and instruments are used etc.

Latest when the proposal has been accepted the provided information is used to register captured metadata following the NFFA metadata model at the NFFA Information and Data Platform automatically. Regardless the generic nature of the NFFA metadata model this pre-registration provides enough information to be able to identify and to retrieve the respective measurement later. Missing entries are mainly related to data assets or additional, experiment-specific content metadata and can be added after the measurement.

After the proposal has been accepted, Bobby (or any registered co-proposer) may travel to F1 and F2 to perform the proposed measurements. At this point, the scientist can concentrate on performing experiment(s) rather than taking care of providing/completing metadata. Furthermore, the experimental phase is assumed to be fully independent from the IDRP. This is because experiments will typically be carried out using facility-specific user identities that are (in most cases) different between F1 and F2. Furthermore, most facilities are providing local data archives for storing output data using either the before-mentioned facility-specific user identities or archive-specific identities. As these facility-specific workflows should not (and probably cannot) be affected by NFFA, they need special attention elaborated later in this document and within JRA3-WP8. As soon as the experiment(s) at one facility have finished Bobby or a valid representative has to register the

experiment and its measurements at the IDRP, either by using the NFFA portal or via a custom (command line or graphical) client. As most of the proposal-related metadata already exist, experiment specific parameters, e.g. the used facility or which kind of data archive is connected to this facility, are already known. The only information that is missing are details about the outcome of the according measurement(s) of the experiment(s), e.g. references to produced data assets or additional, custom metadata documents available for the according measurement. After the successful registration of missing metadata the IDRP takes care of actually ingesting the information into the repository platform, including e.g. extraction and indexing of custom metadata, assigning access permissions to all participants of the proposal or the validation of the provided information. As soon as the ingestion has finished, the registered data assets are retrievable via the NFFA portal by all registered participants of the proposal.

Such a semplified scenario makes evident the need of a single Authorization and Authentication Infrastructure for the NFFA because is actually cumbersome that an user should have three different authentication mechanisms.

3.2 AAI Requirements for NFFA

In the NFFA distributed infrastructure we have different access points, for the user as well as for machines on behalf of the user, all of them having slightly different requirements/possibilities. Figure 1 gives an overall impression on the involved systems (blue boxes) and interface layers needing some kind of authentication and/or authorization (green boxes). The type of communication (user-machine or machine-machine) is written in the according box. The direction and location of the green boxes applies to the basic use case and might be different for other use cases, e.g. the access to the IDRP from a facility can be also done via REST interfaces or via the portal.

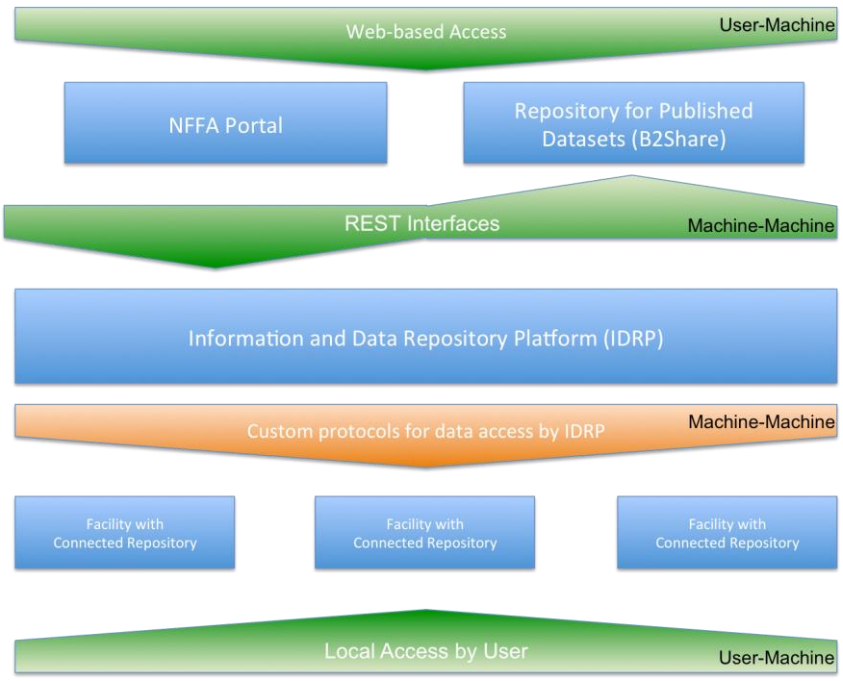

Figure 1: Involved systems (blue boxes) and interface layers needing some kind of authentication and/or authorization (green boxes)

In summary, the first basic needs for the NFFA distributed facilities are:

- A primary entry point for NFFA users, identified so far by nffa.eu the portal of the project (registration, login, changes)
- A centralized mechanism to automatically propagate the registration/changes from nffa.eu primary entry point to the IDRP.
- A mechanism to associate/link/replace local user authentication of each facility to an integrated and federated distributed AAI
- A mechanism to allow NFFA and external user not yet registered to browse IDRP directly without having to first register on the nffa.eu portal

4. NFFA-EUROPE design of the Trust source for authentication

In this section we will present the final design of our AAI architecture which is described in detail in the following subsections.

The requirements and the needs discussed in the previous section are here taken into account and solutions will be presented. The solutions have been elaborated after several in depth discussions where we considered and analyzed carefully already in place solutions (see appendixes and references).

We will illustrate them and the overall AAI architecture by means of the following three different actions a NFFA user will accomplish within the distribute NFFA infrastructure with respect to AAI procedures:

1. CREATION: THE CREATION OF ONE ACCOUNT AND THE ASSOCIATED OPERATIONS THAT WILL GRANT PERMISSIONS TO IDRP.
2. ACCESS: THE ACTION TO PROVIDE ACCESS FROM DIFFERENT FACILITIES TO THE IDRP ONCE DATA HAVE BEEN COLLECTED AND/OR MEASURED.
3. PUBLICATION: THE USER, ONCE COMPLETED ALL THE ANALYSIS, IS READY TO PUBLISH HIS/HER DATASET ON THE SPECIFIC PUBLIC DATA SERVICES.

For each of the above actions we will provide the workflow the AAI infrastructure will accomplish with some technical details.

Creation

This action is performed on the www.nffa.eu portal. User will find a registration procedure on the portal that will provide a username/password pair and an associated userID.

Such information is then kept locally in the portal itself. The portal will then also provides a centralized authentication system by means of an API (Application program interface) where all distributed applications (and users) can access to check if username/password pair exists or not.

In case of positive answer the application will retrieve the User ID (and some more personal data if needed).

This means that when a generic user will browse the IDRP repository and then tries to authenticate there a simple authentication application will issue a request to the NFFA portal to check the username/password pair and if the answer is positive access is granted. At this point either the IDRP or the Portal creates a token used to identify the user for REST service access and during proposal creation basic metadata is created and stored via REST interfaces authorized by user token at the IDRP. This metadata also contains information about co-proposers allowed to access the proposal (its metadata and data).

We incidentally note that a generic user that will land on the IDRP and wants to create an account there, will find a link that redirects him/herself on the nffa.eu portal where the account will be actually created.

Access

This operation involves users at different NFFA facilities. The users store measurement data in connected repositories, either located at the facility or somewhere else; currently known repositories that are used are ICAT[4], KIT Data Manager[5], and AIIDA[6].

It is worth noticing that NFFA AAI has no influence on access to those local infrastructures; it should only take care of the permissions needed by NFFA users to store all the collected and/or measured data on the IDRP.

In this case the user (or an ad-hoc application) will log in by means of the centralized authentication service on nffa.eu portal to retrieve the username/password pair. This will also allow user to retrieve metadata of a proposal via REST interface. For access to data assets (still only available at the repository where they have been stored by the user) the IDRP can use the registered data asset reference, e.g. an URL or identifier. At this point there are multiple options:

1. THE OTHER REPOSITORY TRUSTS THE IDRP AND PROVIDES READ ACCESS TO DATA ASSET BASED ON IP/MAC/HOST CERTIFICATE
2. THERE IS A CREDENTIAL DELEGATED TO THE IDRP THAT CAN BE USED TO READ THE DATA ASSET FROM THE OTHER REPOSITORY ON BEHALF OF THE USER.
3. THE IDRP JUST PROVIDES THE REFERENCE AND LEAVES IT TO THE USER TO AUTHENTICATE OR TO CONTACT THE USER WHO OWNS THE DATA ASSET.

With this respect we note that UMBRELLA [7] tool will be integrated into the nffa.eu portal. This will allow NFFA users, once registered, to associate his/her NFFA userID to the Umbrella federated identify system in order to allow a Single Sign-On (SSO) on all the NFFA facilities which are federated within the Umbrella project.

Moreover the IDRP will integrate the Umbrella tool and therefore, once the user has associated his/her NFFA-userID with Umbrella, a single sign-on mechanism will be active. This means that once the NFFA users is logged in on one site all the other sites (NFFA-facilities, IDRP) can be reached without any further authentication.

We stress here however the fact that the SSO policy is avaiable only if the user associates manually his/her account to the Umbrella federated mechanism. We will provide training sessions and tutorial to introduce this practice to the NFFA community.

Publication

This operation involves NFFA users at their own location. Users log in to the NFFA portal (or directly on the IDRP portal) and select a measurement to publish on an external public data service. With this respect NFFA-EUROPE became pilot project for EUDAT with the goal to integrate EUDAT services within our infrastructure.

As a first service we identified the B2Share one. In order to be able to publish to B2Share, the user should also provide a simple access token for B2Share or the NFFA portal has to obtain an OAuth2 token from B2Access; how to implement this point is still under discussion with EUDAT supporting team and a full integration of the EUDAT B2share service will be considered at a later stage.

More details about the integration the EUDAT B2Share service for data publication are also given in the "Integration of EUDAT services" section in the twin deliverable D8.2.

Security issues

The AAI implementation described above should be trusted and secure. We mention here that any site which will use the centralized authentication system will be requested to send username/password pair to the nffa.eu portal. This of course requires to take into account some security aspects/issues.

The following list of requirements provides us with an adequate level of security:

1. the nffa.eu portal will be exposed by means of https protocol only;

2. all the sites which plan to use the Authentication mechanism should be on https protocol as well;

3. the server-server connection to verify the credentials should be protected and encrypted.

In order to avoid misuse and potential abuse of our server we also maintain a specific list of servers or sites allowed to use the authentication services. This will be implemented by mean of the `Access-Control-Allow-Origin` header that allow us to restrict access to one or more domain specified by us.

5. Conclusions

This document presents a general AAI for the NFFA-EUROPE user community. The proposed infrastructure builds on the current AAI landscape and is driven by the requirements identified by several discussions within the JRA3 and the overall NFFA communities. The AAI infrastructure has been devised in strict cooperation with the design of the repository architecture described in an appropriate deliverable. The proposed architecture is far from being the ultimate solution for our needs: however it already addresses a number of the requirements coming from the NFFA community and implements a reasonable solution. More of the remaining challenges will be incrementally covered in the next releases of the repository architecture, taking into account the work done also by other European project like AARC[1]. An overview of the analyzed AAI architectures and tools available is briefly presented and compared in the Appendixes.

Appendices

In these appendices we shortly review some of the available tools to provide Authentication and Authorization mechanism on distributed infrastructures. We refer to specific web sites and documents there for more detailed information. The appendix A3 reports a bullet point comparison of the tools as analyzed in several online discussions with JRA3-WP8 participants.

A1: Umbrella

Umbrella is the AAI developed by PaNdata (Photon and Neutron Data Infrastructure) for the European Neutron and Photon community.

Umbrella defines its own SAML-based federation. It is not associated with any national identity federation or eduGAIN. Umbrella provides one global, replicated IdP that provides a persistent and unique user identifier over the lifetime of a user. Users can sign up in a self-service fashion for an account in this IdP. Authentication information is provided and updated by the users themselves (self assertion). The Umbrella account needs to be linked to a trusted one at each facility. This avoids central handling of a complex general model for trust relations and procedures between participating facilities (sites), but multiplies work for user ID vetting by delegating this to each facility.

Due to the initial low level of assurance and due to the nature of workflows in the targeted lightsource community (users travel to facilities), their identity has to be vetted upon initial visit at each facility, e.g. by providing a photo-ID for verification. Login to the Umbrella SP-IdP-proxy using a home identity is on their roadmap.

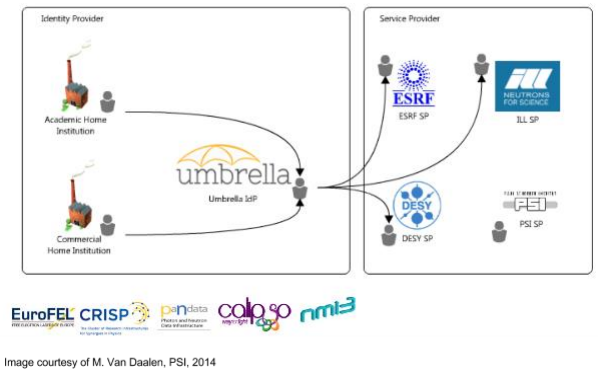

Figure 2: Umbrella

A2: EUDAT B2Access

EUDAT2020 is a Horizon 2020 project building a "collaborative data infrastructure," i.e. an infrastructure which offers data services to researchers. It offers services for replication (B2SAFE), sharing (B2SHARE), personal storage (B2DROP), moving to and from other infrastructures such as PRACE or EGI (B2STAGE), and more. The initial user communities were linguistics (CLARIN), climate (IS-ENES), Earth observation (EPOS), and human physiology (VPH), but the communities have since expanded beyond this.

The EUDAT infrastructure is implementing an AAI solution called B2ACCESS [8]. It serves as a central proxy component which bridges various AAI technologies used within the user communities (SAML/eduGAIN, OpenID Connect) to various technologies used by the service providers within the infrastructure. B2ACCESS also manages additional attributes for the users, some of which are provided by the home IdP, some are self-asserted by the user or allocated by a group manager. The collected attributes are passed on (pushed) to the services in the created token or made available (pull): the token translation component creates the tokens required for accessing a specific service. Supported are X.509 (push), OAuth2/OIDC (pull), and SAML (push), as well as an API for account synchronisation with services (pull).

EUDAT IDs are created by the B2ACCESS upon registration. Therefore B2ACCESS is an Identity Providers for the users that do not have neither a Google account nor a Home Organization Identity Provider. In these cases, B2ACCESS offers also the tool for the managements of the EUDAT IDs.

A3: a short comparison

A short comparison among AAI infrastructure available is here reported as output of an internal discussion within JRA3 members;

Umbrella [7]

Pros:

- SUPPORTS ACCESS TO ESRF, PSI, ELETTRA, DESY AND ISIS
- PARTLY CENTRALIZED SERVICE: HAS NOT TO BE OPERATED BY NFFA

Cons:

- WEB-BASED AUTHORIZATION ONLY
- SEEMS NOT TO SUPPORT ANY STANDARD
- PROCEDURE HOW TO INTEGRATE NEW FACILITIES UNCLEAR/MIGHT BE DIFFICULT TO ACHIEVE
- UNCLEAR STATUS, "FUTURE" GOALS OF 2015 NOT REACHED.
- PARTLY CENTRALIZED SERVICE: IS NOT UNDER FULL CONTROL

B2acces [8]/Unity [9]

Pros:

- OPEN SOURCE, GOOD EXTENSIBILITY
- ALLOWS AUTHENTIFICATION AND AUTHORIZATION VIA MANY STANDARDS AND STANDARD TECHNOLOGIES (OPENID, OAUTH 2, SAML, CERTIFICATES)
- FEDERATION SUPPORT VIA SAML (SHIBBOLETH),SUPPORTING AUTHENTIFICATION OF MEMBERS OF MANY EUROPEAN INSTITUTIONS
- ALLOWS AUTHENTIFICATION CALLOUTS TO EXTERNAL SYSTEMS, E.G. LDAP, KERBEROS
- ADOPTED AND SUPPORTED BY BIG PLAYERS, E.G. EUDAT AND RDA (CLARIN)
- POTENTIAL CHANGE TO CENTRALIZED INSTANCE IN FUTURE WITH MINOR EFFORT
- B2ACCESS AS CENTRALIZED INSTANCE NOT USABLE YET; OWN INSTANCE PROVIDES US WITH FULL CONTROL
- PARTLY CENTRALIZED SERVICE: HAS NOT TO BE OPERATED BY NFFA

Cons:

- SERVICE HAS TO BE OPERATED AND MAINTAINED BY NFFA
- ROADMAP LATE WITH RESPECT TO OUR SCHEDULE

LDAP [10]

Pros:

- ESTABLISHED AND WIDELY ADOPTED TECHNOLOGY
- LDAP ADAPTERS OFTEN AVAILABLE, E.G. WEB PORTALS, DATA TRANSFER SERVICES
- INFORMATION FROM LOCAL LDAP INSTANCES COULD BE INTEGRATED
- NO CENTRALIZED INSTANCE AVAILABLE: OWN INSTANCE PROVIDES US WITH FULL CONTROL

Cons:

- PROBABLY INTEGRATION NOT POSSIBLE DUE TO DATA PROTECTION POLICIES AND PERSONAL DATA
- NO CENTRALIZED INSTANCE AVAILABLE: SERVICE HAS TO BE OPERATED AND MAINTAINED BY NFFA
- RATHER LOW-LEVEL SOLUTION
- RESPONSIBILITY FOR USER MANAGEMENT FULLY IN OUR HANDS, NO PROSPECT OF IMPROVEMENT

References

[1] AARC project: www.aarc.eu

[2] Deliverable DJRA1.1: Analysis of user-community and service providers' requirements Document Code: DJRA1.1. Available at www.aarc.eu

[3] NFFA WP11. Draft metadata standard for nanoscience data. Available at http://intranet.nffa.eu/DocumentRepository?folderHash=MTA5NQ2

[4] ICAT: metadata, data and processing: https://icatproject.org/

[5] KIT Data Manager: http://datamanager.kit.edu

[6] AIIDA: http://www.aiida.net

[7] Umbrella: https://umbrellaid.org/euu/

[8] EUDAT B2Access: http://b2access.eudat.eu/

[9] Unity: http://www.unity-idm.eu/

[10] LDAP protocol: https://en.wikipedia.org/wiki/Lightweight_Directory_Access_Protocol