Advancing Technologies and Federating Communities

A Study on Authentication and Authorisation Platforms For Scientific Resources in Europe

Contents

Executive Summary

Supporting and promoting scientific research and innovation, as well as enabling access to scientific information, have always been key priorities for the European Commission and the Member States. It is widely acknowledged that Authentication and Authorisation Infrastructures (AAIs) play a crucial role in supporting research and in providing a distributed virtual environment where scientific resources can be stored, accessed and shared. More interactive, collaborative approaches to research in conjunction with the deluge of data are opening new frontiers to data processing, storing and preservation; this also poses new requirements and challenges for existing AAIs across Europe.

The goal of this study, prepared for the European Commission, is to evaluate the feasibility of delivering an integrated Authentication and Authorisation Infrastructure, AAI, to help the emergence of a robust platform for access to and preservation of scientific information within a Scientific Data Infrastructure (SDI).

The output of the study consists of a set of recommendations for the delivery of an integrated AAI for the European SDI. The recommendations target different stakeholders; the European Commission for the definition of a possible directive; developers to encourage them to use standard technologies to achieve interoperability; Member States for creating the conditions for such an infrastructure at a national level; and policy makers, particularly those involved in the Data Protection Directive, to create awareness of the impact of legislation on cross-boundary access management.

"The report also describes some target scenarios (for expected/suggested future use of SDI) that should allow the identification of requirements to future AAI/AAA."

This document focuses on **three key areas**:

1. Presenting the requirements for the AAI for SDI, as derived by a collection of use-cases identified among different communities. The use-cases call for:

a. federated access;
b. a trust infrastructure to motivate researchers to share their research environment with other researchers;
c. policies (and consequently proper authorisation mechanisms) to protect data ownership and intellectual property rights.

The paper 'Federated Identity Management for Scientific Collaborations' [FIM] is recommended reading for a more in-depth technical analysis of the e-Research requirements.

2. Analysing the results from a state-of-the-art survey of existing AAIs. Investments have been made over the last ten years to deploy AAIs to serve different purposes; examples of this are eduroam, eduGAIN, EGI, PRACE, and EUDAT (under development). The overview and analysis provided focusses on the infrastructures currently managed by and used in the research and education sector, their underlying technologies and standards, and the use-cases they support. For completeness, STORK (the infrastructure to achieve recognition of electronic identities among Member States) is also described.

This section of the study provides a high-level overview of the implications of data protection laws on the international transfer of personal information. A more extended document is available online [data-protection]. Because of the diversity of the requirements coming from the various communities and because of some limitations within the current technologies, it is impossible to have a one-size-fits-all infrastructure. However, some trends can be observed:

a. All infrastructures evaluated provide Single Sign-On for the users, although the technology used varies;
b. No single AA technology can be adopted universally, but mechanisms must be provided to allow for the integration of different technologies;
c. There is an increased interest in using federated access, although enhancements to the current identity federations are needed to better address e-Research requirements or new requirements;
d. There is an increased interest in cloud computing, which may offer customised and scalable solutions for the data deluge problem; however, users' management, security considerations, legal implications particularly for the public clouds and cost-models associated with deploying large scale cloud solutions are not fully clear and deserve further investigations.

3. Presenting the main challenges and recommendations that the European Commission and other relevant stakeholders should address to develop an open and sustainable AAI for the SDI. The recommendations have been organised into:

a. Technical Recommendations;
b. Policy and Practice Recommendations;
c. Legal Recommendations;
d. Recommendations for Funding Agencies, the EC and Member States.

The general assumption confirmed by this study is that an AAI for SDI should be built on standard technologies, using translation mechanisms among various technologies and that federated access plays an important role. To fully benefit from federated access however, more funding is needed to improve the coverage of national identity federations and more research to enhance authorisation and accounting mechanisms. Support for the development of a common policy and trust framework for Identity Management is needed.

REFEDS (Research and Education FEDerations), the international body led by TERENA to coordinate Identity Federation processes, practices and policies and to discuss ways to manage inter-federation work, could play a pivotal role in this process, in cooperation with the e-Infrastructure Reflection Group [eIRG] and the European Commission. Collaboration and communication between REFEDS, the European Commission, IGTF, eIRG, ESFRI, datacentres and libraries should be improved; dedicated funding to support this should be provided.

Lastly, consistent implementation and interpretation of the legal requirements in the data protection area is essential when building an international infrastructure.

This version of the report includes input received during the final study workshop held on 12 July 2012 in Brussels.

1 Motivations for the Study

1.1 Background

Supporting and promoting scientific research and innovation, as well as enabling access to scientific information, have always been key priorities for the European Commission and the Member States. Rapid developments in Information and Communication Technologies (ICT) have made the Internet much more pervasive and have changed the way in which researchers work. Scientific research has become extremely data intensive and much more interdisciplinary, international and real time.

A consequence of these changes is the deluge of data generated by scientific experiments in various disciplines, produced by wide-scale observational data collection, as well as by the digitisation of content in the arts, humanities and sciences in general. For example, the Large Hadron Collider (LHC), built to advance research in the area of particle physics, will produce roughly 15 Petabytes (15 million Gigabytes) of data annually. This is just the data generated by one research activity in a single discipline. The Genome research at the cutting edge of modern research requires access to a data volume of Terabyte scale, assurance of data integrity, and around-the-clock availability of data. It is also estimated that digitising the whole of the currently available paper-based content and artefacts in the humanities (history, literature, behavioural science) and arts will produce around 2-3 Petabytes of information monthly.

Over the years, thanks to the funds made available by the European Commission and the NRENs, researchers have enjoyed a high-speed network (provided by GÉANT), an infrastructure to access supercomputing resources (offered by PRACE), a federated wireless infrastructure to allow for seamless network access [eduroam] and online tools (i.e., wikis, chats etc.) to create, share and consume digital information in real time.

Whilst there is not yet a single, coordinated, European data infrastructure serving multiple disciplines, there are a number of projects that offer a solution for specific user communities. Examples of these projects are EURO-VO, which offers access to astronomical data archives, OpenAIRE, which supports the implementation of Open Access publishing in Europe, and APARSEN, which is concerned with the preservation of the record of science.

Projects such as ODE, which engages with different stakeholders to work towards an interoperable data sharing and preservation infrastructure, show that libraries and datacentres are committed to providing access to research data and organising, linking and storing them in a trustworthy environment. This project also highlights that the potential of the data deluge can only be unlocked by complementary network and computational facilities supported by interoperable data sharing, reuse and preservation services.

The data produced by research are very heterogeneous, as is the demand to access, store, protect and preserve them. This clearly represents both an opportunity and a challenge. Whilst technologies for 'Big Data' advance and empower users to access an unprecedented abundance of content, they also raise issues concerning authenticity, quality of data and copyright for existing e-Infrastructures. As indicated in the Strategy for a European Data Infrastructure [PARADE] "building of mutual trust among all stakeholders is of utmost importance to the realisation of the SDI".

" Collaboration and partnering are essential in the e-Research environment. While some organizations will specialize in building tools and others in building relationships, both are required."

Rick Luce, 'No Brief Candle: Reconceiving Research Libraries for the 21st Century' (2008)

" ... to collect, curate, preserve and make available ever-increasing amounts of scientific data, new types of infrastructures will be needed."

High Level Expert Group on Research Data, 'Riding the Wave' (2011).

As acknowledged by the Digital Agenda for Europe 2020 and the e-Infrastructure Reflection Group [e-IRG] white papers, there is a need for harmonisation of existing e-Infrastructures. The High Level Expert Group [HLEG] on Scientific Data goes one step further in the report 'Riding the Wave' by stating that: "*to collect, curate, preserve and make available ever-increasing amounts of scientific data, new types of infrastructures will be needed*".

To unlock all the benefits of data-centric research for the knowledge society, Europe needs to build a modern trans-European Scientific Data Infrastructure [SDI] to integrate existing research infrastructures, connect all scientific communities to a high-performance network, and provide access to high-performance computing. New types of (data-centric) infrastructure require new types of access control and security infrastructure capable of answering the challenges of data persistency, authenticity, long-term preservation, and privacy.

All signs point in the same direction: the underpinning infrastructures to rapidly transmit (high-speed networks) and process data (high-performance and high throughput computing facilities) should evolve into a next-generation infrastructure that offers scientists and citizens alike the opportunity and means by which to harness the potential of data. What this infrastructure should look like and the conditions necessary for its implementation are key questions which this study sets out to answer.

1.2 Objectives of the Study

The goal of this study, prepared for the European Commission, is to evaluate the feasibility of delivering an integrated Authentication, Authorisation and Accounting Infrastructure (AAAI) to support the development of a robust platform for access to and preservation of scientific information (SDI).

The goal has been broken down into **two objectives**:

1. A collection of **users' access requirements** coming from different communities;
2. A **gap analysis of the existing AAIs** used in the realm of research and education, the use-cases they support and the associated challenges.

The **output** of the study consists of a **set of recommendations** (Chapter 4) for the delivery of an integrated AAI for the European SDI. The recommendations target different stakeholders:

- The European Commission for the definition of a possible directive to set a unified AAI system and to allocate funding to address specific areas as indicated in the recommendations;
- Developers and AAI operators to encourage them to use specific standards to achieve interoperability;
- Member States for creating the conditions for such an infrastructure at a national level;
- Policy Makers, particularly those involved in the Data Protection Directive, to create awareness of the impact of legislation on cross-border access management.

Because of the multiplicity of requirements, such as support for different user communities across different countries, support for cross-disciplinary data sharing to protect data integrity and ownership, and support for different access levels, the AAAI for the SDI needs to be designed to offer flexible and scalable access-control mechanisms. Clearly this AAAI has to ensure that resources and facilities are used in the correct way and that data are accessed only by users authorised to do so. It is also important to ensure that policies are implemented to deliver a trusted environment for researchers and for data to be processed, stored and shared safely.

Figure 1.1 shows the structure of the study, the role of the partners and experts in relation to the study objectives, and the final output of the study.

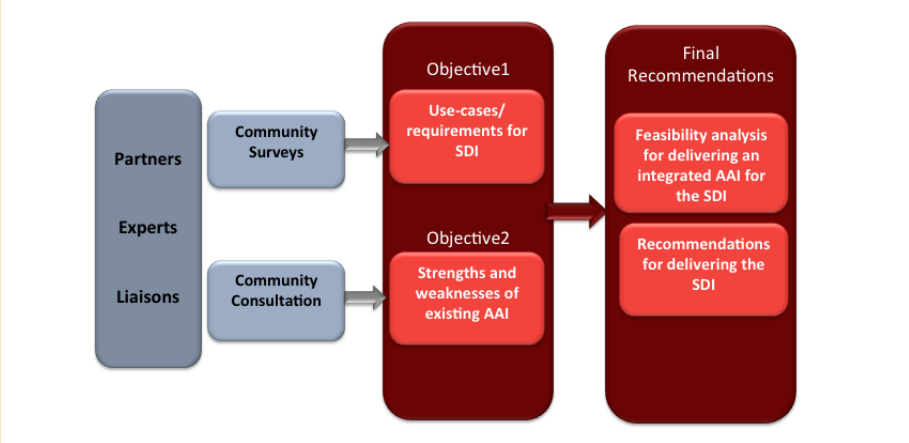


Figure 1.1: Structure of the Study

1.3 Methodology

Utilising the diversity and strengths of the partnership, the study took a dual approach to exploring the challenges and opportunities associated with implementing an Authentication and Authorisation Infrastructure for the future SDI:

1. **Use-cases** have been derived from interviews with stakeholders within the e-Research, datacentre, networking, and library communities. The selected use-cases reflect issues such as data sharing, persistent access, data curation, data management and governance, and long-term preservation. These interviews and use-cases have been instrumental for assessing how existing initiatives can meet the resulting requirements and in describing future scenarios that would benefit from the SDI.

2. **Existing and emerging infrastructures in the realm of research and education have been surveyed** in order to assess how well they meet the requirements identified through the use-cases. The survey provides a complete overview of the AAI landscape in Europe and identifies interoperability features.

" Use-cases have been derived from interviews with stakeholders within the e-Research, datacentre, networking, and library communities."

Finally the study proposes recommendations for the integration of existing research and education e-Infrastructures in order to deliver an appropriate AAI for the SDI. It highlights issues and identifies technical, organisational, regulatory and legal obstacles to pan-European AAI platforms.

FIgure 1.2 depicts the approach followed by the partners in delivering the final recommendations.

Figure 1.2: Methodology

Community feedback has been sought on all preliminary results of the study. Draft reports have been circulated to the partners' communities for review. The results presented in this final report have been validated and can therefore be considered representative of the current situation.

1.4 The Partners

The study, led by TERENA, was carried out by four partners representing the networking, library, and e-Science communities (see table 1.1). The partners were supported by external experts throughout the study. These experts provided contributions on specific topics as well as general comments on the overall study.

The strength of the study lies in the diversity of the partners (NRENs and Identity Federation operators, libraries and e-Science communities) involved and the variety of expertise they contribute. The combination of the four partners provided access to a wide cross-section of stakeholders and relevant networks as well as a unique insight into the issues, both human and technical, associated with implementing and deploying an Authentication and Authorisation Infrastructure in cutting-edge environments.

" Central to this study is an understanding of the human aspects of access to, and use of, research information."

Central to this study is an understanding of the human aspects of access to, and use of, research information. Libraries have been a traditional intermediary between researchers and sources of research information. With the increasing requirement for Open Access (particularly for publically funded research) and increasing need for digital preservation services, many libraries have also been tasked with the management of institutional repositories. This has not only provided insight into how researchers access information, but has also revealed the barriers and drivers of data deposit.

This study marries the insight of the library and related research infrastructure communities with the technical expertise of groups already active in the AAI area. By bringing these different but complementary communities together to work towards a common goal, the study helps to build a common language and understanding of the required collaborative data infrastructure.

Partners	Description
TERENA Trans European Research and Education Networking Association	TERENA, the association of National Research and Education Networks (NRENs) in Europe, has approximately 60 members including international members (CERN and ESA), and a number of industrial companies that are associate members. The mission of TERENA is to offer a forum to collaborate, innovate and share knowledge in order to foster the development of Internet technology and services to be used by the research and education community. TERENA is involved in eduGAIN, eduroam and other GEANT activities. Licia Florio coordinated the overall study on behalf of TERENA and provided contributions concerning the survey of AAIs and the final recommendations.
LIBER Association of European Research Libraries	LIBER is the main research libraries network in Europe. It has over 430 members from national, university and other research libraries across 45 countries. LIBER is actively working to promote the role of libraries within the European research infrastructure, in digital curation, research data sharing, and Open Access. Susan Reilly coordinated the user-community requirements for the library sector.

Table 1.1: Partners

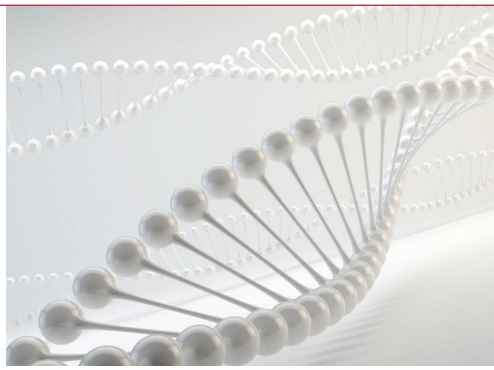
Partners	Description
UvA University of Amsterdam	The System and Network Engineering (SNE) Research group at the University of Amsterdam researches cross-domain interaction between Grid resource providers, optical and hybrid networking, resource descriptions using semantic web, and programmable networks for the Future Internet. SNE has expertise in Cloud architecture and security infrastructure research and development, generic AAA architecture and AAA framework implementation. Yuri Demchenko provided input from the e-Science perspective.
DEENK University and National Library of Debrecen	DEENK is the largest research university in Hungary and has a strong tradition of international collaboration in science and scholarly communication. It provides technical support for the Hungarian Open Repository Network. It also hosts a digital archive for PEER (Publishing and the Ecology for European Research). Tamás Varga and Gabriella Harangi ran the surveys of the libraries.

Table 1.1: Partners - continued

External Experts	Expertise
Nicole Harris (JISC Advance)	Nicole Harris has extensive experience in the education sector as an advisor and project manager with a focus on access and identity management. Nicole coordinated the user-communities' requirements concerning the networking community and provided her expertise in finalising the recommendations.
Diego Lopez (Telefonica I+D)	Diego Lopez has been involved in a number of projects and initiatives at RedIRIS and TERENA while working at RedIRIS; he chaired the TERENA Task Force on Middleware (TF-EMC2). Currently working at Telefonica I+D, he focuses on advanced network infrastructures and security. He contributed to the 'Riding the Wave' report as one of the experts in federated access technologies. Diego reviewed the survey on AAIs and provided feedback on the final recommendations.
Klaas Wierenga (Cisco Systems)	Klaas Wierenga is considered the 'creator' of eduroam, the federated infrastructure for network access. He has been involved in several projects both as a SURFnet employee and in his function in Cisco Systems. Klaas is currently chairing the TERENA Task Force on Mobility and Network Middleware. At Cisco he focuses on security, (federated) identity and mobility. Klaas contributed the eduroam and Project Moonshot sections of the AAI survey and informed the final recommendations for the report.
Torbjörn Wiberg (University of Umeå)	Torbjörn Wiberg has been involved in deploying authentication and authorisation infrastructures at campus level. Torbjörn has been also actively involved in the e-Infrastructure Reflection Group for several years.
David Groep (Nikhef)	David Groep is a senior research scientist at Nikhef, the Dutch national institute for sub-atomic physics; he also chairs the EUGridPMA and IGTF, which are responsible for coordinating the global authentication trust fabric for e-Science. David reviewed the Grid section of this document and provided a more extended text about Grid AAI which is available on the study website.
Andrew Cormack (Janet)	Andrew Cormack works as Chief Regulatory Adviser at Janet, the NREN of the United Kingdom, dealing with regulatory and policy issues of running and developing the network and its services. Andrew advised the study team on Data Protection topics.

Table 1.2: Experts

2 Community Requirements



2.1 Introduction

If the purpose of a Scientific Data Infrastructure (SDI) is to enable researchers to create, store and share the data resulting from their experiments, and to find, access and process the data they need, then the practices and concerns of researchers must be central to the definition of requirements for this study. An SDI and any plans or directive for its implementation must accommodate current practices, address current and future needs, and take into account the concerns of researchers and information providers in relation to data sharing and the use of an SDI.

" An SDI and any plans or directive for its implementation must accommodate current practices, address current and future needs, and take into account the concerns of researchers and information providers in relation to data sharing and the use of an SDI."

There are several efforts already underway to investigate how existing Authentication and Authorisation Infrastructures (AAIs) can be adapted or extended to better meet the requirements of diverse research communities. An excellent example of this is the paper 'Federated Identity Management for Scientific Collaborations', coordinated by CERN with input from a range of research organisations.

The Research and Education FEDerations (REFEDS) group has been coordinating the efforts of national identity federations since 2004. REFEDS has been actively working with the FIM report mentioned above and with a wide variety of Virtual Organisation groups across the research space to explore and develop solutions to AAI problems.

In its aim to develop a Pan-European collaborative data infrastructure to serve the needs of different communities and to tackle the specific challenges of data management, EUDAT is working to deliver an AAI that enables federated access for EUDAT partners and beyond. EUDAT explores different technologies to implement their AAI, to leverage existing Identity Federations and to implement credential-conversion mechanisms.

The New Global Data Generation Manifesto [DASISH-manifesto], signed by several members of the European Strategy Forum for Research Infrastructures (ESFRI), highlights the need for an authentication infrastructure at the institutional and national levels, and for the harmonisation of related policies.

2.2 The Data Sharing Community

The data sharing community is heterogeneous in nature. And requirements vary according to discipline, the nature of the data to be shared, its scope for reuse, methods for collaboration, and culture. It is not within the scope of this study to provide an in-depth analysis of the types of requirements that could be derived from individual disciplines, as some disciplines and data sets have very complex and unique requirements. However, some more generic issues can be derived from a brief analysis of the high-level requirements of broad disciplinary areas. The following are the scientific areas defined by ESFRI and an outline of their community requirements for an AAI:

- In b**iological and medical sciences**, a discipline that generates huge volumes of data, issues of data sensitivity are common and any data sharing must adhere to data privacy laws and policy. In the social sciences and humanities, whilst the

data may be less sensitive, they may still be subject to license and issues around the 'long tail' of managing smaller data sets;

- **Environment and earth sciences** have a strong tradition of data sharing, generate high volumes of data, and are more advanced in terms of the technology used to exploit and interact with data than other disciplines;
- **Materials science, analytical and low-energy physics** are characterised by short projects and experiments, leading to a highly dynamic user community. This community would benefit from a collaborative infrastructure that would allow for both local participation and remote access. The community expresses interest in using federated identity management to reduce administrative overhead and in tools to manage ad-hoc collaborative user groups via virtual organisations or federations;
- Lastly it is worth mentioning the ENVRI project and the LifeWatch project as examples where data sharing amongst different disciplines is a key enabler for system-wide science. The main challenge for these groups relates to data capture from distributed sensors, metadata standardisation, management of high-volume data, workflow execution and data visualisation. The EUDAT project is positioned to offer generic data-sharing services to such projects.

2.3 The Nature of Data

The Data Publication Pyramid (Figure 2.1) illustrates a growing problem that an SDI and AAI could help address. The pyramid visualises the ways in which research data can be made available.

The base of the pyramid represents data stored locally in its raw form on hard drives and disks; these data are typically the result of scientific experiments or analysis. There are several reasons why these research data are not shared, varying from intellectual property protection concerns to ethical, technical or cultural reasons. An AAI is one of the mechanisms that can help improve data sharing practices among researchers by providing technical support for addressing data access control and policy-related issues. An AAI could make sharing data more straightforward technically and safer for research, and also help ensure the integrity and security of the data. It can also be argued that by making data sharing simpler and potentially more commonplace, an AAI will facilitate a cultural shift in terms of data sharing and collaboration.

The second layer of the pyramid visualises data that are already stored in repositories. This data is available for use and reuse. Here, an AAI can facilitate collaboration and provide authorised access to the wider scientific community. Although a certain amount of this data may be open, some may require authorisation for ethical, regulatory and

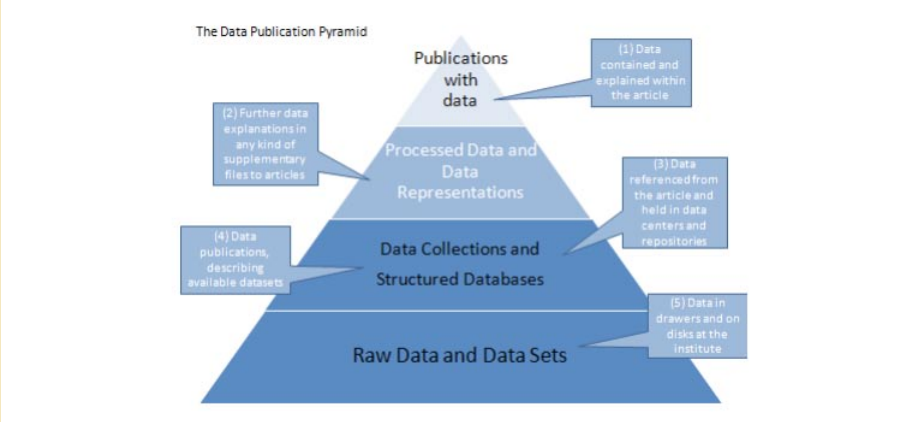


Figure 2.1: The Data Publication Pyramid (source ODE Report in Integration of Data and Publications [ODE])

security reasons. It is also the case that licensed or copyrighted publications may link to open research data and vice versa.

The third layer of data shows publications as supplemental files to articles. Again, these articles and files are located on a publisher platform and may be Open Access or licensed. The top layer is the traditional view of an article or publication with the data embedded within.

For a truly collaborative data infrastructure, researchers need to be able to work with, and collaborate using, data from all of these layers. They need to be able to share their data easily and securely, and they also need seamless access to licensed and Open Access content, from data sets to the finished article.

2.4 Data Management in Big Data Science

The increased availability of computational resources has transformed the way in which research is done and the resulting scientific data used. As indicated in the Scientific Data Life Cycle Management picture (Figure 2.2) four different types of data can be identified:

- **Raw data** collected, produced during experiments, surveys or observations of different phenomena (according to an initial research model); the data is consequently analysed and the findings published. A preservation process is needed during all these stages;
- **Structured data** and datasets resulting from data filtering and processing (supporting some particular formal model);
- **Published data** organised in a way to support a scientific theory and/or research results;
- **Data publishing** to support research consolidation, integration, and openness.

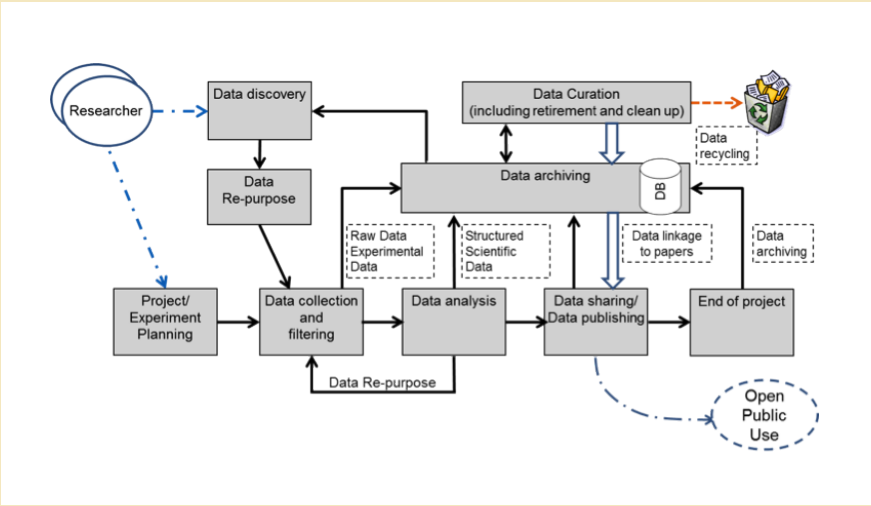


Figure 2.2: Scientific Data Life Cycle Management (source Data Lifecycle Models and Concepts)

Once the data is published, it is essential to allow other scientists to validate and reproduce the data they are interested in, and possibly, to contribute new results. Capturing information about the processes involved in transformation from raw data up to the generation of published data, becomes an important aspect of scientific data management. Scientific data provenance becomes an issue that needs to be taken into consideration by SDI providers[1].

[1] See also http://vgc.poly.edu/~juliana/pub/vistrails-executable-paper.pdf

To enable data finding and consequently data reuse/re-purposing, descriptive metadata should be associated with the data at the instant when the data are created. Understanding the semantics of published data becomes an important issue for reusability; this process has traditionally been performed manually. Best practices generated by the semantic web community on how to provide reusable published data should be considered when developing the SDI.

Different preservation plans should also be defined, based on the nature of the data; for instance the format in which the data is stored should be reviewed in line with the relevant technical developments and the right protocols to access the data should be applied. Access to data may be organised following different policies according to the nature of the data; these policies have an implication on the authentication and authorisation processes and level of trust associated with them (Level of Assurance or LoA).

The AAI for SDI must work seamlessly across the whole data lifecycle and address key issues related to access control.

Future SDI should support all stages of the data lifecycle and allow for multipurpose data collection, use and advanced data processing.

Big Data needs to be collected (sometimes in a time-sensitive way or with other environmental attributes), distributed and/or replicated. Linking distributed data is one of the problems to be addressed by SDI. Facilitation of the storage of initial data sets and all intermediate results will allow for future data use, in particular data re-purposing and secondary research, as the technology and scientific methods develop. As the aim is to make data generated from research available for future reuse, it is important that data remains authentic, reliable and usable; tools, policies and procedures for protecting legitimate privacy, confidentiality, intellectual property, or other security needs should be implemented.

Support for advanced lifecycle functions will become a requirement for the scientific data providers and thus shape new roles for IT/datacentres, archives and libraries.

2.5 What Do We Know about Researchers' Requirements?

The growth in cross-boundary collaboration and cross-disciplinary research has changed researchers' behaviour and their methods of accessing and retrieving. Researchers have less time for complex information seeking. The rise of the 'Google Generation' [google-generation] means changes in expectations and the way information is accessed.

Current practice as to how researchers find, access and process research information and data has been drawn from a survey sent out to research librarians from the 430 libraries within the LIBER network and other communities. The libraries received two surveys. One survey was aimed at the librarians themselves and contained questions relating to how their resources were authenticated and the behaviour of their users. The other survey was sent to the libraries' research communities and explored practices and preferences relating to authentication, as well as attitudes towards the use of information resources and data sharing.

Roughly 100 librarians responded to the survey and the response represented a fairly even geographical spread. The researcher survey had 600 responses. Neither survey was designed to be statistically representative of nationality or discipline, but they do provide an insight into relevant practices and attitudes.

The survey reveals the following:

Researchers primarily use their institutional credentials for authentication (Figure 2.3), although a not insignificant number (19%), use their social network account credentials to access scientific information.

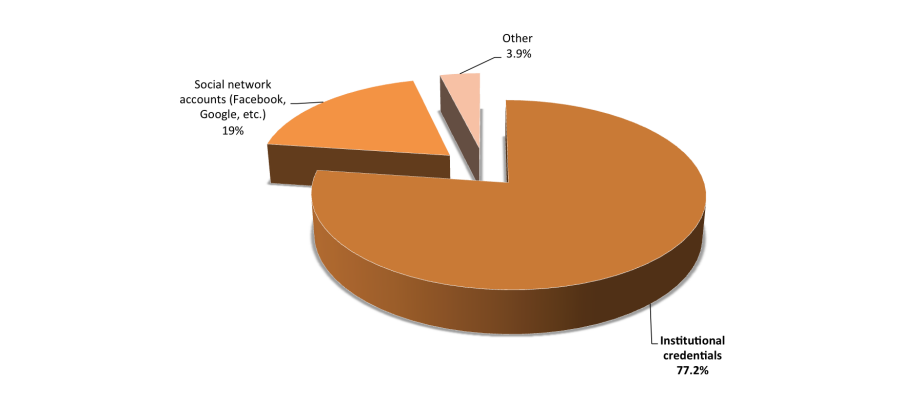


Figure 2.3: Credentials used by researchers

Nearly half of researchers use more than one credential, but a large majority would prefer to access all resources using their institutional credentials.

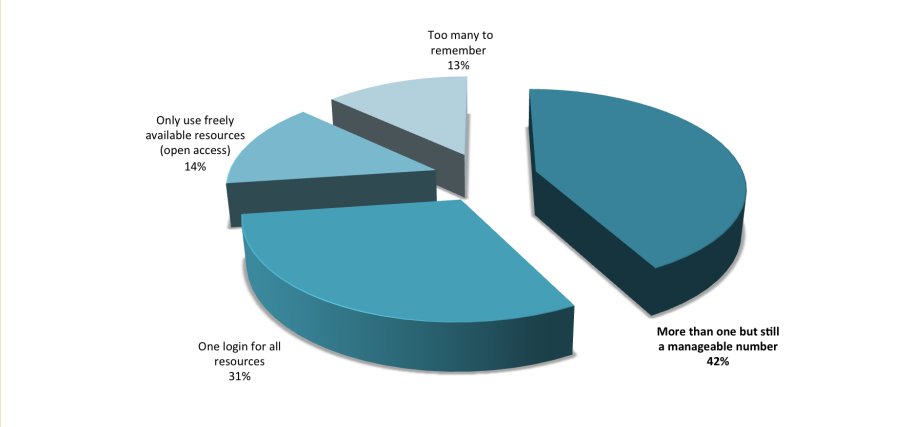


Figure 2.4: Number of credentials used by researchers

IP-based authentication (Figure 2.5) is still the most widely used method of providing researchers with access to information resources subscribed to by institutions.

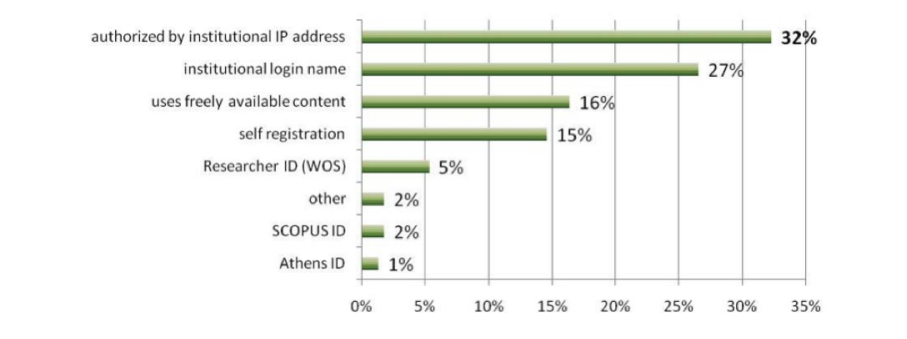


Figure 2.5: Access to institutional subscriptions

The number of researchers depositing or sharing their data in repositories is growing (Figure 2.6) but there is a large percentage of researchers who are still not depositing data. This is down to issues such as trust, IPR and also the fact that researchers cannot always deposit their material directly - they must go through an intermediary, e.g., a librarian.

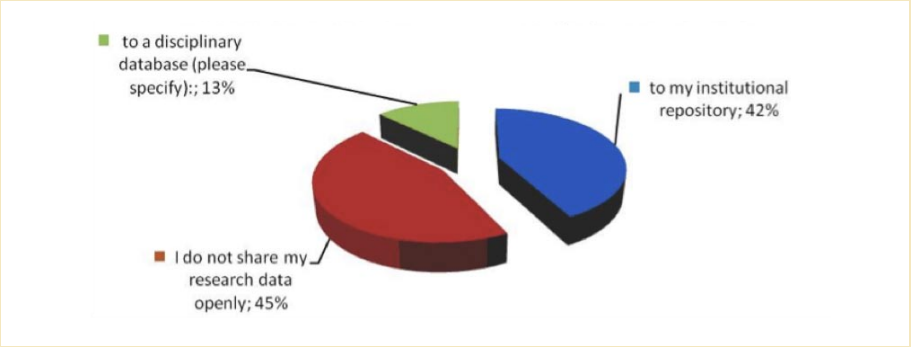


Figure 2.6: Depositing of data

The table below summarises the findings of the survey.

Users'/Researchers' Requirements	Implications
Wish to find: Information in databases (free or commercial); Scientific articles stored in repositories maintained by libraries; (Raw) data generated by previous research or experiments hosted in special datacentres.	Using standardised metadata to facilitate the search and retrieval of information; Motivating researchers to share (raw) data - recent studies[2] show that only very a limited number of researchers share their data even if they are solicited to do so; Enabling an infrastructure to facilitate data sharing among different administrative domains.
Use their institutional credentials to access more services.	Extending the current federated framework or enabling user credentials translation mechanisms to support access to different sets of information with the same users credentials.
Share material (including the licensed material) with other international researchers that collaborate on the same project.	Negotiating different types of licenses.
Trust the infrastructure before depositing their data.	Building and promoting an infrastructure that is able to handle IPR, copyrights and ownership in a simple and reliable way.
Access their content anywhere, anytime.	An infrastructure able to manage the data life cycle and to support different type of access and LoAs.
Preserve and access data generated from previous experiments.	Implementing data curation and access policy protocols.
Facilitate participation of different users to existing access management solutions.	Simplifying procedures and technical access to current access management solutions.

Table 2.1: Overview of users'/researchers' requirements

2.6 Accessing Scientific Data and Information

Currently researchers are largely reliant on institutional credentials[3] and digital certificates[4] for authentication purposes, but it is possible that, as a new generation of researchers enters the fray, the percentage of researchers who use social media for sharing and sourcing information, and for collaboration, will increase.

[2] http://www.plosone.org/article/info:doi/10.1371/journal.pone.0018657
[3] The number of users who possess an institutional account varies from country to country, ranging between 200,000 to millions of users for the big identity federations (source: https://refeds.terena.org/index.php/Federations)

As the survey indicates, the use of IP addresses to identify institutional users is still common practice for accessing online material to which institutions subscribe. This method, characterised by its simplicity of implementation, allows users to seamlessly access resources without needing to go through an explicit log-in process. However, this method has serious shortfalls as it neither identifies the user nor provides permissions, but instead ascertains that the device used to access the resources is in a certain IP range at the time of access. Beside the security considerations (i.e., IP addresses can be easily forged), it only allows users to access a resource when they are physically on campus, which is clearly a limitation in today's mobile world.

Despite its wide adoption, this approach does not provide any features that are desirable in any AA(A)I for the SDI. In fact, it is envisaged that systems like Shibboleth will completely replace IP-based access as a means of providing access to information behind paywalls.

The trend to mandate Open Access (OA), a practice of making more publications publically available, is growing across various communities. This development will need to be supported by Open Access technologies that will require new functionalities, beyond the current institutional subscription or individual paid access to licensed content, to be implemented within the AAI.

Persistent identifiers will play a more important role for data and for researchers as authors and users of data. Persistent identifiers will enable an 'online identity trail' and will improve accounting and statistical analysis of scientific information, publication usage and their inter-relation. This is important, both in terms of ensuring that researchers as creators get recognition for their work and in terms of the preservation of the record of science.

In order to discover the relevant current and future needs and concerns of the researchers in data sharing, use-cases have been drawn from consultations with stakeholders from the library, datacentre, e-Science, and research infrastructure communities. The following are a selection of the use-cases collected through the study. They are a snapshot of the wide array of uses and define some of the requirements for the AAI for SDI in relation to supporting the scientific community in the use and reuse of research data.

Use-case 1
Creating Data

Researcher identification: the ORCID initiative

A fundamental question for researchers, publishers and funding bodies is how to trace a researcher's publications and other research contributions back to the correct person[5]. Researchers frequently move between research groups, change institution affiliations, and even their name. Furthermore, the conventional way to identify a researcher is by last name and initials of the first name, leading to ambiguities since natural names are by no means unique.

The Open Researcher & Contributor ID Initiative (ORCID) is an international, cross-community and interdisciplinary initiative dedicated to solving the name ambiguity

[4] Digital certificates are typically used by the high energy physics, photonic communities, astrophysics and more in general Grid communities. EGI for instance has issued about has 21,000 certificates issued.
[5] Beside ORCID, the PersID initiative aims to solve the problem of assigning persistent identifiers to scholarly and cultural information.

problem in scholarly communication. ORCID will work to support the creation of a permanent, clear and unambiguous record of scholarly communication by enabling reliable attribution of authors and contributors through unique identifiers.

The process of 'asserting identities' will involve, in the simplest scenario, researchers registering themselves to create a profile from scratch, then adding publications to their profile. These claims are self-asserted, both biographical and bibliographical. A small number of universities and other research institutions will bulk-create ORCID profiles, and subsequently those researchers may claim that record from ORCID. After claiming the profile, the researcher will then have control of the profile data.

Tracking provenance is a priority for phase two of ORCID. The aim is to enable consumers of profile data to see where each piece of information came from, and decide, based on the provenance, whether or not to trust that claim.

Requirements:

- Tracking of provenance, authenticity, integrity of the material;
- Integration of researcher ID with institutional credentials;
- Researchers' self registration;
- Securely linking researcher and data identifiers for tracking provenance.

Use-case 2
Processing Data

CERN: enabling the physics community

CERN has approximately 2,250 staff and welcomes annually more than 10,000 guest researchers and scientists from all over the world. These 'users' are detached from their home institute - over 600 different universities worldwide - and use CERN's facilities like physics experiments or accelerators in order to conduct their research.

" CERN is international in every aspect: research & technology, collaborating countries & institutes, participating users, worldwide access to data & computing resources. It is still an oddity that authentication and identities are managed independently at each single institution. We should make this system international, too."

Stefan Lüders,
Head of Computer Security, CERN

For example, CERN runs the Large Hadron Collider [LHC] from which a set of major experiments produce more than 20 Petabyte of physics data per year. These data are subsequently analysed by the international community of high-energy physicists and they require a robust infrastructure to store, manage, and access the data, protect its integrity and support the whole scientific data life cycle. To this end, the LHC experiments and community are supported by the Worldwide LHC Grid [WLCG]. User cooperation and access is organised in the form of Virtual Organisations (VO) which manage user roles and issue user certificates in the form of X.509 certificates. In the context of Identity Federation, the VOs effectively act as attribute providers.

In order to facilitate their work, CERN offers a multitude of additional services to its user community like technical workshops, a library service, and dozens of computing services (file stores, web services, mail services, document and journal services, conferencing systems, etc.). In particular, the computer facilities are accessed on-site and also from abroad, e.g., from a researcher's home institute. These users would benefit from being able to access both their home institute's and CERN's computing services seamlessly. Instead of dealing with multiple identities, a single identity issued by the home institute and accepted by CERN could suffice.

A priority for CERN is that the right to access computing resources is only given to authorised individuals and that any action to access or change data can be traced back to these individuals.

Requirements:

- Delegation of identity management to home institutes;
- Attribute provisioning for users participating in specific research projects managed by the specific research groups (VOs);
- Attribute aggregation;
- Unification and homogenisation of identity federations´ attributes and agreed levels of assurance in order to facilitate authorisation;
- Accreditation of trusted identity Providers (IdPs), based on international standards, depending on the required level of assurance;
- Entitlement management to minimise the occurrence of events where license monies are being paid twice without necessity (e.g., for access to scientific journals).

Use-case 3
Sharing Data

a) Working with communities of practice: DARIAH

The Digital Research Infrastructure for the Arts and Humanities (DARIAH) aims to enhance and support digitally enabled research across the humanities and arts by developing, maintaining and operating an infrastructure in support of ICT-based research practices. DARIAH will work with communities of practice to bring together individual state-of-the-art digital humanities and arts activities across Europe.

DARIAH will operate through its Europe-wide network of Virtual Competency Centres (VCCs). Each VCC is centred on a specific area of expertise. VCCs are interdisciplinary, multi-institutional and international. An AAI service, along with a Persistent Identifier (PID) resolver service, is a core technical service that DARIAH will guarantee as part of its digital research infrastructure.

For researchers, it will be a significant advantage to have easy, yet secure, access to all the resources, data, tools and services that they need to undertake research. In today's time-challenged society, being able to log in easily, using one set of credentials, rather than having to negotiate through a multitude of different authentication systems, will provide significant savings.

A Europe-wide Single Sign-On service will encourage researchers to share their work within a secure and trusted environment. For this, authorisation granularity is essential. With such a service, authorisation could be granted to a secure personal environment for an individual's own research, to secure spaces for sharing with a closed research group, or more widely, to large research communities. As more, often national, AAI infrastructures are developed, the technical challenges of making them work together in a seamless and user-friendly way will need to be addressed.

" A Europe-wide Single Sign-On service will encourage researchers to share their work within a secure and trusted environment. For this, authorisation granularity is essential. With such a service, authorisation could be granted to a secure personal environment for an individual's own research, to secure spaces for sharing with a closed research group, or more widely, to large research communities."

Sally Chambers
Secretary General, DARIAH

The sharing of research data across borders presents legal issues. Even if the technical solutions are in place, resource licenses are typically negotiated at a national level and will not allow access to resources to be authorised on a pan-European level. Conversely, being able to ensure a secure environment for Europe's higher-education community may encourage resource owners to consider licensing resources at a pan-European level.

Requirements:

- Authorisation granularity to groups of users from multiple organisations at a variety of levels;
- Delegated rights to authorise users to become members of specific groups;
- Increase the number of participating institutions in current access management solutions;
- An environment secure enough for resource providers to consider a pan-European approach to resource licencing.

b) The Life Sciences

Biological and medical sciences (also defined as life sciences) focus generally on health, drug development, new species identification, and new instrument development. Life Sciences generate massive amounts of data and pose new demands for computing power, storage capacity, and network performance for distributed processes, data sharing and collaboration. The amount of data that the life science community generates requires powerful datacentres with high-performance networks for data transfer.

There are existing datasets, generated by previous research in this field; one of the main challenges is to interlink them and enable authorised people to access them remotely. Connecting existing data to each other would require new fine-grained access control policy and a consistent enforcement system.

Furthermore, biomedical data (healthcare, clinical case data) are privacy-sensitive data and must be handled according to the specific provisions in the European Policy on Personal Data Processing for such information.

Requirements:

- A trusted environment for data storage and processing;
- User-friendly data encryption;
- Fine-grained access control policy;
- Possibility to filter data based on the applied policy;
- Policy-binding to data in long-term storage to protect privacy;
- The tracking of data usage.

Use-case 4
Preserving Data

Enhancing Publications: the SURFshare programme[6]

An Enhanced Publication is a new type of publication. It links typically text-based publications with additional material, such as research data, models, algorithms, illustrative images, metadata sets or post-publication data like comments or rankings. The option of changing post-publication data, allows an Enhanced Publication to develop over the course of time.

There are several baseline criteria for an Enhanced Publication that are of interest to an AAI, such as recording authorship of the publication and its components.

Enhanced Publications (EP) is a core activity in the SURFshare programme run by the SURF Foundation. Projects range across disciplines, from the humanities to the 'hard' sciences. The technical infrastructure is similar across the different disciplines, facilitating the easy exchange of information across systems. Different disciplines have different habits and needs; to serve these wide-ranging needs, the SURFshare programme uses customised tools that support individual workflows.

Repository infrastructures are being upgraded to support the creation, storage, visualisation and exchange of Enhanced Publications. A common data model is used in the development of the customised tools required in the various EP projects. Eventually all Enhanced Publications will be aggregated in Narcis, the open access portal for scientific output in the Netherlands.

Another focus of the SURFshare programme is permanent access to research data. SURF started with Enhanced Publications, but quickly realised that this style of publication could not happen without proper data preservation and data access models. Licensing and related aspects play an important role in data access; furthermore there are several baseline criteria for an Enhanced Publication [ESEPRT] that are of interest to an AAI, such as recording authorship of the publication and its components.

A precedent for this kind of dialogue and cooperation between publishers and libraries has already been set in the area of Orphan Works, resulting in the 'Accessible Registries of Rights Information and Orphan Works towards Europeana' [ARROW] project and a proposal for a Directive on Orphan Works.

The challenges to be addressed in the establishment of a pan-European licensing initiative are:

1. Identifying the users of the service and thus working out what rights should be negotiated;
2. Identifying what content should be included, and thus who to negotiate with;
3. The identification of requirements of different nations and their researchers. Some countries may already have certain content covered by other strategies. Do all countries have the same sort of set up in terms of library structure;
4. Who pays for what;
5. Convincing publishers that they will benefit from extending their reach and their subscription revenues would not be reduced.

" The advancement of data sharing remains a big challenge. Researchers hesitate to publish data. This is a barrier for both national and international initiatives. This raises some hard questions, such as what licences should be in place? One proposition could be open access where possible, closed when needed."

Wilma Mossink
Project Manager, SURF

[6] Adapted from Ten Tales of Drivers & Barriers in Data Sharing

A pan-European approach to licensing is one solution to simplifying access to information resources via an AAI. Another solution is to use the model developed through the SCOAP3 project. SCOAP3 aims to convert to Open Access the peer-reviewed literature in the field of High-Energy Physics by re-directing, on a global level, funds currently used to subscribe to these journals. It is unclear whether this solution can be scaled up to other disciplines, but in principle the model proposed by SCOAP3 could circumvent the licensing issues that could prevent access to information resources through the AAI.

Requirements:

- Support access management policies to ensure that specific content is protected even when Open Access is offered;
- A support accounting mechanism for tracking changes to data post-publication;
- Offer permanent access to research data by implementing sustainable data preservation mechanisms and data access models.

Use-case 5
Multi-disciplinary Data Services

EUDAT: towards a European collaborative data infrastructure

Data are increasingly being considered as not just a product of research projects but as a fundamental constituent of scientific practice. New methodologies are being developed to discover unexpected phenomena based on large data sets combined from different disciplines. The Digital Agenda for Europe 2020 indicates the crucial importance of enabling and facilitating access to data across scientific domains.

EUDAT is one of the first EC-funded multi-domain projects to tackle the problem of the data deluge. The mission of the consortium is to contribute to the development of a pan-European data infrastructure for research communities from many different scientific domains.

Obviously, such a goal is facing the difficulty that, on the one hand, the scientific stakeholder communities have already invested much effort into either building their own AAIs or using existing identity federations and adapting their identity and authorisation management accordingly. Understandably there is a trend that to further use these existing systems, procedures and policies for federated access management, access control and accounting are needed.

Figure 2.7 outlines the ranges of community-specific, multi-community, domain-specific and cross-domain data services, suggesting that common requirements can be met by generic services on different cross-domain scales. It further indicates that authorisation services should be considered to be domain-specific, whereas authentication services should be as generic as possible.

During the first year of the project, EUDAT reviewed the approaches and requirements of a first subset of communities: linguistics (CLARIN), earth sciences (EPOS), climate sciences (ENES), environmental sciences (LIFEWATCH), biological and medical sciences (VPH), regarding the deployment and use of a cross-disciplinary and persistent data

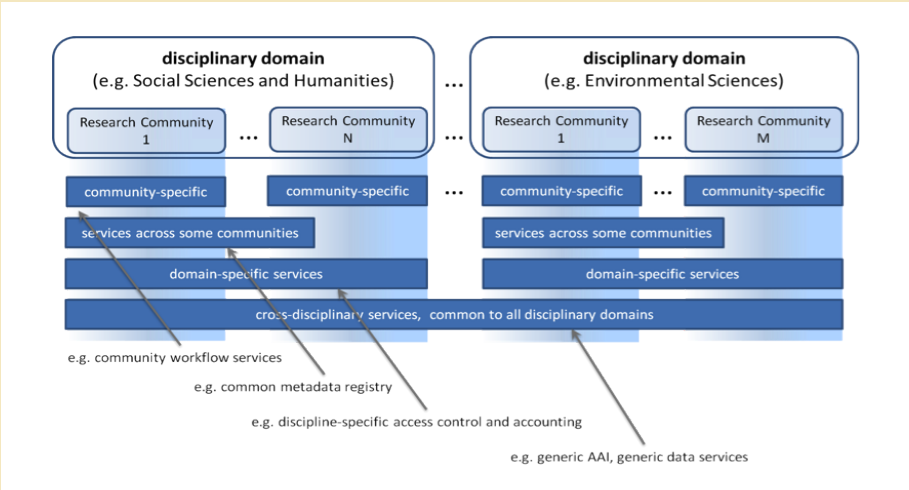


Figure 2.7: Generic data services of different cross-domain and cross community scales

e-Infrastructure. From these initial assessments, it is clear that there is a common need for data services in the area of **safe data replication** (replicating and safely keeping track of multiple copies of registered digital objects across different administrative domains for the purpose of persistency and distributed access), **dynamic data staging** (moving data to and from data processing and analysing facilities on demand), **joint meta data services** to enable cross-disciplinary research and a **simple store** as collaborative workspace for depositing, registering, enriching and sharing the long tail of small scientific datasets in an easy way.

Requirements:

- Separate Identities from authorisation: Use and support of various means for authentication, particularly from OpenID, X.509 (PKI) and Shibboleth identity providing services to serve both institutional and citizen scientists; base authorisation on attributes provided both by users' home organisations and research communities;
- Level of Assurance and Trust: concepts such as the **Level of Assurance** (service provider point of view: can I trust the means for authentication) and Level of Trust (user point of view: can I trust the service provider) must be introduced;
- Enable Single Sign-On: once an identity has been authenticated, it should be able to seamless access on any other service within the federation (unless different level of assurances are required by a specific service);
- Community-managed authorisation and identity credential delegation: The authorisation to deposit, access and enrich data should be managed based on attributes (including roles) provided by the communities;
- Accounting: enable the collection and exchange (with regard to technical and semantic interoperability and the legal policies) of accounting information with regard to storage and network usage across multiple administrative domains.

Use-case 6
Analysing Data

Facilitating new research environments: Göttingen State University Library

Göttingen State University Library is a central unit on the university campus. The university provides programmes across almost all academic disciplines and incorporates the Academia of Science. The University Library is also the State Library for Lower Saxony, houses the national collection of 18th century material, and is the biggest scientific library in Germany.

" A central process for authorisation and authentication needs to be agreed across centres and legal issues need to be harmonised. Researchers will not use these new research environments if they have to agree terms of use with each individual party."

Heinke Neuroth
Head of Research and Development
Göttingen State & University Library

The principle of subject collection is employed in Germany and the Library has seventeen special subject collections; the library collects relevant material from all over the world for these subject collections. The nature of the collections in the library attracts an international user community of researchers and scientists.

Because the collections are of international significance, both local and international researchers need access. Researchers are increasingly working collaboratively and there is a need for virtual research environments to help these groups to share software, disk space, and content. There is a need for a single access to these resources and an AAI to facilitate this approach. Göttingen is one of the main supporters of the Research Infrastructures Manifesto [DARIAH Manifesto] led by the DARIAH project, which champions the need for such an AAI.

Textgrid is an example of a Virtual Research Environment used by researchers at Göttingen that allows for Single Sign-On, but only for resources deployed by the project. This means that it is not possible to exchange research objects or have shared storage outside of the immediate infrastructure of the project.

Supercomputing centres in Germany are also considering a new approach to Single Sign-On. Currently every researcher must agree to the terms of use with individual centres. There are terms of use to be adhered to at every level - European, governmental, state and institutional - which are not easy to harmonise. These terms of use take data protection into account, as well as access to licensed content, creating a confusing picture of access rights and permissions across the centres.

Requirements:

- Access for both local and international researchers in a virtual environment;
- SSO that is not specific to single projects or platforms;
- Reduce requirements for researchers to sign up to multiple agreements.

Use-case 7
Accessing Data

Seamless access to research information resources and repositories: University of Edinburgh Library

The University of Edinburgh Library is part of the Information Services division at the University of Edinburgh. It has a data repository service called Edinburgh DataShare. University of Edinburgh researchers access library resources in a common way, using their institutional login.

The library uses both EZproxy and Shibboleth for authentication and authorisation. There may be more than one set of credentials behind this, but for the end-user the authentication experience is the same. Users log in to resources with their institutional account both on and off campus. An example of this joined up approach to authentication is the user experience of accessi Google Scholar. If users find an article through Google Scholar they can use the institutional link resolver and log in using their institutional credentials. This is not always possible when accessing content directly through a publisher's website where federated access has not been deployed.

The benefit of an AAI for University of Edinburgh researchers is that it would enable easier collaboration across institutions. The value exists in the provision of access rights to the same resources and in the ability to share information. Sharing disk space and software has been achieved with relative ease but an AAI for researchers also requires a new approach to the licensing of e-Resources.

An AAI could also help reduce the administrative work for authorisation of submissions to the institutional repository by allowing co-authors from other institutions to access the repository to update works. Copyright and IPR are a major barrier to achieving the full potential of an AAI.

Licensing is costly to negotiate with publishers and the solutions negotiated must be sustainable. The view of the librarians is that Open Access solutions are key to the success and sustainability of an AAI in the long term. This necessitates investment, incentivisation and a culture change amongst researchers and institutions.

Requirements:

- Simplified licensed conditions;
- Support for author identifiers;
- Support for open-access resources.

Use-case 8
Accessing Experiments and Data

User experimental access, data access and data analysis services: performing experiments at large photon / neutron facilities

There are more than a dozen photon / neutron Research Infrastructures (RI) in Europe. They count for over 30,000 users and offer world-class beam lines and instruments to perform experiments. Beam time at these RIs is so popular that there is always an overbooking of a factor 2-3. The users are coming from various disciplines, such as, life sciences, chemistry, surface and materials science, environmental science, as well as medical applications.

Around 40% of the users use different facilities in Europe for their experiments, which often have an iterative nature. This means that the user has to collect his/her data from different RIs before being able to analyse the data and finally produce a publication.

Users of the neutron / photon RIs submit a scientific proposal to apply for beam time at the specific RI of their choice. In these scenarios, an external review committee reviews the proposal made by the researcher. Once a proposal is accepted, beam time is allocated and the user comes to do the experiment at the RI. Scheduled beam time ranges from a few hours to weeks, depending on the type of experiment.

The amount of raw data produced by the experiments ranges from several Megabytes to Terabytes. Currently there is no common data policy regarding the storage of these data. The users take data home regularly. Recently, there has been an enormous boost in data volume, because of a new generation of photon light sources and because of novel two-dimensional detectors producing more data. This makes data transfer more complicated in view of a limited bandwidth between infrastructures and the users' home institutions, together with the limited computing and analysis capacities at the home institutions there is an increased interest in central services for storage, data archiving and data analysis of the recorded experimental data. Furthermore there is an increasing need from the community for online data analysis during the measurement.

A solution for these various aspects implies efficient tools for remote data access and remote experiment access, as well as a common pan-European authentication system as basis. The Umbrella project (Chapter 3) is working to provide a solution for this use-case.

Requirements:

- A European authorisation and authentication platform (addressed by the Umbrella project);
- Data policy on data storage (addressed by the PaNData Europe project);
- Data analysis centres;
- Remote data and experiment access tools;
- Immediate online data analysis.

2.7 Conclusions

The review of the various user community needs shows some overlap of requirements, such as the need for an infrastructure that eases collaboration whilst at the same time preserving ownership of data and authenticity. Mechanisms are needed to allow researchers to safely link their scientific results with initial data (sets) and with intermediate data to allow for future data re-use.

Research generated in such a manner creates complex intellectual and usage rights, meaning that sophisticated tools to enforce policies on how data can be accessed and processed are needed. In a highly distributed and dynamic environment; researchers (and communities) want to be sure that data held remotely are not compromised, not altered and remain under the users' control. At the same time, these rights need to be managed in such a way as to not become a deterrent to the use and reuse of such data.

As mentioned above in this chapter, the institutional survey showed that most of the users would like to use their institutional credentials to access services. This may reflect the increased penetration of systems like Shibboleth, which enable federated access. Challenges and opportunities to use federated-access technologies to support access to different applications are presented in Chapter 3 of this report.

Users are also using their social media accounts to share and access information and this trend should not be ignored.

The changing approach to mass digitisation and open availability of publicly funded content and resources also implies that libraries, as intermediaries, will need to evolve as they move from managing subscriptions towards enabling open scholarship and curation of data. Just as the library moves towards embedding itself in the research workflow by supporting researchers in managing their data, an AAI must also fit easily in to the researchers workflow in terms of facilitating easy depositing of data. An AAI for the SDI should mediate between Open Access and the need to protect some content for ethical or privacy reasons.

Access to licensed information resources is still a core requirement for many researchers. In fact, as researchers increasingly collaborate across and move between institutions, facilitating access to these resources is becoming more complex. Researchers naturally wish to share research content with their colleagues but licensing issues restrict this or a researcher may unwittingly end up breaching copyright. The AAI could facilitate more seamless access to these resources, but negotiation must also occur between publishers and libraries around licensing and access to these resources.

In summary, the AAI for the SDI should:

- Enable Single Sign-On across services;
- Empower researchers to deposit, share, use, reuse, and claim credit for their data easily and safely;
- Enhance authorisation to support attributes provided not only by users' home organisations but also by research communities;
- Allow the utilisation of the shared facilities of large datacentres for trusted data depositing and processing, with guaranteed data and information security;
- Support different types of credentials and provide ways to translate them to access different systems;
- Support simplified and seamless access to both licensed information resources and Open Access;
- Motivate researchers to share their research environment with, and open it to, other researchers by providing tools for instantiation of customised pre-configured infrastructures to allow other researchers to work with multiple data sets;

- Engender trust in repositories and data curation by protecting data policies, ownership rights, and by supporting persistent identification;
- Plan for flexibility to accommodate constantly and rapidly evolving research behaviour, which is becoming more collaborative, open and social;
- Help those needing and operating AA infrastructures to avoid duplication of expenditure on information resources.

3 Survey of the AAIs

3.1 Introduction

The development and deployment of Authentication and Authorisation Infrastructures (AAI) has taken place in different research and education environments as well as in the private and public sector. National Research and Education Networks (NRENs) have been developing and operating AAIs for over ten years; these AAIs provide services to a great number of users within the academic and research community. The Grid community has developed its own AAI. Cloud infrastructures are becoming increasingly popular and so is the need for a reliable and trustworthy AAI that can be used for public and private institutional cloud. Lastly, governments are trying to deploy an AAI to support business and public transactions.

An AAI is an infrastructure to verify a user's identity (authentication) and to verify that a user has the rights to access the service the user has requested (authorisation); often these infrastructures offer accounting mechanisms to determine how much resources users consume, to collect statistics data, and to record authentication failure and other diagnostics.

The overview and analysis provided in this section focus on the infrastructures currently used in the research and education sector, their underlying technologies and standards, and the use-cases they support. Some emerging technologies and infrastructures are also mentioned because of the impact they are expected to have on the existing AAIs. This report does not address accounting, because in the AAIs that were assessed, only limited facilities are available for monitoring and statistics.

3.2 AAI and NRENs: the Federated Approach

The need to access resources in different administrative domains in combination with the evolution of web technologies, collaborative and international research, and the increasing number of systems requiring authentication has imposed new requirements on access management technologies for education and research. Provisioning user accounts for each application that users wish to access does not scale well in a highly distributed and collaborative environment that crosses multiple administrative domains and national boundaries.

3.2.1 Identity Federations: AAI for the Web

The current best practice to meet this need is based on what is known as **Federated Access**, or **Federated Access Management**, or **Federated Identity Management** or simply **Identity Federation** (IDF).

An Identity Federation is an infrastructure in which:

1. **Authentication** is controlled by the user's **Identity Provider**, also referred to as IdP (typically the institution that the user is affiliated with), which verifies the user's identity and issues access credentials (i.e., username and passwords, and X.509 certificates);
2. **Authorisation** is controlled by the resource provider, also referred to as **Service Provider** (SP) or **Relying Party** (RP) that relies on the authentication done by the IdP and the information (attributes) received about the user from the IdP and possibly from other attribute providers within the Federation;
3. **Policy or legal agreements** are in place among the entities participating in the federation to achieve a trust relationship between the parties.

IDFs enable users to access applications and resources operated in different domains with the same set of credentials issued by the users' IdP. In other words, IDFs enable Single Sign-On: log in once to access multiple services. This model allows institutions to offer a richer service portfolio, reducing the need for bilateral agreements between each institution and each service. Typically, the federation operator handles the agreements for all parties participating in the federation and provides the technical support to enable the communication among all parties.

As reported in the TERENA Compendium and by the REFEDS group, the number of federated access infrastructures in the research and education community has been growing constantly since 2005. To date, the majority of the NRENs in Europe offer (directly or via a third party) federated access for their users. However, the level of deployment, the participation of institutions and the amount of services available via different federations vary significantly from country to country. For instance, not all research institutions, libraries and community datacentres are connected to national federations. IDFs particularly cater for users affiliated with an institution. Users without an affiliation (for example, because their home organisation has not joined an IDF) or users affiliated with multiple institutions or 'nomadic users' (i.e., persons who move from one institution to another), cannot be easily supported by IDFs at this point in time. The 'nomadic users' pose an interesting challenge to the Identity Federations, particularly when they tend to be identified with researchers; for instance, access to the researchers' publications or researchers' data may become unavailable to the owners when they move to another institution.

This problem could be solved with the introduction of a **persistent identifier** that would follow researchers when they move between institutions. Research in this area is being carried out by, among others, the Open Researcher and Contributor ID [ORCID] community. At the moment however there is no universally deployed solution to this problem.

Underlying Technology

The Security Assertion Markup Language, SAML2.0, is the open standard used to build IDF systems. The SAML protocol supports the secure exchange of authentication and authorisation data between identity providers and service providers or relying parties. In the research and (higher) education sector, the first SAML-based IDFs were introduced in 2005 by the NRENs, which have been driving the development of IDFs ever since. Well-known IDF products used in the higher-education sector (and beyond) are Shibboleth, the open-source software developed by Internet2, SimpleSAMLphp, the open-source community driven product developed by UNINETT, and the commercial Active Directory Federation Service (ADFS) of Microsoft.

A challenge to the current IDFs comes from the rapid development of user-centric technologies that are widely adopted by Web 2.0 applications such as social networks (i.e., Facebook and Google). Web users are becoming more and more accustomed to the access control management used in social networks. Increasingly these credentials are federated: instead of creating new user accounts, users can - in principle - use their

existing 'social credentials' to sign in to a range of commercial, third-party services. The penetration of social networking has increased the demand to use social network asserted credentials for libraries and/or university resources. In this scenario, rather than using their institutional credentials users would log in using their preferred social network credential. This model has some implications:

1. The identity vetting and the authority of the information associated with these identities are still a concern; practically those identities are self-asserted and therefore the level of trust associated with them is low. However, they could be used to provide access to services where authentication is not very important, but personalisation of the service is key;
2. These 'social identities' do not carry additional verified information regarding the role of the user (i.e., student, researcher); this information is particularly relevant in the research and education context, for example, to access services to apply for a grant, handle users' exams, and so on. This role can only be provided by the institution(s) the user is affiliated with or community-wide user attribute providers (the latter is gaining more and more support as an approach that is more scalable and future-proof). Therefore a mechanism to link the user's social identity with an institutional identity should be in place. This approach implies a further separation between authentication of the users and effective management of authorisation rights (attributes) and would require some security mechanisms to prevent inaccurate linking.

Trust Model

The trust model in IDFs has different aspects:

1. The **Relying Party must trust the Identity Provider** to authenticate the users as agreed and to ensure that the users' information (attributes) is up-to-date;
2. The **Identity Provider must trust the Relying Party** to process and protect any personal data received from the Identity Provider in a way that conforms to data protection laws;
3. The **users must trust their Identity Provider and Service Provider** to protect personal information.

Clearly the user's IdP plays a very important role in vetting the user's identity (typically when the user enrols), in updating the user's information (for instance, if the user enrols for another course, graduates etc.) and in de-provisioning the account when the user departs. However, this model does not address the needs of very dynamic and cross-boundary research, where researchers from several institutions working on the same discipline and participating in a research collaboration would like to manage additional users' attributes (for instance, to indicate their collaboration-specific roles, group memberships and authorisations). It is difficult to have these attributes maintained by the Identity Provider, because each collaboration group maintains specific information about the participating users that are unknown to the user's Identity Provider and **collaborations often have users from several Identity Providers**.

A complementary approach is that the research collaboration, also referred to as **Virtual Organisation**, maintains the additional user attributes and has related policies, processes and tools for assigning the proper attributes to the members of the collaboration. Technically, this extends the bilateral Identity Provider - Relying Party relationship to a triangle, where the Resource Provider uses an Identity Provider to authenticate the user, and subsequently fetches his/her additional collaboration-specific attributes from a server maintained by the collaborating community, which in effect acts as an Attribute Provider (AP). **Although it is generally agreed that the model above offers a solution to the problem, more research and development is needed to provide easy-to-use and multi-community group management solutions.**

Lastly there is not yet a standardised way among different federations to express that an institution has performed additional verifications on the users' identity (Level of Assurance). The REFEDS group is tackling this problem. Progress in this area will be driven by the increasing requirement for security of the services.

In summary, federated access brings a number of benefits, both for the users (reduced number of credentials, possibility of SSO), for the services (reduced bilateral agreements between the service and each institution) and for the institutions (more services can become available to the users without institutions having to operate them directly). Cost saving benefits have been demonstrated by using federated access, such as reduced usage of helpdesk to reset passwords, less overlapping of work to maintain users' accounts and attributes and reduced costs for users' administration for service providers.

Federated Access Management and the Law

Currently, the exchange of information that relates to the processing of personal data and the free movement of such data is regulated by the European Data Protection Directive, also called European Directive (95/46/EC); the directive was approved in 1995 by the European Parliament. This directive regulates the way in which personal data can be gathered in the EU, the rights of EU established citizens with respect to their personal data, and the transfer of personal data to non-EU (EEA) countries. The directive identifies three main players:

- **Data Subject**: an identified or identifiable natural person;
- **Data Controller**: the natural or legal person, public authority, agency or any other body that alone or jointly with others determines the purposes and means of the processing of personal data;
- **Data Processor**: a natural or legal person, public authority, agency or any other body that processes personal data on behalf of the controller.

Federated Access Management involves a complex exchange of personal information between people and organisations who may not have any direct relationship, often taking place across international borders (or even involving entities whose geographical location is unclear), and often including types of identifiers that were not contemplated in 1995. It is therefore inevitable that it will present challenges of interpretation and application of data protection laws that have their origin in the European Directive (95/46/EC) of that year.

Having educational organisations act as identity and service providers for their members, as will be required for e-Research, raises new legal issues because, unlike other sectors, the individual's relationship to the organisation is not just that of customer of an access management service. Any legal framework for access management in research and education must take into account the existing relationships and contracts associated with employment, education and the provision of services.

It is important to note that the European Directive leaves to the national authorities the freedom to apply the methods that best suit that National Member State to enforce the directive. This lack of clarity and the resulting variation in Member State interpretations make it difficult to exchange users' information (needed to grant access to a service) between countries as it is very likely that these attributes are considered personal data.

However, the directive indicates when personal data can be processed; particularly **Article 7** of the directive states that personal data can be processed if "*processing is necessary for the purposes of legitimate interests pursued by the controller or by the third*

party or parties to whom the data are disclosed, except where such interests are overridden by the interests for fundamental rights and freedoms of the data subject".

For the identity provider the legitimate interest is to support its members in obtaining secure access to services they wish to use; for the service provider, the legitimate interest is to provide the service that has been requested by the individual. This enables both identity provider and service provider to process personal data without having to determine whether the particular user is in a position to grant consent or whether the access is necessary for their employment/study.

Unfortunately, both justifications for processing personal data and relationships between organisations exchanging data have been implemented very differently in different Member States. **This level of divergence makes it hard to find a legal framework that will work in all Member States**; without such a common framework it seems inevitable that the establishment of international federated access management will be hindered by the different laws and expectations in different countries. **Consistent implementation and interpretation of the legal requirements across countries is essential; the revision of the European Directive should improve this aspect**.

Access management for teaching and learning generally involves three parties – the individual, their identity provider and their service provider – but research often adds a fourth party, the project, to further complicate the legal and technical arrangements. In research, it is common for access to resources such as computers, experiments and instruments to be granted to a particular research project (Virtual Organisation), with the project then determining how that access will be allocated among individual project members. Although some virtual organisations are long-term projects with significant technical and legal resources, there will also be many that are small groups of researchers who may, for example, decide at a conference to collaborate on a single research paper. Such ad hoc groups could obtain great benefit from a general-purpose e-Research infrastructure. However any requirement for heavyweight technical or legal arrangements will be impossible for them to satisfy. The legal framework for a successful e-Research infrastructure must, therefore, be based almost entirely on the existing relationships between individuals, organisations and services to avoid creating barriers to entry that these groups will find impossible to meet.

International Transfers

Many applications of federated access management in education and research involve parties outside the European Economic Area (EEA). Educational resources or teaching may be provided by publishers or universities in other countries, while research collaborations often include both researchers and instruments on other continents. Since these overseas participants are likely to play the same roles in education and research as their European peers, it is highly desirable to include them within a single legal framework and agreement, rather than having to maintain two (or more) different legal arrangements among partners who are otherwise treated alike.

Unfortunately, the options provided by the data protection law to achieve this are very limited. Under the current Directive, the only justification that can be used for transfers of personal data both within the EEA and overseas is the 'legitimate interest' of the Service Provider to provide the service as the legal grounds for attribute release. Despite what may be the common belief, user consent is not the appropriate solution for choices that an individual may be compelled to make to continue their employment or study. Since many current research partners are in the USA, the Safe Harbor Agreement might be an option for these, but it cannot be used by US universities as they are not covered by the relevant regulators.

Article 26(2) of the Directive permits transfers of personal data outside the EEA where the data controller "adduces adequate safeguards with respect to the protection of the privacy and fundamental rights and freedoms of individuals and as regards the exercise of the corresponding rights". Although 'legitimate interests' is not currently listed in Article 26(1) as permitting exports, it seems that it might, under Article 26(2), offer a single legal framework that could cover transfers of personal data both within and outside the EEA while ensuring adequate protection of those personal data.

Unfortunately, there is some divergence between Member States in the implementation of Article 26(2). The UK Information Commissioner encourages Data Controllers to base exports on their own assessment of risk, while other countries require that all such exports receive prior approval from the Data Protection Authority. Given the number of identity providers and service providers involved in e-Research, it seems unlikely that Data Protection Authorities would want to receive requests to authorise every such transfer. As with the relationship between identity providers and service providers above, a common approach that scales effectively to large numbers of relationships is essential.

From the European Directive to the European Data Protection Regulation

In 2012, the European Commission proposed a reform of the EU's 1995 data protection rules to strengthen online privacy rights boost Europe's digital economy and to address the current fragmentation resulting in the implementation of the European Directive.

The revised data protection framework, published as a draft in January 2012, will take the form of a Regulation; a 'Regulation' unlike a 'Directive' will be directly applicable in all EU member states without the need for national implementing legislation. The proposal that in the future the data protection laws should take the form of a Regulation rather than a Directive ought to increase harmonisation. However, the draft text of Recital 24 ("online identifiers ... need not necessarily be considered as personal data in all circumstances") seems certain to perpetuate the current, unclear and un-harmonised position.

The proposed Data Protection Regulation makes few changes to the relationship between the parties, although there is a concern that increasing the formal requirements on either of the parties or the relationship between them may make it impractical for small service or attribute providers to participate. The proposed Data Protection Regulation also foresees specific rules that may ease the transfer of personal data outside the EU but still ensure the best possible protection of users' data when it is exported abroad.

It is unlikely that the Data Protection Regulation will be in force before 2015; until then the existing Data Protection Directive will be the applicable law.

Table 3.1 summarises the strengths and the challenges related to Identity Federations.

Federated Access Strengths	Federated Access Challenges
Enables users to access a wide range of (Web) resources, using the same credentials; users log in and gain access to all resources that are part of the federation without being prompted to log in again at each of them. This approach is known as (federated) Single Sign-On (SSO).	Moving beyond web-based resources and addressing the specific needs of researcher groups (VOs) in terms of attribute management and delegation (particularly in the case in which attributes are provided by third-party attribute authorities).
Particularly useful for service providers, to relieve them from the 'burden' related to users' administration.	The deployment and the number of applications available in a federation vary from country to country; in fact not all countries offer a federation.
Good security: information exchange about users between IdPs and RPs takes place through secure channels. Furthermore, the IdPs guarantee that users' personal data are protected. In the event of compromised accounts, logs can be used to find out where the problem occurred.	Stepping up to allow stronger authentication verifications where necessary. Ensuring the appropriate Level of Assurance for provided credentials as defined by applications or resource providers.
Based on standard technologies to achieve cross-platform and cross-domain interoperability.	Working with emerging, possibly competing, standard approaches.

Table 3.1: Identity Federation overview

3.2.2 Interfederation: eduGAIN and Kalmar2

eduGAIN is an infrastructure developed in the context of the GÉANT project to enable trustworthy exchange of information for authentication and authorisation purposes among the GÉANT partners and other cooperating parties.

eduGAIN has been designed to address inter-federation, i.e., to enable users from one federation to access services provided by another federation. This approach requires an infrastructure that supports the exchange of information between different entities (often located in different countries) and a legal framework (such as a contractual agreement) in line with the Data Protection Directive to ensure that the users' personal data are securely handled.

eduGAIN policy [GN3-10-354] allows participant federations to interoperate on a partial and voluntary basis.

The eduGAIN policy consists of:

- eduGAIN Policy Declaration - Each applicant federation signs the eduGAIN Policy Declaration, which is a relatively short legal document that binds the applicant. This document addresses the most fundamental issues of the eduGAIN Policy Framework;
- eduGAIN Policy Constitution - This part of the policy defines the governance of the eduGAIN service and the requirements for the Participant Federations.

eduGAIN builds on existing national federations, therefore in order to participate in eduGAIN an existing infrastructure is needed.

The eduGAIN service offers a solution for Web Single Sign-On (WebSSO), which enables users to log in to multiple services, provided by different federations, using a single, one-step login process. There is a strong demand to extend the SSO to other applications and services areas. A typical example is a researcher who needs access to Grid-based services and scientific instruments that do not use web browser clients and protocols. Development is ongoing in the eduGAIN team to support these use-cases.

Figure 3.1 provides a high-level overview of the eduGAIN model, illustrating the basic eduGAIN components and their relationship. It is important to note that eduGAIN builds on top of national IDFs and that different IDFs can chose which of their entities can take part in the eduGAIN service.

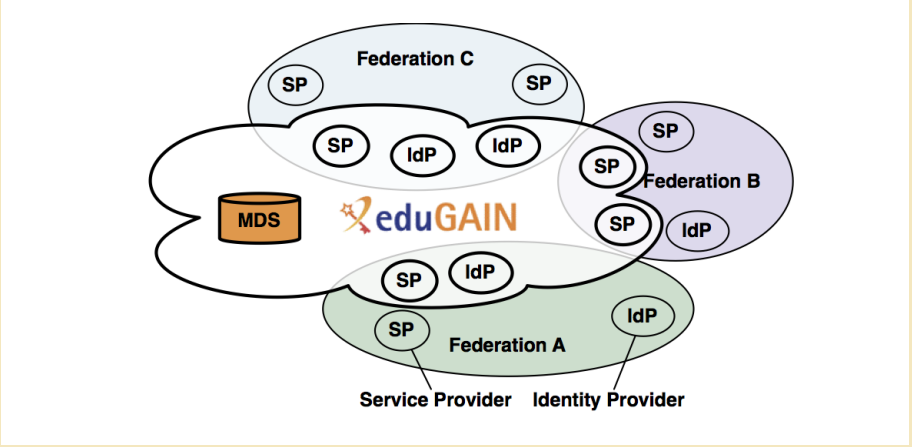


Figure 3.1: eduGAIN's technical architecture[7] (courtesy of eduGAIN)

An area addressed by the eduGAIN team relates to transferring Personal Data needed by services to authorise users, which typically means across national borders. As mentioned above, this exchange of users' information is regulated under the Data Protection Directive as well as by national laws.

The eduGAIN team, together with the REFEDS community, implemented a scalable solution (known as Code of Conduct) to provide the relevant information to resource providers to enable the service delivery without violating the data protection laws mentioned above. More information about the proposed eduGAIN Code of Conduct and the motivations for it can be found online [Code of Conduct]. eduGAIN has sought legal support to ensure that the Code of Conduct satisfies the requirements of the Data Protection Directive. The eduGAIN Code of Conduct implements **Article 7** of the directive.

The 'Kalmar Union' is an inter-federation collaboration with participants from the five Nordic countries. Any federations which act under EU legislation and comply with the Kalmar2 criteria may participate in Kalmar2. The primary goal of Kalmar2 is to serve research and education purposes and federations.

Kalmar2 has been operational since September 2009 and enables service providers to accept users from institutions scattered across the entire Nordic area, thereby broadening the use and scope of their federated technologies previously implemented.

The first demonstration of Nordic inter-federation was presented in 2006; Kalmar2 has later received financial support for development by NordForsk (Nordic Council of Ministers).

One critical design criteria was to support different federation architectures, because the Nordic area is highly heterogeneous in the ways that national federations implement the technical setup of their federations. One consequence has been the focus on SAML2, which has since then evolved to be the de-facto standard for Web Single Sign-On.

The policy framework developed for Kalmar2 provides a common denominator for the quality of the identity management of the participating institutions and sets a standard for protecting the users when releasing personal data across borders. Technical profiles were

[7] More information on eduGAIN technical architecture can be found online [DS3.3.1]

developed, some of which have later been adopted by the Kantara eGovernment Working Group[8] as the suggested standard for scenarios where interoperability must be ensured.

Operations of the central inter-federation services are done by one of the participating federations, thereby minimizing the overhead and ensuring that the required know-how is maintained and available.

Many Kalmar2 solutions, both at policy and technical level, have been re-used in eduGAIN, although eduGAIN decided to lower the entry requirements compared to Kalmar2.

Table 3.2 provides a summary of the interfederation approach; it is important to note that eduGAIN and Kalmar build on the existing IDFs and as such they inherit most of the features described in the previous section.

Interfederation Strengths	Interfederation Challenges
Infrastructure built on standard technologies to address the inter-federation requirement.	Extending technology and protocols to support non-web applications.
Privacy-aware environment complying to the EU Data Protection Directive.	Motivating entities involved in interfederation to meet data protection adherence.
Many important use-cases for the educational and research community addressed (in the case of eduGAIN as identified in the context of the GÉANT Project).	Extending membership and involvement from non-NREN federations.
Limited requirements for national federations to participate.	Requiremening harmonisation of higher Levels of Assurance.
Inclusion of multiple metadata streams provides federated access to more services offered by the participating federations.	Participating federations required to modify their metadata management practices to match inter-federation agreement.
Single identity from the home organisation is used to access a range of federated resources and applications.	Increasing the number of services available via interfederation agreements, by engaging more with relying parties.

Table 3.2: Interfederation overview

3.2.3 eduroam: Federated Access to the Network

The aim of eduroam is to enable federated access to the network: users with valid eduroam credentials can get online on any eduroam network in the world. eduroam is available in Europe, Canada, the USA, in some countries in the Asia-Pacific region (Australia, Japan, Taiwan, New Zealand and others), Africa and Latin America.

eduroam, which in 2012 celebrates its 10th anniversary, is the most successful example of a federated infrastructure used in the academic community (and in some cases is extended to other communities). The eduroam infrastructure is built upon two main technologies: the IEEE 802.1X authentication standard, to securely handle users' credentials and a hierarchy of RADIUS proxy servers, to transport users' credentials.

Figure 3.2 depicts how eduroam works when a user coming from University B (unib.nl) in the Netherlands tries to get connected to the eduroam network of University A, still in the Netherlands. Upon successful authentication of the user, which takes place at the user's home university (University B), the user can get online.

The hierarchical model followed by eduroam, mimicking the DNS hierarchy, has issues with domains that do not fit into this model, like the ".org" or ".eu" or ".edu" domains that are used by some organisations. RADIUS over TLS, an IETF standard since May 2012, offers a solution to this problem. The new features in RADIUS over TLS allow the use of a more dynamic trust model, where connections can be established between the users' home institutions and the visited institutions.

[8] http://kantarainitiative.org/confluence/display/eGov/Home

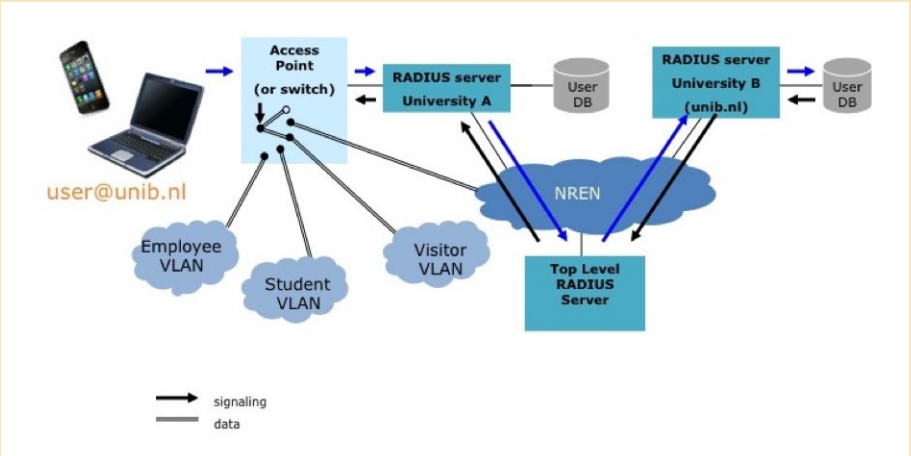


Figure 3.2: eduroam architecture (courtesy of SURFnet)

The eduroam service in Europe, which operates in the context of the GÉANT project, has evolved into a confederation: a federation of federations. In each country in Europe, an organisation (usually the NREN) is responsible for the national eduroam operations.

A European eduroam Operational Team is in place to ensure that international roaming can take place. Collaboration and coordination with other countries has initially been rather informal. Because of the wider deployment and growing use of eduroam, the community has requested a firmer basis for eduroam governance worldwide. The Global eduroam Governance Committee was constituted in November 2010 and at the moment comprises eleven senior representatives for Europe, North America, Asia-Pacific, Latin America and Africa as well as a TERENA-appointed, independent expert. TERENA operates the secretariat.

Table 3.3 summarises eduroam's main features:

eduroam Strengths	eduroam Challenges
Infrastructure to offer secure access to wireless (and wired) networks based on well-established and stable protocols (802.1X and RADIUS).	Extending beyond the educational domain and NREN and university federations.
Privacy preserving technology (users' personal information is not forwarded to the visiting institution).	Addressing fine-grained authorisation requirements.
Offers access to all eduroam networks with the same credentials (no need to request additional credentials when moving to another institution).	Deploying solution to address the current limitations inherent with RADIUS.
Scales to a large number of connected institutions and works on all kind of portable devices.	Improving the consistency of the user experience when configuring eduroam on different devices.

Table 3.3: eduroam overview

3.2.4 Project Moonshot: Federating Non-Web Applications

Project Moonshot is a project originating from Janet, the NREN in the United Kingdom, that aims to address solutions to enhance Identity Federations as they operate today. Moonshot aims to provide a solution to the following problems:

1. **Support for Non-Web Applications**
   Even though Web browsers provide a de-facto interface to the majority of Internet services, many applications are either not Web-based or are more effectively used through a native application. Examples are Outlook access to an IMAP server, Shell access to High-Performance Computing clusters or access to chat and calendar services.

It is worth noting that the problem of providing federated access to both Web and non-Web applications is also being investigated outside Moonshot. There are several other approaches based on the usage of secure tokens, such as OAuth, OpenID Connect, WS-trust, OpenStack and others. Because these initiatives are not, for the most part, led by academic communities, and are not primarily focused on federations, they are not covered in detail in this report.

2. **Addressing scalability issues in discovering the Identity Provider of a user**
   Federations that nowadays have hundreds of Identity Providers are struggling to offer the user a convenient interface for selecting his home organisation Identity Provider. If federations couple with other federations, this problem only gets bigger.

3. **Support for multiple affiliations and federations**
   A user can at the same time be a student at one school and a teacher at another, can both be a teacher and a parent, and can belong to the identity federation of the research network as well as to that of a professional or societal organisation. Current trust fabrics have difficulty distinguishing these roles when contacting the appropriate Identity Provider for assertions about the user.

Project Moonshot addresses these issues by proposing a new architecture (Figure 3.3) that builds on the foundational technologies of two successful federated identity infrastructures described above: Identity Federations and eduroam.

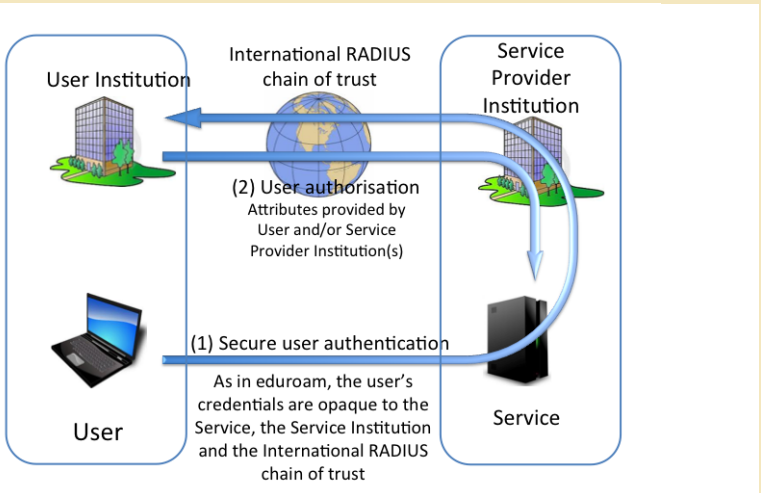


Figure 3.3: Moonshot architecture

Project Moonshot combines the eduroam trust model (built using RADIUS servers) with the Extensible Authentication Protocol [EAP] authentication standards for confidentiality of user credentials and SAML standards for the exchange of information about the user. In addition, a standard API (GSS-API) is used to provide a common interface to applications. This makes for a scalable solution that can be used for both Web and non-Web applications. Project Moonshot also aims to provide solutions to the requirement for multiple affiliations and to define a model 'trust router' that performs a selection of potential trust 'paths' between an Identity Provider and a Relying Party.

The technology developed in Project Moonshot is being standardised in the Application Bridging for Federated Authentication Beyond Web-SSo [Abfab] Working Group in the IETF.

At the time of writing, Project Moonshot is still in a pilot phase; the success of the initial pilots seem to indicate that Moonshot production is not too far from its goal.

Table 3.4 provides a summary of Moonshot features.

Moonshot Strengths	Moonshot Challenges
Attempts to solve problems with accessing cross-domain non-web resources.	Requires updating of the standards on which it is based, which may create difficulties for wide adoption.
Builds on long-term AAI community experience and well-defined goals.	Moving the communities forward with new developments is challenging.
Is an IETF proposed standard track activity. Benefits from the well-established IETF standardisation process.	Needs development of new software to be widely adopted.

Table 3.4: Moonshot overview

3.3 Grid Infrastructures and Research Communities

The past decade has seen the emergence of an e-Science infrastructure in Europe, encompassing resources from a large number of different providers and concurrently used by many different research communities. This type of e-Infrastructure, commonly referred to as the term 'Grid' is also used to refer to any distributed infrastructure that is federated to combine resources from multiple organisations managed by different administrative domains.

The Grid aims to coordinate the sharing of resources in a dynamic and multi-institutional setting to provide additional functionality beyond its constituent parts: brokering, workflow coordination, integration of computing and storage. In order for this to happen, interoperability and standards need to be defined at various levels: for resource access, for coordination and business logic, for data storage and management, for network access and so forth. Given the intent of the infrastructure to span multiple organisational domains, two areas in particular have attracted attention:

- **Technology** – authentication, authorisation and accounting protocols, mechanisms to organise users in communities and express collective attributes like membership and roles, and a resource access methodology that emphasises local resource autonomy and independence;
- **Policy** – authentication assurance levels, quality of identity vetting, traceability, acceptable use, and incident response.

The provisioning of collective services in the Grid has been the driving force for one of the most characteristic aspects of the Grid AAI: delegation of credentials to agents and software services. In order to provision collective services, participating agents need to authenticate in order to deliver that service to the user. Given the potentially large number of agents involved, and the current lack of a single administrative domain to manage these agents, they agents require credentials to authenticate amongst each other.

3.3.1 Public Key Infrastructure and Certificate Authorities

At the technical level, users and resources in the Grid today are identified using a public-key infrastructure (PKI), based on a trusted third-party scheme of Certification Authorities (CAs).

The user is either explicitly issued with such a certificate, or can obtain one as needed based on another form of authentication, e.g., by logging in to a (academic) federation or by using an institutional account. The digital certificate binds the user or the resource 'name' to the data used to prove possession of this private key, thereby providing them a

digital identity. The names thus issued are the key to access control (and data ownership) in the Grid system and must be assumed to be unique within the scope of a Grid.

In recent years, several such systems have been deployed by the research and education community; examples of these services are the Short Lived Credential Service [SLCS] developed by SWITCH, and TERENA's Certificate Service [TCS].

Although access to a resource can be based on explicit user lists, such an identity-based access control system does not scale. Grid software has therefore introduced VO-centred credential providers that can assert community membership, sub-groups, and roles of the user. The most-used mechanism for expressing VO membership is the Virtual Organisation Management System [VOMS], expressing VO attributes either as RFC3281 attribute certificates (ACs), or as SAML statements. A user would present an identity certificate as well as a VOMS statement linked to this identity in order to gain access to the service, and the access control decision is then primarily based on the VOMS.

The user can now direct the service to perform a specific task, i.e., coordinate a workflow or compute and store the results elsewhere. In order for the service to perform these tasks, it has to be able to act in coordination with other services. At the end of a computational task, the resulting data set must be stored elsewhere, or the broker will have to split the task in many individual jobs to be executed all across the Grid. Except in some specific cases, the service should prove that any such actions originate with the actual user request. This is done using a delegated credential that carries the verifiable identity and attributes of the original user.

The the most-used delegation method in the Grid is an extension of the PKI model, in which the user can 'extend' the certificate path with 'proxy certificates' signed by the user credential, and thereby certify that a specific service can act on his or her behalf. These proxy certificates carry the original identity of the user, but can in addition hold attributes and assertions by others, e.g., credentials issued by a particular research collaboration.

Trust Model

Collaborating across different domains can only work when the organisations involved can ensure the integrity of their own infrastructures, and their own policies and procedures are adhered to. This means establishing a policy and procedural framework around the AAI for authentication, traceability, incident response and accounting, and the production Grid infrastructures both in Europe and elsewhere have established such a framework.

The European Policy Management Authority for Grid Authentication in e-Science [EUGridPMA], working with equivalent bodies in the Asia-Pacific region [APGridPMA] and the Americas [TAGPMA] in the International Grid Trust Federation [IGTF] at the global level, provides persistent and unique naming of all users and resources in the Grid. The assurance level is sufficiently high that the names can be used to determine long-term authorisation decisions, as well as to assign ownership of data in long-term storage archives. The minimum requirements for joining the IGTF as an identity provider are driven directly by the security needs of the participating Grid infrastructures (relying parties or RPs), which include EGI.eu, and PRACE-RI in Europe, but also Open Science Grid, XSEDE, NAREGI, PRAGMA, as well as most national e-Infrastructures around the world. The RPs are directly represented as stakeholders in the trust federation and, jointly with identity providers, ensure the integrity of the trust fabric.

A policy framework that is, for the most part, common between all major infrastructures has been established to address community (VO) membership enrolment and management requirements, a harmonised incident response process, and a common end-user acceptable use policy (AUP). This common set of policies, which is fostered by the Security for

Collaborating Infrastructures [SCI] group, allows users of the e-Infrastructure to enrol once and gain access to a global set of resources without the need to register and accept the policies and conditions of each and every participant. Grid resource providers accept the common AUP and VO policies as sufficient, relieving the users of the need to sign large sets of terms and conditions.

In authorising access to resources, the research collaboration plays a very specific role: it is the research collaboration (and not the user's identity provider or 'home organisation') that determines the roles, memberships and rights of the user in a particular session. The authorisation decision at the site is therefore based on a combination of identity (for individual traceability) and attributes relating to the collaboration, coming from **different authoritative sources**.

With the general advancement in AAI technologies, the users' perception of certificate-based access control changed significantly. More and more existing and potential new user communities look at certificate-based access as a barrier to accessing resources, which is only partially addressed by the portal policy and current federated certificate systems.

As confirmed by the Federated Identity Workshop, organised under the auspices of the European E-Infrastructure Forum [EEF], there is consensus to work towards increased use of Federated Identity Management [FIM] and federated services in Grid infrastructure, although there are technical as well as policy issues still to be addressed.

Privacy Aspects

The minimum requirements for identity vetting in the IGTF federation today include face-to-face identity validation, and related controls address the need to provide globally unique and long-term persistent names (subject identifiers), as well as traceability to individuals.

Sufficient information about the participants must be recorded and retained by the identity provider in order to allow for the renewal of credentials. Since re-issuance of the identifiers to assuredly the same individual must be possible both after expiration of the certificate and following a revocation, this involves retention of identity information outside the digital domain.

Although this does imply a trade-off between user privacy and the resource provider's need for traceability (especially for data storage), regulatory requirements and the need to respond to incidents in a high-value, high-risk infrastructure have, up to now, excluded the use of anonymous or pseudonymous identifiers in the Grid. Data confidentiality for end-users can be ensured by existing technical means, using distributed key management systems.

Grid AAI Strength	Grid AAI Challenges
Support for strong authentication (identity of users generally requires face-to-face vetting).	Managing large numbers of digital certificates raises usability issues; approaches to improve the user's experience using digital certificates are being proposed.
Separation between authentication (based on digital certificates) and authorisation (based on attributes controlled by the research collaboration).	Ensuring that attributes maintained by different research collaborations are not community-specific.
Grid infrastructure is cross-border and it can be considered as an example of inter-federation.	Leveraging identity federation infrastructures with Grid AAIs (IDF to identify and authenticate Grid users and Grid infrastructures to rely on the IDF authentication process).

Table 3.5: Grid infrastructures overview

3.3.2 The EGI approach

The European Grid Infrastructure [EGI] is the result of pioneering work that has, through the federation of national resource providers over the last decade, built a collaborative production infrastructure of uniform services, which supports multi-disciplinary science across Europe and around the world.

Today EGI is an e-Infrastructure for e-Science, supporting over 21,000 researchers across many fields of research with a wide range of technical and infrastructure services. EGI's services are distributed across Europe and beyond over more than 350 resource centres, supporting in excess of 1.2 million computing jobs per day, as well as data storage, transfer and open access.

In EGI, users[9] in research collaboration are organised in virtual research communities (VRCs) and virtual organisations (VOs), which are logically distinct from resource providers. A VRC may contain a number of science-specific VOs using different (and even overlapping) sets of physical resources. Although access to a resource can be based on explicit user lists, such an identity-based access control system does not scale to the global research collaborations needed within EGI. EGI has therefore introduced the Virtual Organisation-centred credential providers that can assert community-membership sub-groups, and the roles of the users, based on their personal X.509 certificate issued by an IGTF accredited Certificate Authority. VOMS, is used to retrieve VO membership information.

The user can access a service hosted by one of EGI's resource centres by presenting the identity certificate as well as a VOMS statement linked to this identity. The service uses the VOMS statement to perform the access control decision. In line with the Grid model, the service performs the requested task(s) in coordination with other services. Delegates' certificates are used by the service to demonstrate that all the actions originate with the actual user request.

The functional services (workflows, job execution, file movement and storage) provided by the resource centres or the research community are supplemented once by services provided by EGI.eu and its distributed partners on behalf of the whole community. These coordination services include providing a federated helpdesk to help with user support queries that may span multiple organisations and integrating the use of functional services by individuals from a resource centre, national resource provider or research collaboration perspective.

EGI has recently undertaken a study [EGI-AAA-Study] on the use of traditional and new technologies for access control and identity management. The EGI study concluded that federated identity management could facilitate access to its resources for users if broader access to federated identity systems could be achieved within EGI's targeted user communities.

The broad adoption of federated identity management technologies and solutions within European research communities still provide some technological challenges for EGI in their adoption to access EGI's resources. EGI is exploring where it needs to invest in order to achieve the wider and more harmonised adoption of federated identity management systems on EGI and between NGIs.

EGI Strengths	EGI Challenges
Standards-driven independent authentication token.	High-value authentication token (X.509) provides an entry barrier to too many research communities.
Globally scalable community-driven authorisation model.	Integrating the Grid model with the IDF model requires further research.

Table 3.6: EGI overview

[9] The EGI Case Study collection compiles a growing number of examples of the benefits of grid computing to science and research: http://www.egi.eu/case-studies/

3.3.3 PRACE: Access to European Supercomputing Facilities

PRACE, the Partnership for Advanced Computing in Europe, is a project co-funded by the EC which started in 2008. PRACE offers world-class computing, data management resources and services open to all European publicly-funded researchers. The need to maximise the usage of the facilities and to minimise the costs has necessitated a distributed Research Infrastructure. No single site can host all the necessary systems required because of limitations of space, power, and cooling facilities.

PRACE has a mixed AA infrastructure. The basic service model enables users to have interactive access to sites on which they can run jobs. In this case, a user's access is handled on a per-site basis and requires SSH (Secure Shell) credentials; each set of SSH credentials can only be used for the site for which they have been issued.

However, PRACE also provides remote Single Sign-On features, implemented via the GSISSH protocol (the modified version of OpenSSH to support the Grid Security Infrastructure) and the usage of personal certificates (X.509 certificates), which must be issued by CAs accredited by the IGTF. Besides GSISSH, the PRACE infrastructure includes other services like GridFTP, UNICORE (Uniform Interface to Computing Resources), a helpdesk, and accounting. GSISSH and the other services enable users to run jobs on different systems and to transfer data between sites.

Information about users, needed for the operation of the integrated services (those available via the SSO), is managed in a shared LDAP-based repository, where each partner only has access to its own domain, but partners can access all information stored in the database for infrastructure management purposes. This approach defines the basis of the PRACE AAI federation: each partner relies on and trusts the information provided by the others.

In addition to basic account information, the repository contains further additional attributes for authorisation purposes; for instance, it gives information about which systems the user can access to run jobs. The attribute information is based on IETF standard schemas and a PRACE-specific schema that includes attributes specific for the PRACE application domain. This means that when interoperating with other AAIs (for instance IDFs) the information received from those AAIs needs to be complemented by the attribute information from the PRACE AAI.

PRACE Strengths	PRACE Challenges
Builds on long-term Grid AAI experience and well-defined goals.	Integration with AAIs requires specific PRACE attributes.
Requires strong authorisation before X.509 certificates are issued.	General usability issues related to the usage of personal certificates.

Table 3.7: PRACE overview

3.3.4 The Umbrella Project

Umbrella project is the identity system being developed by PaNdata (Photon and Neutron Data Infrastructure) for the users of the European Neutron and Photon sources. Umbrella supports more than 30,000 visiting scientists that, twice per year for a limited period of time, perform their experiments using neutron/photon facilities. The management of these experiments is handled by 'local user offices', a few people at each facility who enable scientists to access the facilities (i.e., by providing necessary support for registration, stockroom, computer accounts, and storage space for experimental data).

Umbrella is a pilot authentication/authorisation infrastructure whose goal is to federate the local user management systems used in the photon community of the participating facilities.

The demand for federated AAI in this community is triggered by an increased need to access remote services, especially remote data access and remote experiment access. Because of the highly competitive type of research, the AAI needs to offer support for confidentiality, fine-grained access control, identity uniqueness and persistence, and also needs to be user-friendly.

The main feature of the Umbrella AAI architecture is that it uses only one IdP, which is used by all facilities to verify the identity of the user; the rest of the information related to the users (which is needed for authorisation purposes) is stored in the databases of the local user offices. The Umbrella IdP contains references to the location (local user offices' databases) from which to retrieve the rest of the information of the users. The main function of this single IdP is to guarantee a unique user identification and therefore a unique and persistent user identifier; this identifier is then used across all facilities.

The communication of the various elements of Umbrella (the Umbrella central IdP and the resource providers) is based on Shibboleth/SAML2.

Implementation within the photon/neutron community is scheduled for early 2013.

Umbrella Strengths	Umbrella Challenges
Serving short-term needs in a scalable way.	Managing the Umbrella IdP in conjunction with local user offices.
Ability to offer and control fine-grained access.	Managing user expectations regarding the reusability of credentials.

Table 3.8: Umbrella overview

3.4 Cloud Infrastructures

Cloud computing technologies are becoming a common way of provisioning infrastructure and services on demand, possibly combining computing and storage facilities as well as dedicated network infrastructure. The most successful form of cloud computing for the general public is Software-as-a-Service, which implies outsourcing some functionalities (e.g., emails, calendars, and storage) to a cloud provider (e.g., iCloud, and Google). A number of commercial cloud providers offer computing services on demand, ranging from virtual machines to platforms for developing user applications.

There are two main types of cloud services: public clouds, typically operated by commercial companies (Amazon, Rackspace etc.) and private clouds, typically operated by a specific user group. At the beginning of 2012, a consortium of IT providers and three Europe's biggest research centres (CERN, EMBL and ESA) announced a partnership to develop Helix Nebula, a European cloud computing platform.

The main benefits of public clouds are on-demand services with pay per-use that do not require users and organisations to own hardware or build their own infrastructure, and offer the possibility to dynamically scale resources required for solving specific tasks. The distributed character of cloud resources means that tasks and applications can run anywhere in the world depending on the physical spread of the infrastructure of the cloud provider.

Security Considerations in Cloud Infrastructures

Cloud technologies are based on hardware virtualisation, which allows for the management of virtual computing resources (scaling, migration, reconfiguration) independent of the applications layer. In theory, cloud-based virtualised applications should run in the same way as non-virtualised applications; in reality, in many cases, moving applications to the clouds requires their redesigning to support dynamic deployment and configuration.

Cloud-based service virtualisation provides an additional level of security due to the separation of applications executing environments, and the possibility of using pre-configured, security-enhanced virtual machines.

A concern with public clouds relates to data security, as in most cases the data is stored 'in the cloud' on servers whose location is conceptually nebulous. Typical questions users ask are "Where are my data? Are they protected? What control has the Cloud Provider over data security and location?", and also, "Who has access to my data? Is the usage statistics collected and how it can be used?"

The distributed nature of cloud computing, in which backup servers and data can be located potentially anywhere causes problems in understanding which of the national data protection laws apply to the cloud concerned and its usage – something often defined by a click through the terms of use, which are often not fully comprehended by the end-user.

Identity Management and Cloud Infrastructures

Clouds are becoming more and more popular among researchers, as they allow them to quickly obtain necessary computing facilities when they are needed (and often without the additional procedures to access organisational or community Grid resources). Research organisations are also using more and more virtualisation platforms offered 'in the cloud', on which to build other infrastructure, services and applications that are specific to a community or a science. At the same time, NRENs are defining strategies to offer cloud services, in some cases by contracting commercial providers. In establishing these arrangements, NRENs are seeking a model that allows cloud services to be used without compromising the AAI arrangements already offered to researchers and research organisations.

In moving out into the cloud, researchers are also moving outside of the AAI that has grown up to support their workflow within the federated and Grid spaces. Many cloud offerings are actually a step backwards in terms of access management and group management, which are specific to the commercial offering. Researchers may find themselves unable to make use of institutional credentials and unable to connect their entire research group to the cloud service without explicitly asking other users to sign up for a new service.

In order to optimise the benefits of cloud approaches, researchers and research organisations will need to review their identity management approaches and decide how much of the AAI infrastructure can be comfortably virtualised, outsourced and managed by third parties. Cloud providers are increasingly talking about Identity-as-a-Service offered as part of the cloud package, and this will challenge all of the approaches to identity discussed in this report. Cloud-type identity services will emerge in two ways – via organisational identity services fully outsourced to cloud providers and through the already established use of social identities as credentials for a range of services.

The added complexities of identity management within a cloud environment have been recognised in the development of new standards to support such processes. The System for Cross-domain Identity Management (SCIM) specification has been introduced to tackle specific workflow problems for cloud identity. SCIM does not specify any particular authentication or authorisation schema, but instead specifies a way for a variety of known endpoints (directories, group management systems, required services) to be seamlessly and easily linked together to provide an AAI. In this sense, SCIM is addressing the provisioning side of the identity management workflow. SCIM is a very new specification, but as the identity space grows it is likely that this type of cross-walking approach to AAI may become more popular and indeed necessary.

In summary, the main issues that should be addressed to make cloud environments and cloud-based infrastructures secure and trustworthy for a wide range of scientific applications are:

- Standards to facilitate interoperability;
- Secure operation of cloud infrastructure in line with national data protection laws and directives (typically addressed by cloud providers);
- Clear and explicit terms of use and licensing for cloud services;
- Integration of cloud-based infrastructure and access control services with existing AAIs.

Legal Aspects of Cloud Computing

In 2011, JISC Legal produced a report that surveyed the legal issues arising from the adoption of cloud computing services [JISCLegalToolKit]. The report concluded that there are no legal outright bars to the adoption of cloud computing solutions; in principle cloud computing services are similar to any other service. One major difference in using a cloud provider arises from the flexibility and movement of data between servers that may be located in various parts of the world. This makes it difficult to identify which law applies, at any given time, to the data.

As a general rule when using a cloud service, an institution will usually be the data controller responsible for compliance with the data protection laws when processing personal data, and the cloud provider will be the data processor. The cloud provider, as data processor, should act in accordance with the agreed terms under the contract with the institution in order to ensure compliance with the data protection laws. The implications of this rule are that data protection compliance remains an institutional obligation in the cloud. Institutions should carefully assess cloud contracts and service conditions to ensure that data protection requirements, data security issues and responsibilities are covered.

In summary, institutions that fail to incorporate the appropriate clauses into their agreements with cloud suppliers could find themselves facing actions for a breach of the law for the failure to impose appropriate obligations on their outsourcing supplier. Although many suppliers have signed up for the US/EU Safe Harbour scheme, their compliance with the scheme should be made clear in the terms and conditions of the services offered.

Cloud AAI Strengths	Cloud AAI Challenges
Ability to view AAI as a commodity service.	Need to address common concerns about data security – both stored data and personal data related to user identity.
Cloud-based service virtualisation provides also an additional level of security by separation of application executing environments and a possibility to use preconfigured security-enhanced virtual machines.	To exploit the potential benefits of cloud virtualisation, new trust and security management mechanisms need to be developed.
Virtualised approaches to cloud AAI may be potentially cost-efficient for organisations.	Architectures and models for access control and trust management in clouds are still evolving. The cost models of public clouds, however, is not yet fully clear and more investigation is needed.
Cloud providers can guarantee a high level of availability, data recovery and security of their infrastructure and platform.	Need to balance integration with campus infrastructure and legacy applications.
Major cloud service providers (such as GoogleApps, Amazon) either use or plan to implement SAML-based federated access to user-deployed infrastructure or applications.	Privacy issues related to the way these providers handle users' data and/or stored data.

Table 3.9: Cloud AAI overview

3.5 AAI for e-Government: STORK

Whilst much research activity happens across institutional infrastructure, with researchers affiliated to educational organisations as part of their work, there is an increasing need to address and interact with the AAI requirements of a non-affilitated citizen. A consistent message from all stakeholders spoken to as part of this study is the challenge of managing users that do not have a natural home organisation to provide credentials and to manage identity information, and the overlap of identities held by researchers as part of their citizenship.

In 2008 the EC funded the Secure idenTity acrOss boRders linKed [STORK] project to achieve the pan-European recognition of electronic identities (e-IDs) among Member States. STORK aims to reduce the barrier encountered by citizens to access public services while working or living abroad. By developing a system that recognises electronic identities and enables citizens to use them to authenticate to different systems, business and several public services will become more accessible.

The goals of STORK are:

- To define a common service architecture allowing citizens to use their national e-IDs to access e-Government portals across borders;
- To offer a platform for safer online communication using e-IDs for children;
- To offer a service facilitating students' mobility across Europe;
- To use of e-ID for cross-border electronic delivery for citizens and businesses;
- To test the electronic process of address change for EU citizens that move to other Member States.

The pan-European infrastructure developed by STORK has been designed to support different national e-ID systems (and policies), rather than asking Member States to adapt their solutions.

To interconnect the e-ID infrastructures of the Member States, a single Interoperability Framework, based on two basic models, is used:

- **Middleware Models (MW)** – In the middleware model the Service Provider uses software components (a middleware 'SPware') that implement direct communication with the foreign e-ID token. The citizen communicates directly with the foreign Service Provider and no intermediaries are in between;
- **Pan European Proxy Services (PEPS) model** - PEPS acts as a protocol gateway for countries that use different technologies for their national e-IDs and also as an intermediary for foreign e-IDs towards its domestic Service Providers. This model foresees a PEPS in each country. The PEPS model resembles the original eduGAIN architecture (based on bridging elements) which was abandoned (in 2009) when operational experience highlighted its scalability and user-interface problems.

SAML2 is the underlying technology, which allows for inter-operability with other infrastructures used, such as those used in the R&E sector (i.e., Shibboleth or eduGAIN).

STORK assumes the user is central to all operations; explicit user-consent is required before sharing mandatory or optional attributes across borders. The use of consent however raises some concerns, because, in federated infrastructure, consent is not freely given and cannot therefore be used. eduGAIN (in light of the feedback received by the Article 29 Working Party [WP29]) decided to use 'legitimate interests' for attribute release. In 2011, the analysis of STORK, carried about by the Article 29 Working Party [WP29-STORK] identified possible issues regarding which bodies collect, process and store which data in STORK in the PEPS model.

In 2011 IEEE published a document [ISBN 978-1-4577-0458-1] that identified key points that the STORK consortium and stakeholders should resolve in order to make the STORK security and privacy framework more robust; some of the issues relate to PEPS security, trustworthiness of providers that issue digital certificates.

Some of the Research and Education national federations are also participating in STORK and information is exchanged between eduGAIN and STORK.

The STORK project was concluded at the end of 2011; in 2012, STORK2, the follow up project started and work is ongoing to offer more pilots.

STORK AAI Strengths	STORK AAI Challenges
Available to all users regardless of affiliation.	Addressing consistency of user experience in a diverse environment.The usage of national ID cards varies from country to country in Europe and therefore not all countries can currently benefit from STORK.
Simplifying access to a potentially large number of services and portals.	Ensuring take-up by services internationally. The usage of digital certificates and the additional related process to manage them may hinder this process.
Ensuring services are offered in line with the privacy laws.	The current approach based on user-consent needs to be reviewed in line with the recommendations of the Article29 Working Party.

Table 3.10: STORK overview

3.6 Data Infrastructure: the Vision for EUDAT AAI

Sharing data infrastructures and enabling collaborations both require federated identity management and access control. As an infrastructure service project, EUDAT decided to make use of the existing AAI solutions, services and policy frameworks rather than building a dedicated solution. This approach ensures that users can rely on existing infrastructures and procedures they are familiar with.

Most of the communities involved in EUDAT, including **citizen scientists** (who are not directly affiliated to a research organisation, but who usually rely on an ID card issued by or in behalf of national authorities or on social network credentials), have established procedures to providing access to applications (e.g., Web bases, portals, and command lines). Some use X.509 certificates, some use OpenID (usually with a restricted set of providers) and some use the academic identity federations (i.e., Shibboleth). In each case, different technologies are used, with different levels of assurance, and different sets of attributes are released. As a service provider, reconciling this is not an easy task. To support multiple technologies, EUDAT decided to choose a credential conversion approach, in which communities can keep their existing AAI and service providers only have to support a few technologies. To this goal, EUDAT is working on Shibbolising services and is evaluating credential conversion and 'Security Token Service' technologies.

Figure 3.5 depicts the AAI architecture proposed by EUDAT. In order to deal with attributes coming from different sources consistency of semantics is important; names, affiliations, contact details, and so on, need to be interpreted in the same way by every attribute provider, and be published in the same schema. Roles need to be named and interpreted the same way across communities, and/or will have to be named uniquely so as to not clash with the same role in a different community. Some projects have chosen to have a central mapping database, where roles published by one attribute provider are mapped to those published by another service provider although this is clearly not a scalable, long-term solution.

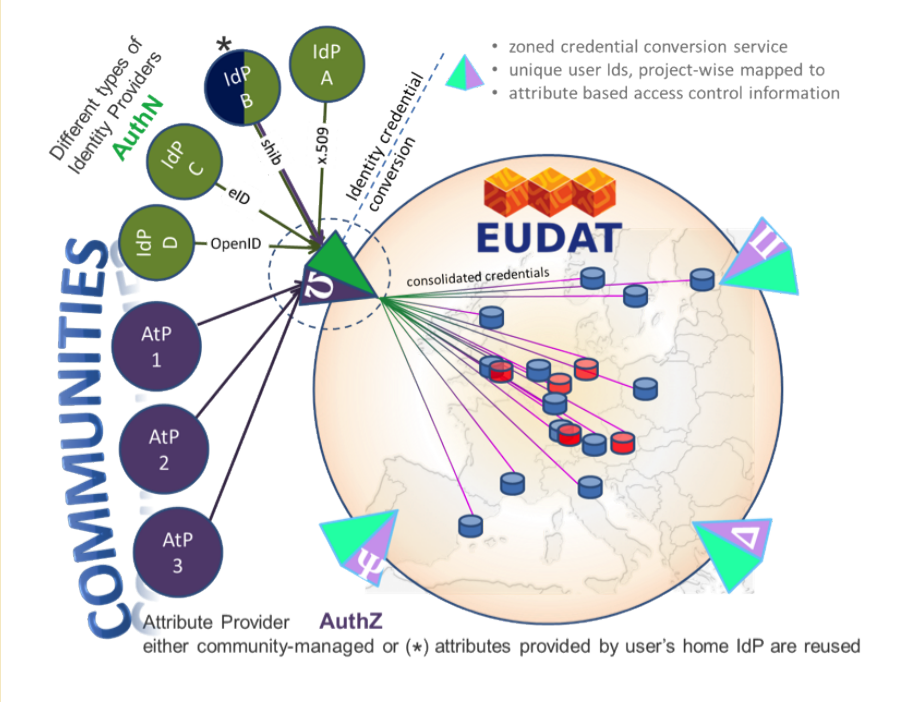


Figure 3.5: Overview of the EUDAT AAI

The work of eduGAIN should evolve in a way that will make it easier to 'con-federate' services, and to access them from multiple federations. Moreover, eduGAIN should make it easier to negotiate the required attributes from identity and attribute providers, possibly even automatically, instead of relying on the provider's lowest common baseline. EUDAT would then benefit enormously from eduGAIN.

Finally, data management policies – not just access rights but also rules for storage, replication, and processing – have to be enforced by the service providers, but based on policies defined by **data owner**s or communities and will have to comply with legal constraints in different countries. In this context, the AAI plays an important role in ensuring that only properly identified and authorised entities are able to use the resources available.

EUDAT AAI Strengths	EUDAT AAI Challenges
Supporting users that currently do not have access to AAIs.	Reconciling multiple authentication methods in to a single approach.
Increasing potential for attribute release through delegation of management.	Ensuring consistency of semantics for attributes.
Flexible environment to allow for the needs of large and small communities.	Need for development of data management policies at a scalable level.

Table 3.11: EUDAT overview

3.7 Conclusions

This chapter has provided an overview and an analysis of the existing infrastructures and of those under development that are considered relevant for this study. All the infrastructures (with the exception of STORK) are used by the research and education community and provide a solution for the specific requirements of the community. Because of the diversity of the requirements coming from the various communities and because of some limitations in the current technologies, it is impossible to have a one-size-fits-all infrastructure. However, some trends can be observed:

- All infrastructures evaluated provide Single Sign-On for the users, although the technology used varies: SAML for Identity Federations, 802.1X + RADIUS for eduroam, X.509 certificates for EGI, PRACE and most of the e-Science infrastructures.

- No single AA(A) technology can be adopted universally, but there should be mechanisms in place to allow for the integration of different technologies. The current trend in the research network environment is to converge Grid identity infrastructures (based on X.509 certificates) and NREN-operated identity infrastructures based on RADIUS and SAML and explore the application of technologies based on secure tokens, such as OAuth and Security Token Services (STSs). There have been already successful initiatives to leverage user credentials issued on campus, for example, via Shibboleth, to issue Grid certificates, such as TCS, and via Moonshot, to implement SSO for different types of applications.

- Authorisation requires mechanisms for the aggregation of identity data (attributes) for authorisation decisions; this is particularly challenging in a heterogeneous environment and it is often implemented via complex systems. There is consensus that current AAIs should evolve to support external attribute authorities that will be in charge of providing the information about the users, rather than relying solely on Identity Providers.

- Accounting remains one the weakest aspects of most of the considered AAIs (with the probable exception of some Grid AAIs). There seems to be consensus that service-oriented architectures can hide complexity while offering rich mechanisms, including better support for accounting.

- Enhancements to the current identity federations are needed to support the e-Science requirements, such as stronger authentication vetting.

- Cloud computing may become a scalable solution for the data deluge problem; however, particularly for public clouds, typically operated by commercial companies, there are still security considerations, such as how to maintain ownership of the data and under which legislation the data are stored. In some cases, cost-related issues need to be addressed. At present, privacy can only be guaranteed by private clouds, which are mostly community-specific. In the research and education community, some NRENs are already investing in private clouds.

- The lack of real standardisation in the cloud technologies, or the lack of inter-operability means that each solution is vendor-specific; this makes the migration to a different provider rather complex.

- Consistent implementation and interpretation of the legal requirements of the Data Protection Directive (this will remain the law in force until the Data Regulation is approved) is essential when building an international infrastructure.

4 Recommendations

4.1 Introduction

The purpose of this study has been to envisage what work is required to create a future-proof AAI for the SDI. In the previous chapters, the study identified different user community requirements and the gaps in the surveyed infrastructures that support different authentication and authorisation services.

Based on the analysis of the requirements and on the AAIs assessed, the study has reached the following conclusions:

1. Existing infrastructures support the user-community requirements for which they have been initially designed. These AAIs are very heterogeneous in terms of service offered, resources required for their operations and their capabilities.

2. Many AAIs have a significant user base and deployment that could be leveraged; hence inter-operability of existing infrastructures should be pursued as a means to deliver the SDI.

3. Currently, there is no existing AAI that can support all requirements. By enabling inter-operability among them it could be possible to offer an integrated and scalable infrastructure where different players control different elements.

3. Continuous enhancement of existing infrastructures in line with emerging requirements should be pursued. The study has identified some gaps in the existing AAIs that, if not addressed, will limit their deployment.

5. There is consensus among different communities on the necessity of adopting federated identity management technologies. For this to happen, some technical aspects need to be enhanced. Identity federations can only be effective if they are as inclusive and as simple to use as possible.

6. The establishment of trust and the fostering of a culture of collaboration are key to the success of an AAI and vice versa.

This chapter provides a set of recommendations to deliver the SDI; the recommendations are addressed to different stakeholders.

4.2 The Vision

In ten years' time, most research data will be readily discoverable and the vast majority of data will be electronically and openly accessible. Data will be used ethically and according to the norms of the research community, including fair attribution.

Mechanisms are needed to support data creation, curation, preservation, as well as the data re-usage process (see Figure 2.2 - Scientific Data Life Cycle Management). The underlying AAI for the SDI should be able to preserve data ownership, to guarantee data authenticity, and to preserve users' privacy whilst enabling them to access a variety of applications and to use the data while protecting any associated IPR.

" In ten years' time, most research data will be readily discoverable and the vast majority of data will be electronically and openly accessible. Data will be used ethically and according to the norms of the research community, including fair attribution."

The following recommendations focus on AAI aspects of the SDI.

The technology alone, however, is not sufficient to ensure the deployment and the sustainability of the SDI. Funding bodies and existing organisations (e.g., NRENs, libraries, and data/IT/research centres) will have to evolve accordingly.

Historically, research data have sometimes been regarded as the researcher's private property, and sometimes as a national asset that should be protected or exploited as a commodity. There is now growing consensus within the research community, and particularly in funding bodies, on the value of sharing data, and there is a general movement towards more openness and less restriction in terms of data usage and availability.

To move towards a vision of wider data sharing and cooperation, trustworthy infrastructures will be an important condition. A successful future infrastructure will need to be able to manage a variety of access policies where there are legitimate restrictions on data access to protect human privacy, and cultural and natural heritage.

However, the trend for more open access to data must take into consideration the high levels of commitment and manpower dedicated to research, and needs to recognise the fundamental drivers of competition and IPR in the research community. The terms of grants given to researchers often dictate and drive the approach taken to these issues. The future SDI should therefore:

- Be trusted by both data and information providers who are concerned about access control to data;
- Be trusted by researchers who are concerned about safety, integrity and confidentiality of their research results.

The recommendations below dictate a pragmatic approach to achieving a flexible AAI for the SDI, building on and improving existing architectures, recognising the diversity of requirements in different research communities and addressing the need to support an open, yet protected, trustworthy infrastructure.

4.3 Technical Recommendations

Technologies to enable an AAI for SDI already exist and are maturing well, while some new technologies are appearing. The core focus for technical development should be on enhancing existing infrastructures, supporting standardisation, interoperability and embedding technologies consistently across EU countries. The resulting infrastructure should be secure but easy to use.

1. **Promote widespread use of existing standards in federated access control technologies for network, service and application**, particularly in geographical areas that have seen slow adoption. Specific support should be given to inter-federation to meet cross-disciplinary and cross-boundary requirements and to create a common, but distributed AAI for SDI.

2. **Enhance existing AA infrastructures to address research communities demands for accessing different type of services in manageable and secure way; the following areas should be addressed**:
   a. Improve AAA support for mobile access and mobile devices;
   b. Improve AA to support non-Web browser access;
   c. Develop security token translation services to enable inter-operability of different AAIs;
   d. Provide guest IdPs for users that cannot rely on an institutional IdP;
   e. Allow for effective resource usage accounting for highly distributed, heterogeneous infrastructures envisaged for global research data;

f. Provide technologies and security mechanisms to enable the use of persistent identifiers for researchers and scientific resources within AAIs;
g. Support social network identities in combination with institutional identities to address specific use-cases for the SDI.

3. **Enhance authorisation in inter-federation scenarios by providing support for distributed attribute management.** A better separation between the authentication process and the retrieval of attributes needed for authorisation should be implemented particularly for Identity Federations. Currently IdPs that belong to an Identity Federation authenticate the users and provide the necessary attributes. This model does not scale for large research projects. As mentioned in the previous chapters research collaborations are better positioned to maintain the additional users attributes, the related policies, processes and tools for releasing them.

4. **Phase-out IP-based Authentication** still in use in the library communities. National federations should support this community to migrate to federated identity technologies. In line with this, libraries and national library consortia should work together with other organisations dealing with licensing at European level to negotiate agreements with publishers regarding the adoption of the AAI for authentication of their resources and also moving towards granting access to these resources to every researcher in Europe. Alternative consortial Open Access solutions, such as SCOAP3 could also be investigated.

5. Harmonisation of attributes has been a long time priority. Reality has shown that a global scale attributes harmonisation is not feasible. **However, well-defined semantic attributes within a community and mapping mechanisms are a possibility.** Different communities are therefore encouraged to work towards this goal.

6. **Leverage cloud virtualisation and on-demand capabilities** to build distributed dynamic and secure environments controlled by institutions but not necessarily hosted within the institutions' boundaries.

4.4 Policy and Practice Recommendations

There is growing realisation that existing policies and practices do not meet the needs of ever evolving technologies and, more critically, are being interpreted and implemented differently across the EU member states.

1. **Facilitate the development of a common policy and trust framework for Identity Management** that involves Identity Federations, e-Research communities, libraries and datacentres. REFEDS (Research and Education FEDerations), the international body led by TERENA, could help coordinate Identity Federation processes, practices and policies and to discuss ways to facilitate inter-federation work. REFEDS should evolve to play a similar role played by the IGTF, in the e-Science community. Communication amongst different groups (eIRG, REFEDS, IGTF, ESFRI, LIBER and policy makers) should be improved.

2. **National federation operators should act to expand the coverage of their identity federations** to be more inclusive. This can be achieved by **allocating national funding to support and train communities to join a national identity federation**. This should also include support for small data/IT centres and any other institutions with few human resources and little expertise to connect to new AAIs.

3. **Implement scalable policy negotiation mechanisms**. Negotiating agreement between IdPs/Federations and SPs does not scale on per IdP-basis. REFEDS should work with national federations to offer a solution to this problem.

4. **Identity Federations should work to harmonise their policies**. EduGAIN experience has demonstrated that lack of harmonisation results in different levels of opt-in, which hinder the large scale inter-operability of different federations. REFEDS could act to coordinate this at European level.

5. **The EC should invest in coordinating actions to develop and implement a pan-European licensing agreement and infrastructure to facilitate the sharing of information resources**. Libraries, publishers and infrastructure providers together should work to support this goal.

6. **Lower the adoption entry level of existing infrastructures for new users and providers and support communities to benefit from existing AAIs**. In the case of Identity Federations, for instance, participation in federations requires a certain degree of technical and policy knowledge, which is not always available in the non-technical sectors. Offer well-documented and standardised practises and if possible ready to use solutions, particularly for services and/or institutions with few human resources and little technical expertise.

7. **Closer collaboration between national federations, national e-Science centres, community datacentres and libraries is needed** to ensure that existing services are offered in a consistent way to users. One way to improve this collaboration may be to organisationally link NRENs and other types of infrastructures (i.e., Grids, clouds), hence offering one interface to both network, grid and cloud services per country. This is the model followed in the Netherlands, in Finland and in other countries.

8. **The EC should coordinate the dialogue between Research and Education pan-European initiatives** (i.e., GÉANT/edUGAIN, EGI, EUDAT, OpenAire) and governmental initiatives (STORK) to ensure future inter-operability between R&E and the public sector in relation to AAI deployment. Facilitating the exchange of scientific data across disciplines and national boundaries is a global challenge and the stakeholder organisations that have been identified in this report (e.g., TERENA, LIBER, EGI, COAR, EUDAT, and EDUGAIN) should support and engage in global forums in order to contribute to the international coordination of access standards.

4.5 Legal Recommendations

Developments in this area should focus on achieving clarity, consistency and user-friendly tools for implementation. The main law to be considered in this area is the European Data Protection Directive (95/46/EC) and its revision, the draft Data Protection Regulation (published in January 2012).

1. **Extend the Legitimate Interests justification** (Article 7f of the Data Protection Law) **to cover international transfers**, as proposed by the draft Data Protection Regulation to permit the use of a common legal framework for all e-Research involving European researchers or services. It has also been suggested that the Regulation might be accompanied by a review of the current arrangements for export of personal data: including provisions suitable for use by overseas universities and public research organisations would further assist collaboration between European and overseas researchers.

2. **The EC should provide clarity about Consent and Legitimate Interest**. The existing relationships (employment, site licences, etc.) between individuals,

identity providers and services in education and research cast doubt on whether Consent (Article 7(a) of the Data Protection Directive) is the appropriate justification for processing in federated access management (the draft Regulation would make the use of Consent within an employment relationship even more questionable). Instead both identity providers and service providers appear to have a legitimate interest in providing access to the services their members seek to use, which justifies them exchanging information necessary to do so. Consent can then be reserved for information that is not necessary to provide the service, but where the user wishes to enhance the service by providing that information.

3. **The EC should fund a study to investigate how adequate protection of personal data can be achieved** by incorporating lightweight agreements into existing relationships between researchers, projects, services and home organisations, whether these are user-mediated, organisation-mediated, (for example, in VO-based authorisation), or both.

4. **The EC, together with the Article 29 Working Party[10], should create and support a clear statement on the legal status of processing opaque identifiers**, implemented consistently across Member States, to support the use of privacy-protecting identifiers in federated access management. This statement should offer the possibility for service providers to treat suitably protected opaque identifiers as non-personal data or, at least of representing a very low risk to privacy with correspondingly light regulatory requirements.

5. Fragmentation in the implementation of the current Data Protection Directive has hindered international collaboration, which is the cornerstone of research. **The EC, together with the Article 29 Working Party, should organise training for the Member States representatives** to avoid cultural interpretations of the Directive and to prepare for a smooth transition from the Directive to the Regulation.

6. **Member State Data Protection Laws should be aligned with EC Data Protection Directives/Laws.** This lack of clarity and the resulting variation in Member State interpretations on what is subject to personal data regulation make it difficult to exchange attributes between countries. If the same attribute is considered personal data in one country but not in another it is unclear whether the attribute can lawfully be transferred between them.

4.6 Recommendations for Funding Agencies, EC and Member States

1. **Funding should be allocated to support interoperability of e-Infrastructures and to enhance the underlying AAIs.** In particular, EC funding should be directed, where possible, to consolidate and harmonise established systems rather than creating new ones. A proliferation of localised solutions unable to leverage existing infrastructures would result in more fragmentation and higher management costs. The deployment of an SDI can only be successful if Member States embrace it and provide the necessary funding to ensure that universities, libraries and research centres can connect to it.

2. **Plans should be made by national funding bodies to ensure that structural funding is available for AAIs.** Whilst this is already the case for the network backbone in many countries and for some grid facilities, mechanisms are not yet in place to ensure long-term support for other types of infrastructures (i.e., Identity Federations and Data Infrastructures).

[10] The Article 29 Data Protection Working Party was set up under the Directive 95/46/EC to protection individuals regarding the processing of personal data.

3. **The EC should bootstrap plans to implement the Digital Agenda**. Strong measures should be taken to highlight the benefits of an integrated SDI and to engage with different stakeholders. The fragmentation of the current landscape and the different needs/interests of the stakeholders involved in different initiatives may lead to a situation that can hinder the actual deployment of an integrated AAI for SDI. This situation can be avoided by ensuring participation of different players (at technical, policy, national and international levels) at the very beginning.

4. **The European Commission should invest in a 'training the trainers' initiative** to ensure that data professionals (i.e., datacentres and libraries), as intermediaries, have the correct knowledge to educate and provide guidance to researchers about issues, such as data privacy and intellectual property as well as to address cultural barriers to collaboration and data sharing via an AAI.

5. **Periodic studies should be funded to assess the emergence of new technologies and the penetration of existing AAIs at national level**. The eIRG, REFEDS and the IGTF are well positioned to be involved in these studies. The results of these studies can be used to inform recommendations, workshops or specific actions.

4.7 Conclusions

Table 4.1 summarises the recommendations that have a higher priority and should therefore be addressed in the short term.

Recommendation	Action Required	Main Stakeholder(s)	Area
Rely on standards for federated technologies for network, service and application access across Europe	Specific support should be given to inter-federation to meet cross-disciplinary and cross-boundary requirements and to create a common access infrastructure.	Developers, eduGAIN, EUDAT, EGI, REFEDS	Technical
Enhance existing AA infrastructures to address the demands of research communities for accessing different types of services in a manageable and secure way.	AAA support for mobile access; Support for non-Web browser applications; Develop security token translation services to enable inter-operability of different AAIs; Provide guest IdPs for users that cannot rely on an institutional IdP; Allow for effective resource usage accounting for distributed and heterogeneous environments; Enable the uptake and use of persistent identifiers within AAIs; Support social network identities in combination with institutional identities to address specific use-cases for the SDI.	National Identity Federations, eduGAIN, Research collaborations (i.e., big scientific projects)	Technical
Enhance authorisation in inter-federations scenarios by providing support for distributed attribute management.	Provide security mechanisms and tools to enable Identity Federations to consume attributes managed by collaboration projects.	National Identity Federations, collaboration projects (eResearch)	Technical
Phase-out IP-based Authentication	Provide support for those institutions relying on IP-based Authentication to migrate to federated access technologies	National Identity Federations, Libraries, National funding bodies	Technical / Funding
Facilitate the development of a common policy and trust framework for Identity Management that involves Identity Federations, eResearchers communities, libraries and datacentres.	a) REFEDS should coordinate Identity Federation processes, practices and policies on behalf of R&E Identity Federations b) EC to facilitate communication among different groups	eIRG, REFEDS, IGTF, ESFRI, EUDAT, LIBER and policy makers)	Policy
The operators of national federations to expand the coverage of their identity federations.	Allocate national funding to support and train communities to join national identity federations	National funding bodies, EC	Policy
Implement scalable policy negotiation mechanisms.	Define ways to simplify the negotiation of service agreement (services should not negotiate with each IdP, but via the federation).	REFEDS, eduGAIN, National Identity Federations	Policy

Table 4.1: High priority recommendations and relevant stakeholders

Recommendation	Action Required	Main Stakeholder(s)	Area
Identity Federation to harmonise their policies.	Define guidelines for R&E Identity Federations policies	REFEDS, eduGAIN	Policy
Lower the adoption entry level of existing infrastructures for new users and providers and support communities to benefit from existing AAIs.	Consider ways to offer ready-to-use solutions that hide technical complexity from the users.	eduGAIN, EGI, EUDAT, National Identity Federations	Policy
Provide clarity about Consent and Legitimate Interest	Provide clear and simple documentations and raise awareness on when consent can or cannot be used.	EC, Member states, eduGAIN/National Identity Federations	Legal
Organise training for the Member States representatives to avoid cultural interpretations of the Directive and to prepare for a smooth transition from the Directive to the Regulation.	Raising awareness.	EC, Member states	Legal
Secure funding to work towards inter-operability of e-Infrastructures and to enhance the corresponding AAIs	EC funding should be directed, where possible, to enhance, consolidate and harmonise established systems rather than creating new ones.	EC, Member States	Funding
Secure sustainable structural funding to support various e-Infrastructures	Provide mechanisms to ensure long-term sustainability for different infrastructures (i.e., Identity Federations, Data Infrastructures and Grid).	EC, National Member States	Funding
Invest in 'train-the-trainers' initiatives	Provide training for data professionals to provide guidance to researchers on issues, such as data privacy and intellectual property as well as to address cultural barriers to collaboration and data sharing.	EC, National Member States	Funding

Table 4.1: High Priority recommendations and relevant stakeholders - continued

References

Abfab	http://tools.ietf.org/wg/abfab/
ADFS	http://en.wikipedia.org/wiki/Active_Directory_Federation_Services
APARSEN	http://www.alliancepermanentaccess.org/index.php/aparsen/
APGridPMA	http://www.apgridpma.org/
ARROW	http://www.arrow-net.eu/
CERN	http://www.cern.ch/
Code of Conduct	https://refeds.terena.org/index.php/Data_protection_coc
Data-protection	https://confluence.terena.org/display/aaastudy/AAA+Study+Home+Page
Data Protection Directive	http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:NOT
Data Protection Regulation	http://ec.europa.eu/justice/newsroom/data-protection/news/120125_en.htm
DARIAH	www.dariah.eu/
DARIAH Manifesto	http://www.dariah.eu/index.php?option=com_content&view=article&id=195:icri-manifesto&catid=3:dariah&Itemid=197
DASISH-manifesto	http://dasish.eu/manifesto/Manifesto2012-03-14.pdf
EAP	http://en.wikipedia.org/wiki/Extensible_Authentication_Protocol
eduroam	http://www.eduroam.org
eduGAIN	http://www.edugain.org
EEF	http://www.einfrastructure-forum.eu/
EGI	http://www.egi.eu/
EGI-AAA-Study	https://documents.egi.eu/public/ShowDocument?docid=1178
e-IRG	http://www.e-irg.eu/publications/white-papers.html
EMBL	http://www.embl.de/
EPIC	http://www.pidconsortium.eu
ESA	http://www.esa.int/
ESEPRT	http://dare.uva.nl/document/150752
ESFRI	http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri
EUDAT	http://www.eudat.eu/
EUGridPMA	http://www.eugridpma.org/
EURO-VO	http://www.euro-vo.org/
ENVRI	http://envri.eu/
Europe2020	http://ec.europa.eu/information_society/digital-agenda/index_en.htm
EZProxy	http://www.oclc.org/ezproxy/
FIM	https://cdsweb.cern.ch/record/1442597/files/CERN-OPEN-2012-006.pdf
FIM Workshops	https://cdsweb.cern.ch/record/1442597
Google-generation	http://www.emeraldinsight.com/journals.htm?articleid=1733495&show=abstract
Google Scholar	http://scholar.google.com/intl/en/scholar/about.html
GÉANT	http://www.geant.net/
Helix Nebula	http://press.web.cern.ch/press/PressReleases/Releases2012/PR03.12E.html
HLEG on scientific data (Riding the Wave Report)	http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf
Janet	https://www.ja.net/
JISCLegalToolKit	http://www.jisclegal.ac.uk/Portals/12/Documents/PDFs/Report%20on%20Cloud%20Computing%20and%20Law%20UKFEHE%20-%202011.pdf

Kalmar2	http://www.kalmar2.org
LifeWatch	http://www.lifewatch.eu/
IEEE 802.1X	http://en.wikipedia.org/wiki/IEEE_802.1X
IGTF	http://www.igtf.net/
ISBN-978-1-4577-0458-1	http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6060006
Manifesto	http://dasish.eu/manifesto/Manifesto2012-03-14.pdf/
NAREGI	http://www.naregi.org/
OAuth	http://oauth.net/
ODE	http://www.alliancepermanentaccess.org/index.php/current-projects/ode
OpenAIRE	http://www.openaire.eu
Open Access (OA)	http://en.wikipedia.org/wiki/Open_access
OpenID	http://www. openid.net/
OpenID Connect	http://www.openid.net/
Open Science Grid	http://www.opensciencegrid.org/
OpenStack	http://www.openstack.org/
ORCID	http://about.orcid.org/
PaNdata	http://pan-data.eu/node/29
PARADE	http://www.cros-portal.eu/page/36-strategy-european-data-infrastructure
PKI	http://en.wikipedia.org/wiki/Public-key_infrastructure
PersID	http://www.persid.org/index.html
PRACE	http://www.prace-project.eu/
PRAGMA	http://www.pragmaworld.net/
Project Moonshoot	http://www.project-moonshot.org/
RADIUS	http://en.wikipedia.org/wiki/RADIUS
REFEDS	https://refeds.org/
RFC3281	http://www.ietf.org/rfc/rfc3281.txt
Safe Harbour Agreement	http://www.aicpa.org/INTERESTAREAS/INFORMATIONTECHNOLOGY/RESOURCES/PRIVACY/FEDERALSTATEANDOTHERPROFESSIONALREGULATIONS/SAFEHARBORAGREEMENT/Pages/default.aspx
SAML	https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=security
SCI	http://www.eugridpma.org/sci
SCOAP3 Project	http://scoap3.org/
SLCS	http://www.switch.ch/grid/slcs/index.html
SimpleSAMLphp	http://simplesamlphp.org/
Shibboleth	http://shibboleth.net/
SWITCH	http://www.switch.ch
TAGPMA	http://www.tagpma.org/
Ten Tales of Drivers & Barriers in Data Sharing	http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2011/10/7836_ODE_brochure_final.pdf
TERENA compendium	http://www.terena.org/activities/compendium/
Textgrid	http://www.textgrid.de/
TCS	http://www.terena.org/tcs/
Umbrella Project	https://umbrella.psi.ch/euu/
UNINETT	http://www.uninett.no
VOMS	http://www.sciencedirect.com/science/article/pii/S0167739X04001682
WP29 (Article 29 Working Party)	http://ec.europa.eu/justice/data-protection/article-29/index_en.htm

WP29-STORK	http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/others/2011_04_15_letter_artwp_atos_origin_annex_en.pdf
WLCG	http://wlcg.web.cern.ch/
Workshop	https://confluence.terena.org/display/aaastudy/AAA+Study+Workshop
XSEDE	https://www.xsede.org/