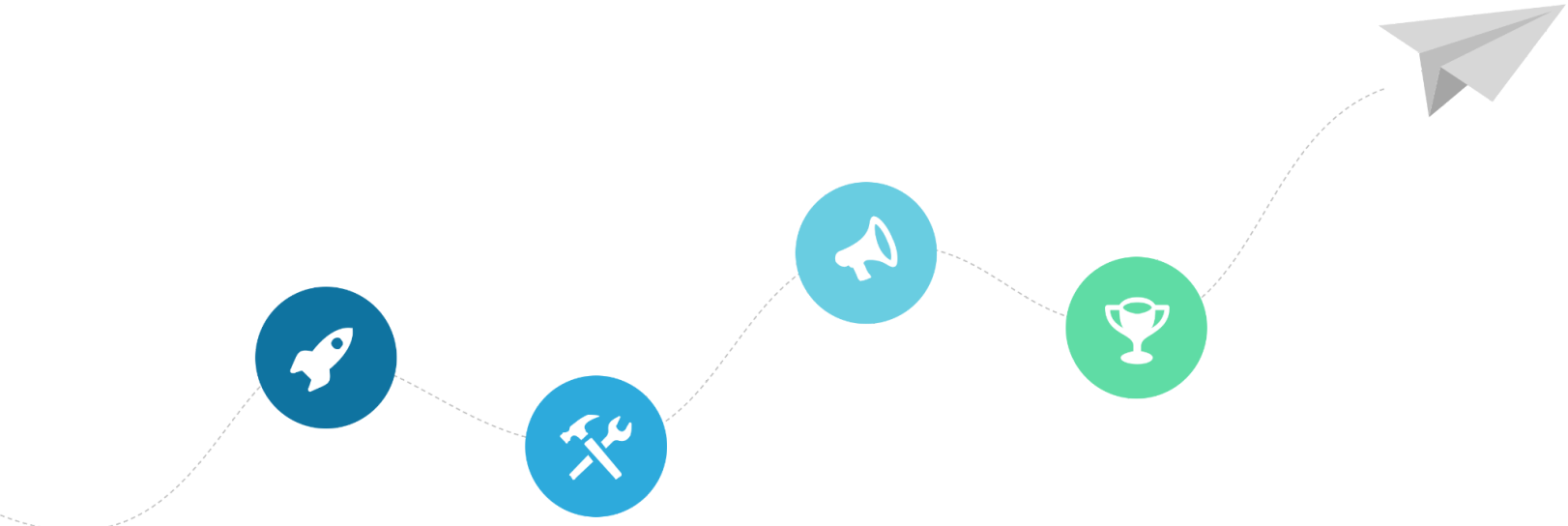
HUMAN GENOMES PLATFORM PROJECT

Federated Identity and Access Management

DISCOVERY PHASE REPORT

version 3.0

National Community Needs



DOI: 10.5281/zenodo.10725133

Table of Contents

Glossary

AAF	Australian Access Federation: national identity federation that simplifies inter-organisational authentication and authorisation for research and education organisations.
Attributes	Identity attributes are the set data about a user. Attributes may include the type of user (staff, student) or personal metadata such as name, email address, phone etc. Can be known as “claims” in international standards documentation such as GA4GH.
Credentials	Combination of inputs that a user enters into an interface to be authorised into a system or service. For example a username and password together are credentials, but they may include biometric data like a fingerprint, a timed code, email address etc.
HGPP	Human Genomes Platform Project
IAM	Identity and Access Management
Open Authorisation (OAuth2)	An open standard for allowing users to login to a service with credentials from an alternate resource
OpenID Connect (OIDC)	Third generation standard of the OpenID Authentication Protocol (adheres to OAuth2
openLDAP	Open-source implementation of Lightweight Directory Access Protocol
Security Assertion Markup Language (SAML)	Open-standard for exchanging authentication and authorization data between an identity provider and a service provider.

Acknowledgements

The HGPP formed part of Australian BioCommons’ Human Genome Informatics initiative and was funded by NCRIS via the Australian Research Data Commons (https://doi.org/10.47486/PL032) and Bioplatforms Australia. Contributions were also made by partner organisations: Australian Access Federation, Garvan Institute for Medical Research, National Computational Infrastructure, QIMR Berghofer Medical Research Institute, The University of Melbourne Centre for Cancer Research, the ZERO Childhood Cancer Program and Children’s Cancer Institute.

Version Control

Date	Version Number	Description of Changes
19 April 2022	1.0	First version after subproject and project discovery work
20 May 2022	2.0	After project reference group secondary review, minor wording and formatting changes (without change of meaning), addition of authors’ list.

Authors

in alphabetical order by surname

Carnuccio, Patrick - AAF

Cowley, Mark - ZERO

[1]Davies, Kylie - AAF

Downton, Matthew - NCI

Dumevska, Biljana - ZERO

Holliday, Jessica - BioCommons

Kummerfeld, Sarah - Garvan

Lin, Angela - ZERO

Monro, David - NCI

Patterson, Andrew - UMCCR

Pope, Bernie - BioCommons

Ravishankar, Shyamsunder - Garvan

Robinson, Andrew - NCI

Scullen, John - AAF

Shadbolt, Marion - BioCommons

Syed, Mustafa - ZERO

Wood, Scott - QIMRB

Wong-Erasmus, Marie - ZERO

1. Introduction

The Human Genomes Platform Project (HGPP) is a nationally-funded collaborative research project aiming to enhance capability for securely and responsibly sharing human genomics research data. National and international connectivity will maximise the utility of these sensitive and valuable assets. The partners on the project represent many of the largest human genome sequencing and analysis efforts in Australia.

At the heart of any technology platform is identity and access management (IAM): a collection of standards, policies and technologies that enable a platform to determine whether to permit access to a user. In a federated environment such as the Australian/global genomics community, IAM is the glue that enables loosely coupled systems to establish strong trust relationships for the purposes of data sharing (Figure 1). Trust relies on technologies such as cryptography but also on coordinated policies outlining shared expectations between federation participants.

The aims of the Federated IAM sub-project within the HGPP are to explore and implement systems that can be used across organisations to confidently ascertain that someone being granted access is, in fact, who they say they are, and that their professional identity and role is considered. A new federated identity and access system will be piloted across participating repositories and in partnership with other sub-projects of the HGPP, to establish the case for national human genome community adoption. Figure 1 illustrates the interrelationships between the subcomponents of the project.

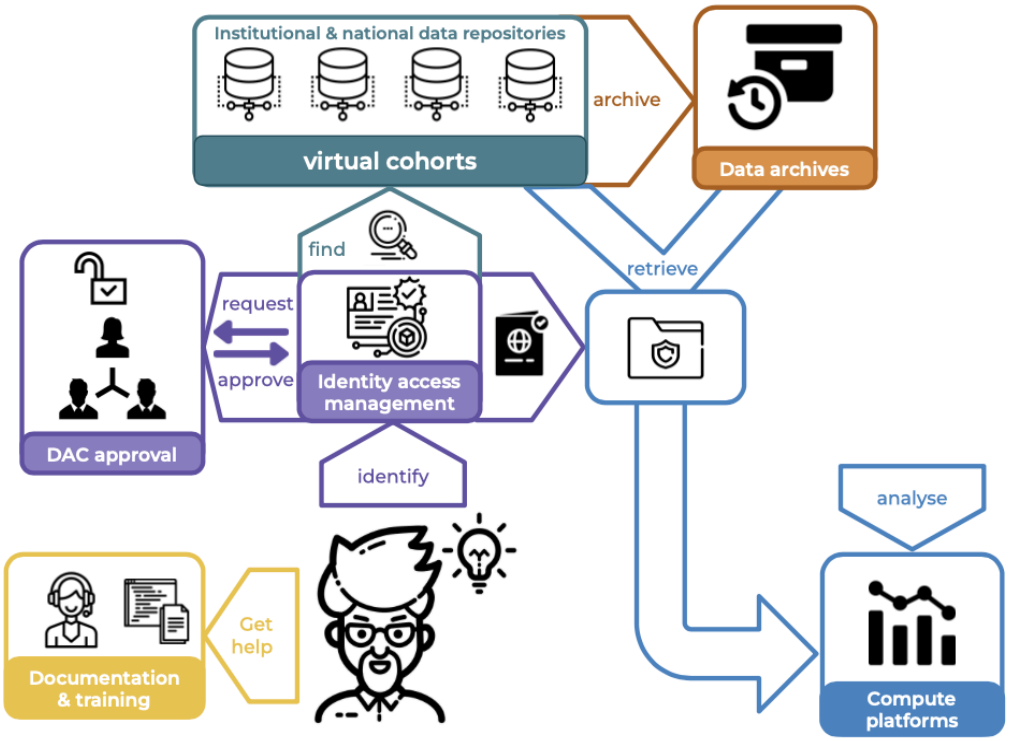


Figure 1 - The HGPP infrastructure ecosystem from the perspective of the research end user showing the key elements of the human genomics data sharing toolbox and data and information flows. (Drafted by Marion Shadbolt, Human genome data specialist, HGPP.[1])

The initial focus of the Federated IAM sub-project team (from here on known as “we”) was a discovery and recording phase to define:

- the current state of identity and access management in the community
- the set of problems that need to be addressed
- key stakeholders and their (likely) requirements (Figure 1).

For an Australian genomics federation to be successful, widespread adoption of the new processes and systems will be needed. To foster widespread adoption, we used a range of techniques to understand the current state and the future needs of the national human genome research community including:

- project partner interviews
- synchronous (workshops, meetings) and asynchronous (communication tools, kanban boards and shared repositories) discussion and review
- consultation with influential stakeholders not involved in the project (MCRI/VCGS and CSIRO)
- a survey of human genome researchers to validate user stories developed by the project team.

Team members are our subject matter experts and they work at the following organisations:

- Australian BioCommons (BioCommons)
- ZERO Childhood Cancer Program of the Children’s Cancer Institute (ZERO)
- University of Melbourne Centre for Cancer Research (UMCCR)
- Garvan Institute of Medical Research (Garvan)
- Queensland Institute of Medical Research Berghofer Medical Research Institute (QIMRB)
- National Computational Infrastructure Australia (NCI).

The initial focus of interviews and sub-project meetings was to ascertain the current state. The Federated IAM Discovery Phase Report (this document) records:

- the current state of processes and tools for identity and access management across the community
- national community needs
- gap analysis
- identification of international projects with components suitable to canvas and potentially pilot.

This document will be used as a reference to plan the pilot for a system that addresses prioritised requirements to create a Minimum Viable Product (MVP). The audience for this document includes the sub-project team, other HGPP stakeholders and the project reference group.

2. Current State Findings

Project research and interviews confirmed that human genome research organisations in Australia use a variety of identity management solutions appropriate to their context. There are differences in compatibility and functionality between them. The project intends to leverage standards to integrate existing technologies within the Australian genomics community to enable secure interoperability while leveraging existing technology investments. This section details current identity and access management processes and technologies and their limitations.

2.1 Authentication Technology Landscape

- Authentication protocols used by the community include SAML 2.0, OIDC and other OAuth2 protocol implementations within systems such as Auth0, Okta and LDAP.
- Sources of identity vary:
    - Windows Active Directory (AD); Azure Active Directory; AAF's hosted Shibboleth endpoint (RapidIdP).
    - two of the project partner organisations have also set up Google Workspace accounts for their home domain and use Google Workspace identity management to support access to Amazon Web Server and other tools.
- The majority of researchers who access data housed at project partner organisations are issued credentials in the domain for the data holding organisation. Exceptions to this are where Garvan issues access to users with external email addresses, who create a new password for access.
- Limited data sharing between ZERO CCI and UMCCR was established in 2021 using federated identity access (via AAF and CILogon).

<<Table 1 summarises aspects of the authentication technology landscape for key organisations within the national human genome research community. This has been redacted for internet publishing to protect the security of the partner organisations..>>

Table 1 - *Authentication Technology Landscape*

Technology Element	Garvan	UMCCR Lab Internal	UMCCR via UniMelb	ZERO CCI	QIMR Berghofer	NCI	MCRI & VCGS
Windows Server/Local Active Directory							
Azure/Cloud based Active Directory							
Okta							
CILogon (BioCommons instance)							
AAF Subscriber							
AAF RapidIdP hosted Shibboleth endpoint							
Google Workspace (G Suite) ID for certain apps							
Authorised external users use their home organisation credentials to access this organisation's resources							
Multi Factor Authentication (MFA)							

2.2 Roles and System Components

Partner organisations currently organise onboarding of users, creating identities and managing access control in a variety of ways. For ease of understanding, role/component names used in this document to describe IAM activities are summarised below (Table 2).

Table 2 - *Roles in the Federated Identity and Access Management Landscape*

Role	***At times also known as:***	***Description in the Context of IAM***
Identity Provider (IdP)	IdP, home login screen, virtual home, home organisation	Component provided by a home organisation for a user. Presents a login page for the user to authenticate to. Also shares user attributes with a service provider so that the user can be authorised to access the service provider's service.
Credential Service Provider (CSP)	Human Resources Officer, Recruitment Officer, IT Officer	Person who verifies identity of people when issuing credentials. May check birth certificate, drivers licence, have a zoom interview, check the validity of ID documents or all of these tasks. This function may be outsourced to a third-party.
Service Provider (SP)	Service, Holding Organisation, Application, Tool, System.	Organisation providing a service to the platform, or the service itself. For example Gen3, REMS, DUOS, JupyterHub, Cloudstor, Galaxy, Beacon, AWS S3, EGA Node - are all examples of services provided to researchers.
Authorised Officer	Institutional Signing Officer, CEO, Company Secretary	Person who is authorised to sign and enter into data sharing agreements with another organisation (e.g. board member, executive officer or CEO).
User		User of the platform. May be any of the roles within the HGPP including Researcher, Bioinformatician, Systems Engineer, Geneticist, DAC Committee Member, DAC Coordinator, Data Owner. Authenticates to systems using credentials.

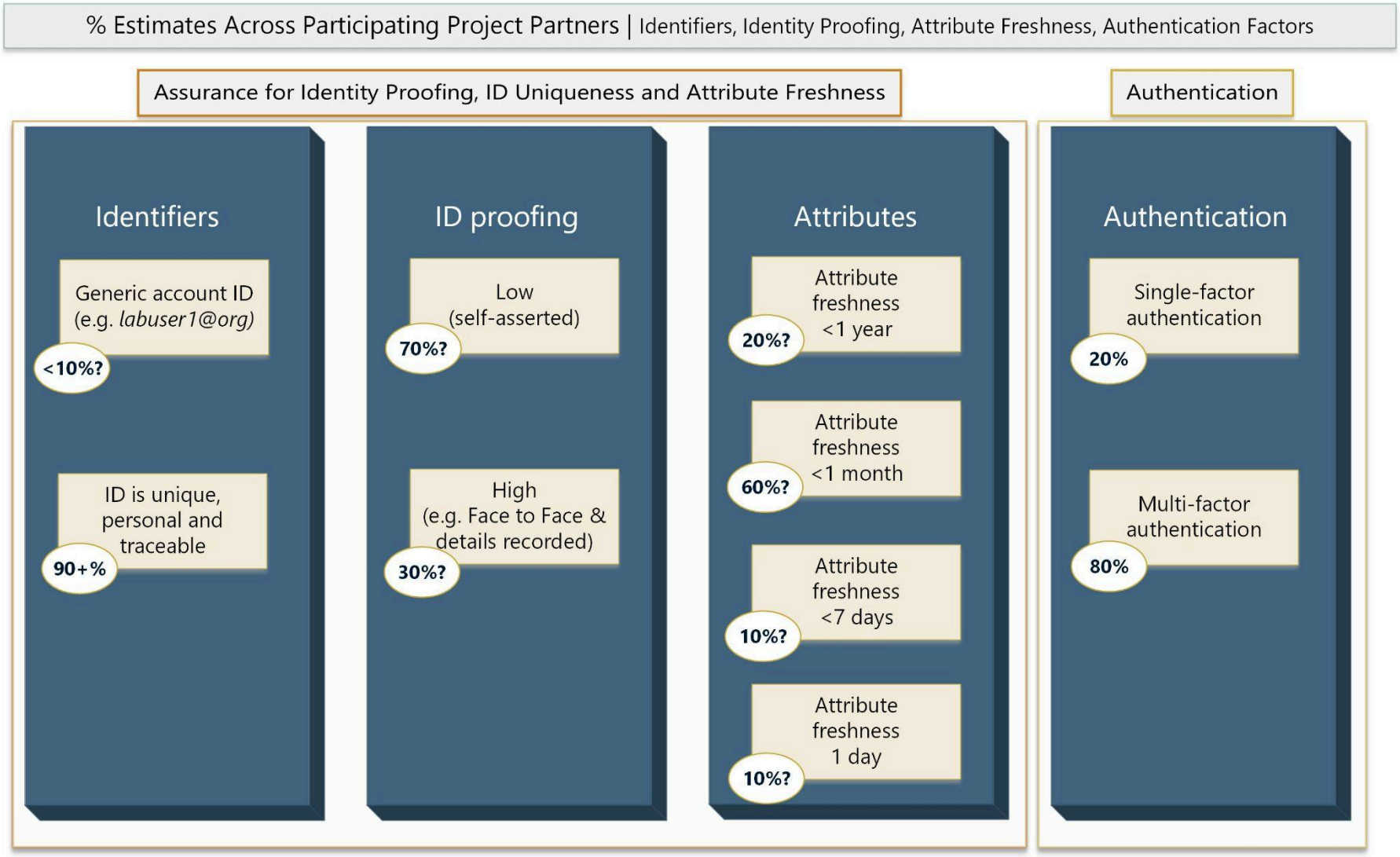


Figure 2 - Identity assurance aspect estimates, across five major project partner organisations interviewed.

2.3 Identity Assurance

The current state of project partners' identity assurance is illustrated in Figure 2, and summarised in the following estimates:

- Between 90-100% of identifiers issued by project partners are unique and belong to a single natural person. Whether users can be contacted by credential providers is unknown. At least one of the project partners confirmed that their organisation identities are never recycled to another person.
- Most of the ID proofing appears to be self asserted, with heavy reliance on trust in a known email domain. For internal users, ID checks may be undertaken by organisation's HR departments however this was not confirmed.
- All partners interviewed undertook steps to confirm attribute freshness. Frequency of attribute freshness checking varied widely, from as little as a day to as long as once annually.
- 80% of project partners interviewed leverage multi-factor authentication. However, this is enforced at varying intervals, as long as 60 days for some organisations. There is also a question as to whether MFA is enforced correctly, where access to one factor does not by itself grant access to others.

Overall, the current state of identity assurance is reasonable in comparison to basic REFEDS standards, apart from identity proofing and some examples of attribute freshness.

Project partners reported a significant manual handling burden when off-boarding users who depart projects or the organisation. Dr Marie Wong-Erasmus, Precision Medicine Informatics Manager at the Childhood Cancer Institute described a key pain point of managing off-boarding:

> *"When someone leaves we try to do it all within 7 days but the problem is a non-central authentication one. Our Information Technology department removes the emails, access to internal storage and then Computational Biology deal with other things like access to apps and Slack etc*
> *… So we have a giant checklist each time which is not ideal."*

2.3 Early Findings - Issues with Current IAM Practices

During current state analysis, several issues became apparent:

Table 3 - ***Early Findings on Issues - Identity and Access Management***

#	*Issues Description*	*Federated IAM Sub-project Mitigation Potential*	*In scope for Federated IAM in HGPP?*
1	Self asserted identity: Data access requests frequently handled via email and reliant on trust in the email domain, introducing risks such as impersonation or person-in-the-middle attacks.	Identity Proofing uplift: registration to platform could require proof of identity to be checked.	Yes
2	Researchers are often low on computational skills required to work with data.	Identity data could contain training details indicating capability to process complex data.	TBD
3	Researchers often do not understand the risk of re-identifiable data and the need for encryption.	Identity data could be dependent on ethics training proof prior to any sensitive data access granted.	Yes
4	Better visibility needed of who has current access to data or who is holding data.	Identity data could be linked to a project which is in turn linked to details of data access granted. A centralised access management system provides visibility and enables auditing.	Yes
5	When researchers leave an organisation there can be a delay before their access is revoked.	Attribute Freshness uplift: candidate solutions will be assessed for their ability to check for attribute freshness and policies tested to mandate user offboarding standards.	Yes
6	Unlinked IDs: REMS - User needs to remember which ID they used to register and its password and they sometimes try to authenticate with a separate ID	Linked IDs via a centralised identity and access management system.	TBD
7	Is the user in possession of their own credentials? Currently unknown	Multi-factor authentication	Yes
8	Manual offboarding is complex, time consuming and difficult with so many different systems for which access must be revoked.	Support multiple pathways to navigate from authentication to access (dependency on partner organisations' own processes being in order.)	Yes

3. User Stories and Survey

Representatives of project partner organisations, with experience working in relevant roles, contributed role-based User Stories to describe their own identity and access management needs.
A survey collected wider feedback on these user stories from the human genome researcher community.

Free text comments from the survey added further information. The following comments were directly relevant to identity and access management:

Principal Investigator User Story Survey Feedback

For "As a Principal investigator, I need to be able to add and remove team members through the life of the project so as to easily establish seamless access for all members of my team" feedback included:
"This is critical and ideally would not require amendments as the team changes, or at most a simple notification."
And from another respondent:
"Ethics application/HREC approvals record who can access the project and be associated with the data involved in a project, not data access applications and their associated committees. Expecting both to have a record doubles up on paperwork. I would remove the whole statement. Unless you propose a user interface for a centralised data repository, not housing a copy on the applicant's organisational infrastructure. The statement should be clearer the use of 'centralised data repository' and 'group permissions'. Remove the term 'data access application."

DAC Committee Member User Story Survey Feedback

For "As a human genomics DAC Committee member, I want to know the verified identity and institution/workgroup details of the applicant, so as to avoid the need for double checking with institutions" feedback included:
"Strongly agree - DAC Committee need to be able to interface with the Data custodians who will (in many cases) have ultimate approval of data access."

4. Requirements Recording and Gap Analysis

Human genome data is inherently sensitive, which elevates the need to confirm the identity of researchers requesting access to data and tools. We recognise that proposed solutions must be able to support more robust identity and authentication assurance where appropriate.

We drafted requirements based on the user stories, the current state, and technical, legislative and operational considerations. The requirements were reviewed and scoped. We grouped related high priority requirements and mapped these to the current state. This work built an outline of the Desired State traced to the Current State and scoped prioritised requirements (shown in Tables 4 to 11).

The mapping exercise also produced a gap analysis, illustrated in Figure 3. Actions to address each gap will form the backlog for the pilot.

we drafted an outline of the Minimum Viable Product for the pilot from these desired state summaries as follows:

Minimum Viable Product Outline

- A dedicated unique immutable user identifier for the human genome research community with suitable attributes for relying services, so that users can authenticate using their home organisation credentials to access data and services hosted at other organisations without needing a new set of credentials *Table 4*;
- Ability to link and unlink external identities to the unique immutable identifier *Table 5*;
- Identity proofing (primary sources) for registration to platform and creation of unique identifier *Table 6*;
- An AAI solution capable of integrating in a technologically diverse landscape *Table 7*;
- Workgroup management capability allowing community managers to arrange their teams, data and tools *Table 8*;
- A more seamless and automated user experience compared to email exchanges, while maintaining rigorous security and access control *Table 9*;
- Infrastructure and community policies are implemented at user enrolment and at authentication points *Table 10*;
- Audit trails for key actions (for DAC as shown in the System Use Diagrams in DAC Automation Discovery report, for other sub-projects this needs defining) linked to verifiable user id, time-stamping and reporting, together with a facility for recording management actions *Table 11*.

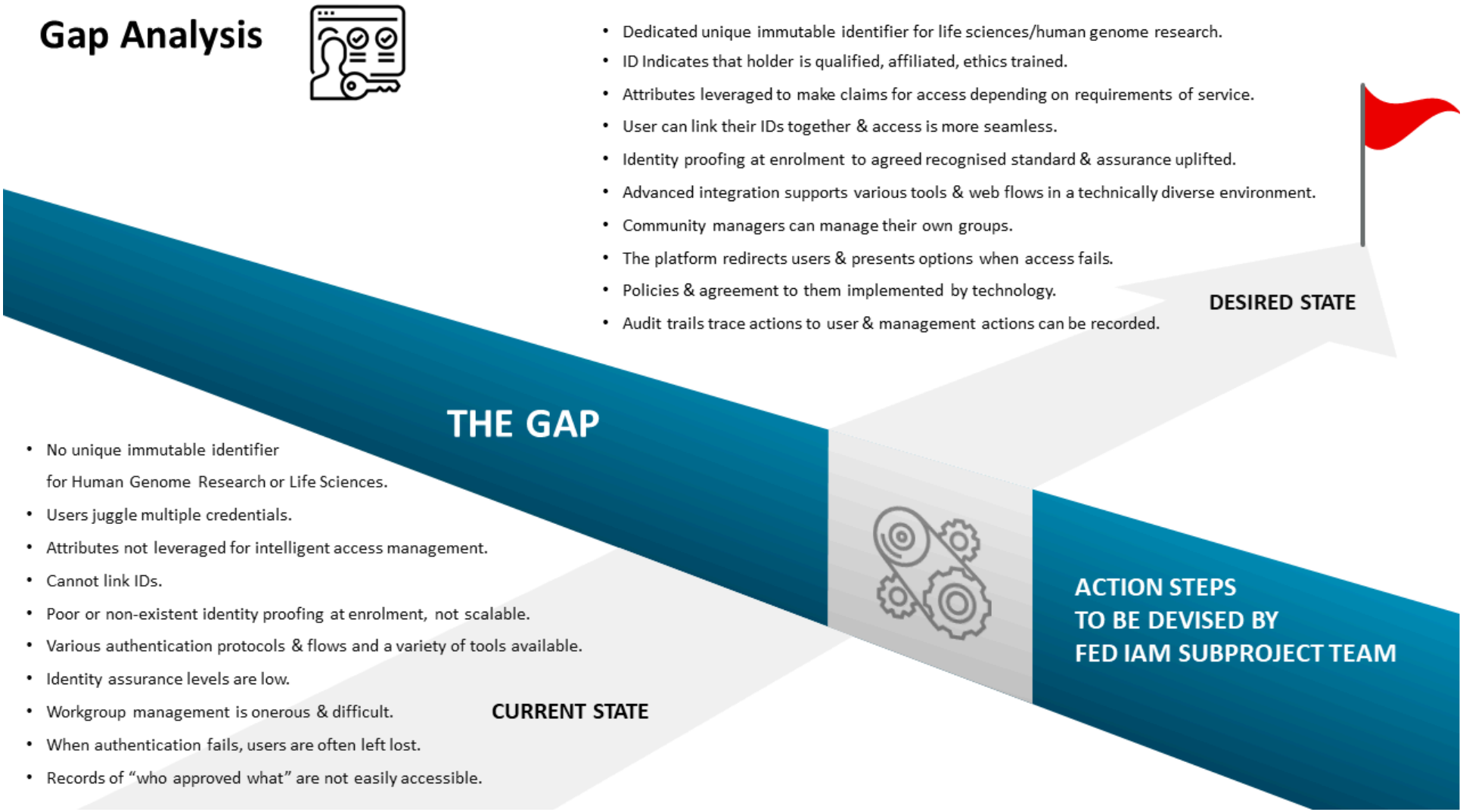


Figure 3 - Summary of Federated Identity and Access Management Gap Analysis

Table 4 *- DEDICATED UNIQUE IMMUTABLE IDENTIFIER for the HUMAN GENOME RESEARCH COMMUNITY & ATTRIBUTES FOR RELYING SERVICES*

HGPP REQ NUMBER	Related Reviewed & Scoped Requirement	Mandatory or Should Have	Current State
HGPPREQ-004	The platform consumes verified identities from users' home organisations and these are central to the authentication process to ensure verified identity.	M	**No unique identifier for life sciences or human genome research in Australia** ● Currently there is no unique ID for the human genome research community. ● Users are given several accounts at various institutions and juggle multiple credentials. ● User accounts are not always traceable to their home organisation affiliation. ● User accounts provide no attributes that may indicate a person's qualifications, affiliations or ethical training. ● When a user leaves an institution, it takes manual effort & time for their credentials to be removed. ● User attributes are not leveraged to manage arbitrary access to data, tools or workgroups. ● Identity standards are not compatible with international standards such as GA4GH.
HGPPREQ-005	The platform consumes verified institutional affiliations from users' home organisations.	M	
HGPPREQ-006	The platform represents and manages working group memberships and can assert these memberships to relying services.	M	
HGPPREQ-034	Data Access team members can be assured that researchers accessing data understand the need to maintain data anonymity and encryption, so as to reduce the need for them to provide one-on-one education and to ensure patient privacy is protected. The platform should have the capability to retain training/ethical quals attribute information and this is available for potential consumption by downstream services.	M	
HGPPREQ-041	The platform should support a qualified human genome researcher ID, indicating the researcher is of sufficient standing to access human genome data.	M	
HGPPREQ-052	The platform should be able to present user attributes to a number of different types of relying services, such as instruments (producing data for research purposes), data archives (managing data for secondary use), computing and cloud services (enabling researchers to compute on data) and various collaborative tools that support researchers' interaction (such as, wikis, content management systems, mailing lists or e-learning environments).	M	
HGPPREQ-074	The identity management system within the platform should incorporate other attributes relevant for relying services. (what attributes/ what method, are to be decided)	M	

Table 5 - *ABILITY TO LINK AND UNLINK EXTERNAL IDENTITIES TO THE UNIQUE IMMUTABLE ID*

HGPP REQ NUMBER	Related Reviewed & Scoped Requirement	Mandatory or Should Have	Current State
HGPPREQ-064	The platform should allow a user to link and unlink multiple external identities to a single Life Sciences/Genome Research Passport. To link additional identity a user should have to prove that they are able to authenticate with a new one as well as the one which is already linked. After the linking the user can use any of the linked identities to authenticate to their Life Sciences/Genome Research ID. (This also helps when a user is moving from one home organisation to another.)	M	**No capacity to link multiple IDs to one person based on a unique human genome research ID** ● As there is currently no unique human genome researcher ID, there is no identity linking capability to a human genome researcher ID.
HGPPREQ-065	The system should PROMPT USERS during registration of a new account when systems discover existing accounts with the same organisational account. THIS HAS A DEPENDENCY ON orgs NEVER REPEATING UNIQUE IDs. This is to prevent people accidentally creating duplicate identities.	M	
HGPPREQ-057	Other user identities may be able to be linked to the primary source of identity, once the primary source of identity has been verified. These could include eduGAIN, ORCID, Google -WHICH LINKED ID types are eligible needs to be determined. DESIGN OF THIS WILL BE NUANCED BECAUSE SOME PEOPLE HOLD TWO PRIMARY IDENTITIES	M	

Table 6 - *IDENTITY PROOFING (PRIMARY SOURCES) FOR REGISTRATION TO PLATFORM & ISSUANCE OF UNIQUE ID*

HGPP REQ NUMBER	Related Reviewed & Scoped Requirement	Mandatory or Should Have	Current State
HGPPREQ-056	Primary sources of users' identities should be the users' external identity providers that are the most authoritative source for their role and affiliations in their home organisation (ie. professional/scientific/clinical affiliation).	M	**Poor or non-existent identity proofing practices, or participants are personally known, which is not scalable.** ● Most identity proofing is limited to the new user being known personally by the person enrolling them to another organisation's system, or a Google/LinkedIn search verifying they are working at the institute they claim to be. There is also a heavy reliance on the correct appearance of the person's email domain.

Table 7 *- AAI SOLUTION CAPABLE OF INTEGRATING IN A TECHNOLOGICALLY DIVERSE LANDSCAPE*

HGPP REQ NUMBER	Related Reviewed & Scoped Requirement	Mandatory or Should Have	Current State
HGPPREQ-044	The authentication service must support CLI tools.	M	**Currently AAI solutions are not integrated between partner organisations. But most partner organisations offer a range of tools (including CLI) and these should be supported by the new system.** ● Authentication protocols and flows vary between organisations (SAML, OAUTH, OIDC, Auth0).
HGPPREQ-045	The authentication service must support web flows (OAuth, OIDC, SAML)	M	
HGPPREQ-053	The platform should be able to handle compatibility issues in a distributed environment, where different actors have adopted heterogeneous technical approaches (this could be addressed via a **proxy**: an adapter between the external identity providers and the end services, able to do protocol and attribute translation when required by their specific needs).	M	
HGPPREQ-067	The system should provide endpoints (SAML 2.0 and OpenID Connect) to which the relying services can integrate to authenticate users and receive their attributes.	M	
HGPPREQ-075	The identity management system within the platform should provide credential translation, enabling an end user to receive an X.509 certificate or other short lived token if required for accessing a particular relying service.	M	
HGPPREQ-076	The identity management system within the platform should support complex protocol flows, such as the device code flow of OpenID Connect that enables human genome research system AAI login with SSH Secure Shell.	SH	

Table 8 *- WORKGROUP MANAGEMENT CAPABILITY ALLOWS COMMUNITY MANAGERS TO ARRANGE THEIR OWN TEAMS, DATA & TOOLS*

HGPP REQ NUMBER	Related Reviewed & Scoped Requirement	Mandatory or Should Have	Current State
HGPPREQ-046	The platform should facilitate arbitrary subsets of users: to limit access to tools, data and storage etc to managed workgroups developed by a person and within the platform.	M	**Current community and workgroup setup is technically complex and not easily managed.** Most organisations currently allow users to be managed in discrete groups along project lines and it is essential that this capability is not lost. But also...community workgroup management is currently: - technically complex - insufficiently granular - lacks a good user interface. Current community management does not allow community managers to define user flows.
HGPPREQ-058	The identity management system within the platform should provide a user interface where delegated community managers can manage their users, organise them in groups and add access to data and objects for them.	M	
HGPPREQ-059	The identity management system within the platform should enable community managers to define registration flows to individual groups including user life cycle in the groups.	M	

***Table 9** - AN IMPROVED USER EXPERIENCE, WHILE MAINTAINING RIGOROUS SECURITY*

HGPP REQ NUMBER	Related Reviewed & Scoped Requirement	Mandatory or Should Have	Current State
HGPPREQ-054	Platform interacts with users to redirect flow. Typical use cases to include: - to redirect new users to registration; - to check if the user is authorised to use the service which they are trying to access; - if the user doesn't fulfil all conditions, the system should offer a solution in the form of immediate action or to a page where they can request access.	SH	**Current levels of identity assurance are low. Also, for the user, a failed authorisation provides few clues as to the cause or links to rectify.** ● Currently, when a researcher wants to access resources at another organisation, the organisation provides the user with a set of credentials specific to the organisation and so they have control over whether there is a requirement to use multi-factor authentication or other method of increased identity assurance. (*For any new solution, it is critical that organisations with "visiting researchers" accessing data or apps are able to specify the level of identity assurance uplift they require. An agreed baseline for the infrastructure will be able to be increased by providers of data and services but not decreased below the agreed minimum baseline.)* ● Currently, because separate credentials are required for sign in to each organisation, single sign on is not available across organisations. ● Currently, organisations leverage their organisation's information technology team or systems engineers within their own teams to manually provision access for users from other organisations.
HGPPREQ-055	If the service a user is trying to access requires stronger authentication/higher levels of assurance and it has not been covered by the external identity provider, the platform should be able to enforce users to provide a higher level of assurance. Use cases include: - MFA; (not centrally managed MFA - rather sending back to IdP to get higher levels of assurance). THIS HAS A DEPENDENCY ON MFA AND HIGHER LEVELS OF ASSURANCE BOTH BEING A REQUIREMENT	SH	
HGPPREQ-062	If a home identity provider supports single sign-on, the platform should allow this to cover authentication to the platform. The service can force a re-authentication if they deem necessary.	M	
HGPPREQ-068	The platform should provide identity provider discovery service(s) for an end user to select their authentication provider.	M	
HGPPREQ-071	The platform should enforce access control, enabling a relying service to request the Proxy Identity Provider to deny access for users who don't qualify to the access policy (e.g. group membership) configured for the relying service.	M	

Table 10 *- INFRASTRUCTURE & COMMUNITY POLICIES ARE IMPLEMENTED AT USER ENROLMENT & AT AUTHENTICATION POINTS*

HGPP REQ NUMBER	Related Reviewed & Scoped Requirement	Mandatory or Should Have	Current State
HGPPREQ-085	The platform must provide a process to ensure that all users are aware of, and accept the requirement to abide by, the Acceptable Use Policy (AUP). One mechanism is: The policy must be presented at the time of registration and the new user must read and accept before completing registration.	M	**The current state of policy enforcement at partner organisations is unknown.** **For any future solution, technical implementation of policies will be needed for:** ● Acceptable use ● Privacy ● Infrastructure standards ● Community specific standards ● Service provider standards and ● Data protection.
HGPPREQ-086	The platform must present to users any changes to the AUP and/or additional restrictions or requirements on acceptable use that arise out of new collaborative partnerships (if any). Users must reaffirm commitment including changes.	M	
HGPPREQ-088	The platform must present to all Collections of users** who use the Infrastructure, and have them accept the need to abide by, applicable Infrastructure policy requirements applicable to Collections of users.	M	
HGPPREQ-102	The platform must display the common aims and purposes, i.e. the research or scholarship goals of Collections of users** and have a mechanism to display changes	M	
HGPPREQ-103	The platform must have a method of displaying the Data Protection Policy and ensuring Constituents, Users and Collections of Users who process personal data in/associated with the platform agree to the policy	M	
HGPPREQ-104	The platform must display a privacy collection notice to users, stating the personal information collected from platform users and how that data will be used.	M	
HGPPREQ-105	The platform must implement identity assurance to the level specified by the Infrastructure's policies for minimum requirements.	M	
HGPPREQ-106	The platform must implement authentication to the standard specified by the Infrastructure's policies for minimum requirements.	M	

Table 11 *- AUDIT TRAILS LINKED TO VERIFIABLE USER ID, TIME-STAMPING & REPORTING*

HGPP REQ NUMBER	Related Reviewed & Scoped Requirement	Mandatory or Should Have	Current State
HGPPREQ-091	The Platform must be able to identify a person that has taken an inappropriate action via a personal traceable unique ID and provide the ability to record membership management actions as these may be needed in security incident response. **(We are imagining something like an audit trail with facility to input notes)**	M	**Currently, records of who made access approvals and what they approved, are not easily accessible.**
NEW	*NOT LISTED AS REQUIREMENT YET IN FEDERATED IAM BUT ARISES FROM ABOVE REQUIREMENT HGPPREQ-091. ACROSS THE PLATFORM THERE WILL BE PLACES AND TIMES THAT THE USER'S UNIQUE ID SHOULD BE CAPTURED IN AN AUDIT TRAIL - FOR EXAMPLE A USER IS A DAC COORDINATOR AND THEY APPROVE THE ACCESS REQUEST - EITHER THEIR ID OR NAME SHOULD BE CAPTURED AS HAVING APPROVED (AND IF CAPTURED AS USER FRIENDLY NAME THEN THE ID SHOULD BE BEHIND THAT) AND MAYBE THE USER SHOULD NEED TO AUTHENTICATE OR VERIFY THEMSELVES TO DO CERTAIN ACTIONS?*	M?	
HGPPREQ-069	The platform should provide service level reporting from e-infrastructures to the user community. Report access should be controlled so that managers at organisations can access reports on their own organisation, without violating the privacy of other organisations.	M	

5. Other Standards and Global Projects for Benchmarking

Europe and the United States are well into multi-year programs to improve secure sharing of human genome data. We have reviewed the activities undertaken by international projects and organisations because:

- It is important to support international standards to ensure compatibility between this work and other international activities; and
- If we can leverage work already completed, we will achieve maturity faster in our own solutions.

As part of the candidate solution analysis phase for the Federated IAM Sub-Project, we will assess the suitability to the Australian human genome research community of selected work packages and products. The selected packages, their source and the reason for canvassing them are summarised below in Table 12.

Work Package/Product	*Source*	*Why Canvas This Item?*	*Who, What and How?*
Passport/Visa 1.2	GA4GH DURI Work Stream	Open-Source, widely used tech (JWT), holds details describing researcher attributes/qualifications in the form of a bearer token that is consumed by a technology solution (Clearing House) to control access to secure resources.	Garvan performed limited initial tests using an earlier specification (1.0). We may leverage this work or run a test alongside the pilot. GA4GH Github documentation is plentiful. We reviewed this specification when drafting our sub-project requirements. Waiting on Passport specification 2.0 to be finalised.
ELIXIR AAI	ELIXIR Led by Mikael Linden and Dominik František Bučík	Open Source, good uptake (6400 logins per month), already in use with ELIXIR Beacons for Human Genome data, already working with GA4GH, uses components with known capability (COManage, Perun). This infrastructure will soon be migrated to operate using the Life Science Login service (https://lifescience-ri.eu/home.html)	Discussions are scheduled with Mikael Linden. A desktop study will be completed then further activity plans drafted.
WP5: User management and access services (AAI) work package	EOSC Life (also co-led by Mikael Linden of ELIXIR Finland)	Real life multi-organisation implementation of a federated Life Sciences AAI including an identifier; based on AARC Blueprint; very well documented so we can leverage to avoid making preventable errors.	We reviewed this specification when drafting our sub-project requirements. Discussions are scheduled with Mikael Linden. A desktop study will be completed then further activity plans drafted.
CILogon	CILogon	Allows researchers simplicity in login but also supports workgroup management. Used by NIH. Proven locally - between ZERO and UMCCR in 2021, still in production.	CILogon has demonstrated their product. we have set up a test instance of CILogon and will be testing through scenarios and personas
AARC BLUEPRINT	IGTF	Comprehensive policy documents applicable to AAI solutions. Mature and openly shared. We can leverage their maturity.	We based policy sub-project requirements partly on specifications from the AARC Blueprint. Policy discussions and workshopping ongoing throughout the HGPP.

Table 12 - *GLOBAL PROJECTS AND PRODUCTS WITH CANDIDATE SOLUTION POTENTIAL*

6. Conclusion

Technical capability, processes and systems are mature in the national human genome research community, however data, tools and resources are highly siloed. Processes and supporting technologies vary considerably between different organisations, often relying on email trails for delivery. The discovery phase of the HGPP has uncovered several contributing factors that fall within the remit of authentication and authorisation expertise to address.

The absence of a trust network between the flagship genome research organisations impacts their ability to leverage maximum value from their own and each other's holdings. Reasons for this absence are articulated fully in scoped requirements. High priority current state issues include:

- inability to present attributes necessary to confirm the identity of a researcher and make decisions about access to services;
- low or unknown levels of identity proofing;
- low and variable levels of identity assurance;
- absence of easy-to-reference records and management policies that would ensure users acted appropriately and records of previous actions were transparent.

We mapped the current state of processes, technology landscape, roles and tools during the discovery phase. Qualified subject matter experts have confirmed their most pressing requirements and produced a gap analysis. The team has produced a traceable outline of the Minimum Viable Product with key features including:

- a dedicated unique immutable identifier with suitable attributes for relying services;
- identity proofing (primary sources) for registration to platform;
- advanced integration capabilities;
- collaborative workgroup management capability;
- improved user experience, while maintaining rigorous security and access control;
- infrastructure and community policies implemented at user enrolment and at authentication points;
- audit trails for key actions linked to verifiable user id, time-stamping and reporting, together with a facility for recording management actions.

Subject matter experts within the team will now leverage their contacts in global projects to benchmark and canvass existing candidate solutions against the mandatory requirements and the known gaps. Components will be selected and we will formulate a backlog for federated identity management system development for the national community based on high priority requirements. This system will be piloted and tested against the requirements in order to recommend pathways for future production implementation.

References and Links

General References

Auth0	https://auth0.com/
AARC Policy Development Kit	https://aarc-project.eu/policies/policy-development-kit/
CILogon	https://www.cilogon.org/
ELIXIR	https://elixir-europe.org/about-us
EOSC-Life	https://www.eosc-life.eu/
GA4GH	https://www.ga4gh.org/
HGPP	https://www.biocommons.org.au/hgpp
NIH RAS	https://datascience.nih.gov/researcher-auth-service-initiative
OAuth	https://en.wikipedia.org/wiki/OAuth
OIDC	https://en.wikipedia.org/wiki/OpenID#OpenID_Connect_(OIDC)
openLDAP	https://en.wikipedia.org/wiki/OpenLDAP
RAF v1.0 (REFEDS Assurance Framework)	https://refeds.org/assurance
REFEDS	https://refeds.org/
SAML	https://en.wikipedia.org/wiki/Security_Assertion_Markup_Language

<<Federated IAM Sub-Project Team Artefact links have been redacted for internet publishing. For a copy of a cleansed artefact example please contact author Kylie Davies via Australian BioCommons.>>

User Stories

User Story Validation Survey Responses

Scoped and Reviewed Requirements

Gap Analysis

Endnotes

1. Icons from the Noun Project: search by Flatart, database by Start Up Graphic Design, identified by Tippawan Sookruay, group by Gregor Cresnar, Data File by Blangcon, Unlock by Arthur Shlain, archive by Adrien Coquet, support by Komkrit Noenpoempisut, documentation by lastspark, Scientist by Maxim Kulikov.)