AUSTRALIAN ACCESS FEDERATION

Authentication and Authorisation Infrastructure Requirements Document

GUARDIANS

12 June 2025

Prepared by: Ryan Paytes, AAF Business Analyst

Revision History

Revision Date	Summary of changes	Reviewer	Approver
07/05/2025	v0.1 Document created based on insights from discovery workshops with all partners and initial requirements validation with UMCCR and Garvan	UMCCR/Garvan/CCI	N/A
12/05/2025	v0.2 Document review by AAF Project team	AAF Internal	John Scullen/Sarah Thomas/Poornima Mani/Patrick Carnuccio
14/05/2025	v1.0 Document created		
14/05/2025	v1.1 Formatting cleanup	Patrick Carnuccio	
9/06/2025	V1.2 Implementing BioCommons feedback	Ryan Paytes	

Contents

Executive Summary

This requirements document has been developed as part of the discovery phase for **Stream 2: Foundational National Infrastructure, Project 2B: Trust & Identity (Long Term)** of the GUARDIANS program. It consolidates input from partner organisations to inform the design and implementation of Authentication and Authorisation Infrastructure (AAI) that will support secure, federated access to genomics services and platforms across Australia.

The discovery phase engaged six key partner organisations—UMCCR, Garvan, CCI, NCI, BioCommons/UoM, and QIMRB to understand current and planned identity, authentication, and authorisation practices, user onboarding and access provisioning workflows, technology stacks, and service integration needs. Workshops focused on identifying both current-state environments and desired future-state interactions with GUARDIANS services.

Key Discovery Findings

- Partners are at varying stages of readiness to deploy services to a test environment. Some partners are using AAF for login, with one partner already using an AAF AAI, albeit a Test environment in a production capacity. Most of the partners report use of internal system for managing users, like EntraID, Auth0 or similar.
- Manual and fragmented onboarding processes are common across partners, particularly for external users. There is strong interest in streamlining and standardising these processes while maintaining necessary governance controls
- All partners support using AAF for user authentication, but there is a shared need to support non-AAF users (e.g., external researchers and collaborators) through alternative trusted identity providers. A seamless onboarding experience for these external users will be critical.
- Partners have indicated that authorisation will be managed at the service level. However, several partners expressed interest in consistent support for group and role-based access models across services.
- Identity validation and assurance are critical for enabling appropriate access for users from non-AAF subscribing institutions. Further alignment and discussion are necessary across GUARDIANS partners to define how identity proofing, organisational affiliation checks, and access governance will work for external users. This same process can be applied to users from AAF subscribing organisations to improve identity validation and assurance for those users if GUARDIANS considers this necessary.
- Two identity strategies emerged through partner discussions, highlighting a key decision point for GUARDIANS. The first involves using persistent institutional identifiers (e.g. AAF-issued eduPersonPrincipalName (ePPN) to provide user access across services. The second supports identity portability, allowing users to retain access as they move between institutions. However, persistent identifiers are not inherently portable, while supporting portability would require identity linking and GUARDIANS minting and using persistent user identifiers across services.
- NCI will enable access to their datasets via their existing identity management platform for phase one of the GUARDIANS project with integration to AAF AAI planned for phase two. High-level requirements will be captured during phase one and AAI integration to support NCI will occur in phase two.

Integration Scope Summary

Below is an overview of each partner's GUARDIANS services and whether integration with AAI solution is required:

Institution	Service	AuthN	AuthZ
UMCCR	Jupyter	Yes	No
	Elsa	Yes	No
	REMS	Yes	No
Garvan	REDCap	Yes	No
	CTRL	Yes	No
	REMS	Yes	No
	RDDP	Yes	No
	Elsa	No	No
CCI	Patient Data Portal	Yes	No
	FHIR	No	No
	Beacon	Yes	No
QIMRB	Globus	Yes	No
	REMS	Yes	No
	Beacon	No	No
BioCommons	OMIX3 (Gen3)	TBC	TBC
	*REMS	Yes	TBC
	Beacon	TBC	TBC
NCI	EGA	TBC	TBC
	Gen3	TBC	TBC
	REMS	TBC	TBC
	Beacon	TBC	TBC

Note: Integration priorities will be refined during the design phase

*REMS service to be implemented on behalf of CCI

Pilot Partner Pathway

Based on implementation timeframes and technical readiness identified during discovery, CCI, Garvan, and UMCCR have been selected as potential pilot partners for phase one. The insights captured through partner discovery sessions—along with upcoming technical design work—will form the foundation for the next phase and support the broader, ongoing development of the GUARDIANS program.

Project Background and Context

Purpose

The GUARDIANS project aims to transform access to Australia's genomics infrastructure from a fragmented system into a coordinated, scalable and interoperable national framework. Australian Access Federation (AAF) is a key Collaborator in the GUARDIANS Program, and it supports Australian researchers in human genomics by providing secure, seamless access to services through an Authentication and Authorisation Infrastructure platform and policy framework.

Objectives

The objective of the project outlines specific, measurable goals that the project seeks to achieve providing a roadmap for the project team to track. The key objectives of this project are to:

- Deploy test and production environments for authentication and authorisation with role-based access control features available to applications in the GUARDIANS ecosystem.
- Implement single sign-on (SSO) capabilities to improve user experience and efficiency.
- Integrate AAI platform with a range of services ensuring successful authentication and authorisation.
- Develop a baseline policy set covering authentication assurance, acceptable use, privacy, membership management, personal data processing and security protocols to set shared expectations for identity and security across services implemented under the GUARDIANS program.
- Enable researchers from non-AAF subscribing organisations to access GUARDIANS infrastructure.

Expected Outcomes

- A fully operational AAI platform supporting access to multiple services connected by partner organisations. A common view of the researcher's identity across the collaboration will enhance research efficiency, collaboration and security for all participants.
- Sustainable infrastructure supported by comprehensive policies for long-term viability and continuous improvement.

Discovery Context

This requirements document has been developed as part of the initial discovery phase for **Stream 2: Foundational National Infrastructure, Project 2B: Trust & Identity (Long Term)** of the GUARDIANS program. It captures input from partner organisations to inform the design and implementation of authentication and authorisation mechanisms that will underpin national-scale genomics infrastructure.

The focus of this discovery is to define how users will be identified, authenticated, and authorised to access services across multiple organisations while accommodating diverse institutional requirements, assurance levels, and governance models.

This document includes a consolidated view of:

- Current state practices for identity, authentication and authorisation across partner environments
- Partner-specific needs, constraints, and expectations
- Future-state requirements to guide design and integration of a federated AAI platform

Discovery Approach

Methods Used

The GUARDIANS project follows the Australian Access Federation's (AAF) **Trust and Identity Framework** implementation approach, which is based on GÉANT's **Authentication and Authorisation for Research and Collaboration (AARC) Blueprint**, an internationally recognised framework for implementing federated access management solutions for international research collaborations. This approach emphasises **co-designing, developing, and testing trust and identity** solutions in partnership with research communities and facilities through an incubator-style program.

The method comprises five key phases:

1. **Discover** – Understand current-state identity, authentication, and authorisation practices; gather stakeholder needs
2. **Define** – Shape and prioritise future-state requirements (this document); begin development of context-level architecture
3. **Design** – Translate requirements into Solution Architecture Design document, define integration patterns, etc.
4. **Develop & Test** – Build and iterate AAI components; test integrations with partner services and validate functionality
5. **Deliver** – Deploy production-ready integrations; establish support and maintenance structures for operational readiness

Stakeholder Engagement

As part of the discovery process, **discovery workshops** were conducted with each participating organisation. These sessions aimed to:

- Identify current-state identity, authentication and authorisation practices
- Understand how user onboarding, credential management and access provisioning are handled today
- Explore partner-specific challenges and pain points related to identity and access
- Understand planned integrations with GUARDIANS services and related future-state needs
- Identify user types, user stories, key journeys, and system touchpoints
- Identity those GUARDIANS partners best placed with the capacity and capability to progress the project

Each session was followed by a playback and validation meetings with selected partners to refine and confirm the information gathered. A discovery session with the BioCloud team is still pending. A summary of the key deliverables and timeframes for each partner is available in the GUARDIANS Deliverables Summary

Transition to Design Phase

The insights gathered during **Discovery** and **Define** phase (current phase) will inform the **Design** phase, where a detailed technical architecture for the AAI solution will be developed. The requirements captured in this document — alongside identified integration points and user journeys — will guide system design, policy development, and the prioritisation of implementation activities. This approach ensures the resulting AAI solution is practical, scalable, and aligned to the operational needs of researchers and service providers across the GUARDIANS ecosystem.

Participants

This discovery phase involved AAF's consultation with several GUARDIANS' partner organisations to understand their identity, authentication, and authorisation needs. The table below outlines the institutions and the stakeholders who participated in discovery workshops.

Institution	Name & Role
BioCommons	Kylie Davies, Senior Business Analyst Conrad Leonard, *Technical Lead* Jesseka Chaddertion, *Project Manager* Yassaman Alavi, Project Manager
Children's Cancer Institute (CCI)	Marie Wong-Erasmus, *Project Manager* Kamile Taouk, *Project Lead*
University of Melbourne Centre for Cancer Research (UMCCR)	Andrew Patterson, Software Engineer Lead Marko Malenic, Software Engineer
Garvan	Sarah Kummerfeld, Chief Scientific Officer Katherine Champ, *Project Manager* Leonard Goldstein, *Data Science Lead* Tim Ho, REMS Lead Ignatius Menzies, *CTRL Lead*
QIMRB	John Pearson, Bioinformatics Manager Scott Wood, Software Engineer
National Computational Infrastructure (NCI)	Warren Kaplan, *Science Lead* Andrew Howard, Associate Director David Monro, Linux Systems Administrator Dmitry Degrave, Principal Software Engineer Pravin Mangukia, Manager, Cloud Services Matthew Downton, Associate Director
Australian Access Federation (AAF)	John Scullen, Collaborator Lead Poornima Mani, Project Manager Anurag Katariya, Solutions Architect Patrick Carnuccio, Senior Solutions Analyst Ryan Paytes, Business Analyst

Discovery Findings

This section summarises key findings from the discovery phase, organised into current state practices and anticipated future state directions across partner organisations. These insights provide the foundation for defining technical and policy requirements for the GUARDIANS trust and identity framework.

Background, Current State and Future State Summary

BioCommons	
Background & Current State	**Proposed Future State**
BioCommons BioCloud team plans to implement Beacon, REMS (shared GUARDIANS instance) and Gen3	Users of OMIX3 platform will explore data, identify useful cohorts and apply for access to underlying data – which will include an application form with purpose of research, preferably followed by a personal contact/onboarding interaction (e.g., Zoom, in-person)
OMIX3 is a data commons platform comprising Gen3, central Beacon & Beacon network for data discovery Gen3, a component within OMIX3, uses OIDC flows, with one microservice for authentication and another for authorisation	
Auth0 is the service currently used for identity and access management (IAM) Current IAM onboarding is a manual process	Anticipates onboarding external users (non-AAF), including Hospitals, MRIs (Medical Research Institutes), possibly individuals unaffiliated with GUARDIAN partners Expects a common AAI platform across all GUARDIANS partners
Identified the following IAM access levels: **Anonymous –** Basic public view (landing page, dataset counts) **Authenticated –** Logs user activity only, minimal data visibility **Authorised –** Get access to bulk and detailed data after approval	Future identity states include: **Unauthenticated** – Minimal access **Authenticated** – Unauthorised with limited metadata visibility **Authorised** – Providing full data access
	AAF subscription model supports onboarding via MRIs' HR processes, pushing onboarding costs to MRIs, who benefit from access, help resolve identity validation burden for those users (though possibly not all).

CCI	
Background & Current State	**Proposed Future State**
Currently uses AzureAD (EntraID) for authentication. Azure directory includes most internal CCI staff and an unknown number of external clinicians/collaborators as guest accounts with a CCI email address. External clinicians are onboarded every two weeks and users sign in via Azure single sign-on to CCI external services.	The AAI discovery service will include social identity providers (for non-AAF users). A Transparent Enrolment process will onboard new users, after authentication to streamline the registration process. Transparent enrolment does display a terms and conditions acceptance UI. The UI will need to configure and deployed at each service.
Several internal CCI users hold hospital or research institute credentials	CCI user base to include AAF subscribers within GUARDIANS, non-AAF researchers, external clinicians and public users. There may be the potential for thousands of public users interacting with the system
CCI will deploy a public-facing data commons portal with a subset of public data and connect to a Beacon network in the next 3 months, hosted by BioCommons Beacon service is ready for AAI integration in both test and production environments. Testing can commence once GUARDIANS-specific customisation is applied.	AAI will be integrated with the data discovery portal and the REMS service. Additional identity validation will occur during the data access request phase in REMS, with possible manual or automated checks
Beacon allows open access to high-level statistical plots without requiring login. CCI currently operate a Beacon node in production against a test CILogon instance to provide access to data stored in Amazon S3	Once access is granted, users will be able to analyse datasets. Access to datasets will have expiry dates
Researchers must undertake Privacy training (e.g., from UNSW or ICH) before access to data is granted	
CCI also operate ZeroDash which supports 100–120 internal users (out of scope)	
	Tiers of access to GUARDIANS services will include public, restricted, and sensitive
	No invite-based access model – users will self-register and request data

UMCCR	
Background & Current State	**Proposed Future State**
Currently uses Google Suite with users under the @umccr.org domain, with internal portals in place for UMCCR user's authorisation process.	Internal portals will be integrated with AAF for authentication (AAF will be used as source of authentication for all internal staff) and manage authorisation internally. User identities will be linked to a stable identifier within the portal for permission control (e.g. ePPN)
UMCCR is onboarding Peter Mac to use GUARDIANS services within the next 6–12 months	Future GUARDIANS users (24 months and beyond) may include Medical Research Institutes (e.g., Royal Children's Hospital), clinicians, and visiting researchers or external collaborators — including non-AAF and Gmail accounts. These users, who may require occasional access, could include international members and non-admins. (CILogon can be used as to accommodate Gmail (personal or organisational) & ORCID linking)
UMCCR staff are using the University of Melbourne's Okta IdP. This leverages UoM identity verification and validation processes. UMCCR will/is not using UM Okta IdP to enable access for PMCC user access	UMCCR will shift identity management to AAF as the central identity source
UMCCR is using htsget, a protocol used to fetch specific genomic data slices from large files. It acts as a layer on top of object storage, serves as an inception point for authorisation, allowing for granular, nuanced authorisation logic, and is commonly used in Gen3, where it works alongside authorisation information to extract only permitted data	Will integrate Jupyter, Elsa, and GUARDIANS' REMS services with the AAI solution
Super Admins are included in the initial Elsa deployment. They are responsible for creating projects, can "bless" or approve users to receive elevated permissions. This model limits broader permission grants, ensuring super admin approval is required for key actions	Internal staff will be required to complete data access requests. Additional governance steps/layers will be
Authorisation decisions (e.g., who can access what data) will be handled by UMCCR within REMS.	External users will be onboarded and allocated an entitlement after identity validation. Once verified, they are granted access to the service.
	There is a long-term vision to scale data sharing nationally — this should be considered in requirements planning
	The Genomic ordering system will intersect with Garvan and QMIR, especially for ICGC data storage. The goal is for Garvan/QMIR to act as a host, with the platform pulling data from them (service-to-service integration)
	UMCCR want to detect a user's organisational affiliation early in the authentication flow. After

	authentication, a check should verify if the user belongs to a known organisation. (For purposes of granting or restricting access as part of transparent enrolment)
	Integrations with GUARDIANS partner services will require consistent user identifiers across systems to support seamless data movement on behalf of users.
	UMCCR will demonstrate first AAI integration capability with portal service by 30 May 2025
Garvan	
Background & Current State	**Proposed Future State**
Garvan is implementing a dynamic consent platform (CTRL) where patients can log in via web interface to upload data and manage consent.	AAI will provide centralised authentication across CTRL (except for patients), REMS, REDcap, and RDDP (non-priority/stretch goal), however authorisation will remain at the service level Garvan will stand up CTRL, REMS, and REDCap ready for integration by November 2025. A full demo of the tech stack (CTRL, REMS, REDCap, and RDDP) with all integrations in place by May 2026
CTRL will support researchers/admins sending surveys and managing participant data.	Intended GUARDIANS service users include Australian institutions, overseas collaborators, hospitals, and commercial researchers. Internal users are sometimes clinicians who move between hospitals and research organisations. Desire to maintain minimal login barriers for patients A process is needed to verify the identity of hospital staff who are not part of an AAF subscriber.
Participants are invited to participate and access CTRL (not open registration)	
CTRL integrates with REDCap, which uses its own authentication (AAF, org AD, or Okta) and is accessed by clinical teams.	CTRL will support MFA (for admins). There is interest in SSO across services, allowing admin and clinician users to access REMS, CTRL, and Elsa seamlessly
Currently, CTRL uses Auth0 for auth, with email as the identifier Most identity references across CTRL, REMS, and Elsa use email addresses, but these are inconsistent and problematic.	Garvan envisions a federated identity platform supporting AAF subscribers, eduGAIN institutions, Gmail, Microsoft and GitHub All platforms must move away from email identifiers toward a persistent identifier Integrations with other GUARDIANS partner services will require a common persistent user identifier.
REMS is used for data access requests, with approval by a data access committee. It is	

locally configured via Auth0, using local and Gmail accounts for authN	
Elsa (data release coordinator) is connected to REMS and automates approved data releases.	
Sensitive data is stored in REDCap	
	Role-based access control is implemented at the service level, with different user roles and permissions per platform
QIMRB	
Background & Current State	**Proposed Future State**
Maintain and manage 10,000 genome equivalents worth of data, 100 GB for one 30x genome (300 GB raw; ~600 GB end state per genome)	
Currently QIMRB store data locally	Not all data sets will live permanently in the encrypted environment; infrequently used data may remain on QIMRB's internal network and only be pushed out via Globus when needed.
Deposit into "journal repositories" is required, typically in Europe	
Data transfers to/from Europe takes several months	
Currently planning a segregated secure network for QIMRB - separate internet connection, separate security devices; No privileged internal connection from QIMRB, they will access the network as an external party—leaving QIMRB's own network and re-entering as an external user	Encrypted data sharing model - encryption inside QIMRB. Encrypted copies pushed into the isolated network. Mechanisms will be implemented to support secure data sharing Services expected to live in the network are Beacon (Top Level), REMS for authorising data requests and Globus for data transfers
Only encrypted copies of data intended for sharing will be moved into the isolated network	Encrypted public/private key access on top of Globus' security layer. QIMRB will import public keys. Keys may be rotated; new encrypted versions re-pushed as needed
EGA (European Genome-phenome Archive) currently acts as an intermediary. Handles user authentication and data access request, confirms with QIMRB before sharing data (reliant on EGA for IAM processes)	REMS will validate all required user credentials, tie credentials directly to the data access request – will require AAI integration Globus will/may also require AAI integration Access model will support global users, with a particular focus on Australia
4-5 data requests per month currently	Anticipate small number of high-volume users, many users requesting small data sets. Access process will remain consistent regardless of data size

All data shared is personal health information (genomic). No unrestricted access — all access is mediated and evaluated.	Research plan required for all access requests; Emphasis on vetting and validating source of users before access is granted; All users (internal & external) go through an access approval workflow
Acceptable use policy exists on the specific data sets	Policy – expectation is that GUARDIANS partners would be treated as any other organisation when it comes to data sharing
	Institutional, academic, or commercial affiliation necessary — Gmail not accepted; non-AAF users expected and will be able to access the platform (reputable users)
	Data classification model - Controlled (personal data), Semi-controlled (e.g., data about cancer), Uncontrolled (rare, limited data in this category)
	Aim to commence Globus and REMS testing as early as July.
NCI	
Background & Current State	**Proposed Future State**
NCI uses LDAP for authentication and identity management where users are assigned persistent Unix identifiers	NCI intends to retain their use of local LDAP accounts in the GUARDIANS future state environment
Users' email addresses are captured at registration and users are issued NCI identifiers and credentials. When users move to a new institution, users may request an update of their NCI details	Persistent NCI credentials will allow users to move institutions and retain access by updating their details. Additional investigation will be necessary to support NCI and develop a process to relink an NCI internal identity to a user's new AAF/GUARDIANS identity to support the change in home organisation.
Neuron cloud service users are manually provisioned and authenticate directly against NCI's LDAP Authorisation is enforced at the service access layer or login point with group and role aggregation managed internally based on user attributes	Service-level authorisation will continue, with access decisions and group memberships managed and enforced by the applications or login nodes
Account lifecycle management disables or deactivates accounts when access is no longer needed, based on expiration of the merit allocation Principle investigator (PI) registration requires a successful merit allocation which leads to project creation and then team members are registered against the projects After user registration through a portal, an approval notification is sent to the PI of the	Existing policies define how users are managed when moving between institutions, including clear ownership by scheme managers who will be responsible for managing user access to projects

project who has authority to admit users into the project	
Role-based access exists for PIs and delegated PIs having the ability to approve users within projects	
Users can belong to multiple projects and use the CHGRP Unix command to switch active project groups controlling which group permissions apply during session activity	
Some public services such as those for Geoscience Australia and the Australian Copernicus Hub are managed using a Keycloak service that is skinned and maintained by NCI on behalf of external partners	External Keycloak services such as those deployed for Geoscience Australia will continue to operate in the future state with no major architectural changes expected for these services
Some services rely on service accounts rather than individual user accounts, for analysis pipelines and automated export of data/results	
NCI must comply with Defence Trade Controls Act obligations requiring positive user identification and affiliation with known trusted institutions	The existing Defence Trade Controls Act restrictions may be revisited and potentially expanded to allow access from users belonging to trusted international partners. Final definitions are pending.
A domain whitelist is relied on for identity proofing during onboarding with a manual vetting process initiated for other users	
Auditing of user activity is mandated, especially for access to HPC systems	NCI aims to maintain physical identification and traceability for all users accessing HPC systems and services ensuring that identity attribution remains strong across all activities
	Services such as REMS will be positioned as sector-wide services whereas internal NCI applications, like analysis pipelines, will be tightly scoped to NCI-approved users
	A consistent approach is needed to reliably identify individual users across multiple services and across different organisations

Ongoing Discovery and Pilot Progression

Partners are at varying stages of readiness in defining and implementing their authentication and authorisation models. The current findings represent a snapshot based on discovery workshops to date; further insights are expected as technical implementation progresses.

Pilot development will continue with the University of Melbourne Centre for Cancer Research (UMCCR), the Garvan Institute, and the Children's Cancer Institute (CCI). As the design phase commences, requirements will be refined, and integration planning will continue in close collaboration with all partners. These updates will be reflected in the Requirements Traceability Matrix, which will be shared with BioCommons and partner organisations prior to the Develop and Deliver phases.

High-Level Requirements

The requirements outlined in this section are high-level and have been captured through discovery sessions with all partners. Requirements from pilot partners (CCI, Garvan, UMCCR) have undergone initial review and validation (CCI pending validation). Requirements from non-pilot partners have been captured but are not yet fully developed and will require further refinement as we initiate the Design phase.

These requirements reflect current priorities and expectations but will be expanded into detailed functional and non-functional requirements. AAF can accommodate new or evolving requirements as the project progresses, with changes managed through an established change control process. The focus of this section is on AAI-specific components such as identity, authentication, and user access, which will be mapped to user stories and design features to ensure traceability throughout implementation.

UMCCR	
Requirement Number	**Requirement Description**
R01	The AAI solution **must** support federated institutional logins, including those from AAF subscribers and eduGAIN-affiliated institutions, to accommodate internal users and affiliated researchers.
R02	The AAI solution **must** support social identity providers, such as Google and Microsoft, to enable access for visiting researchers and unaffiliated international collaborators.
R03	Multi-Factor Authentication (MFA) **should** be supported and configurable per login pathway, with consideration for social or unaffiliated logins.
R04	The AAI solution **must** integrate with GUARDIANS services Jupyter, Elsa, and REMS, providing centralised authentication.
R05	The AAI solution **must** deliver a defined set of identity claims to UMCCR services to support local authorisation decisions within Jupyter, Elsa, and REMS. These claims must include name, email address, identifier (e.g., ePPN – eduPersonPrincipalName), and home organisation.
R06	AAF **should** act as the central identity provider for internal UMCCR users, ensuring consistent authentication across all GUARDIANS services. AAF identities must be link to internal user identities using a stable, persistent identifier.
R07	The AAI solution **should** assert a persistent user identifier (e.g., eduPersonPrincipalName) which supports user identification across services and other GUARDIANS partners
R08	The AAI solution **could** support organisation filtering to limit user access to services (e.g., Peter Mac, UoM)
R09	The AAI solution **must** support OpenID Connect (OIDC) as the protocol for service integrations.
R11	For non-AAF users, AAI onboarding **should** support identity validation and post-verification tagging/entitlement allocation
R12	AAI solution **should** interoperate with tools like htsget, which require fine-grained user-level access enforcement to specific genomic data slices.

Garvan	
Requirement Number	**Requirement Description**
R13	The AAI solution **must** support single sign-on (SSO) for Garvan services – REDCap, CTRL, REMS, RDDP (long-term/stretch goal).
R14	The AAI solution **must** support OpenID Connect (OIDC) as the protocol for service integrations.
R15	The AAI solution **must** support federation with a mix of institutional and non-institutional identity providers, including AAF subscribers, eduGAIN institutions, Google, Microsoft, GitHub, ORCID.
R16	The AAI solution **must** use a persistent, non-email-based identifier (e.g., ePPN or another stable unique ID) that is consistent across services and resilient to changes in user email addresses.
R17	The AAI solution **will not** support portable identifiers. Users must not retain identity continuity when moving institutions. Identity must be re-established under the new institution to meet ethics and audit obligations tied to data ownership by the host organisation.
R18	The AAI solution **must** support MFA signalling, with enforcement configurable per integrated service.
R19	The AAI solution **must** deliver a defined set of identity claims to downstream services, including name, email address, Identifier (ePPN – eduPersonPrincipalName), and home organisation.
R20	The AAI solution **must** support self-registration for non-AAF users, allowing them to select from approved identity providers. Each integrated service **should** support service-specific onboarding, including terms & conditions acceptance.
R21	The AAI solution **could** support Auth0 as an identity provider
R22	Administrative access to the AAI platform **should** be provided, allowing approved admins to track, monitor, and audit user access.
R23	The AAI platform **could** expose backend APIs, allowing integration with internal databases or external service workflows.
R24	The AAI platform **could** provide access to authentication logs, enabling audit trails and security monitoring as needed by participating services.

CCI	
Requirement Number	**Requirement Description**
R25	The AAI solution supports access to the CCI Data Discovery Portal for a broad user base, including AAF researchers, non-AAF researchers, external clinicians, and public users, through federated authentication
R26	The AAI solution supports authentication via multiple Identity Providers (IdPs), including AAF, eduGAIN, Google, and Microsoft.
R27	The AAI solution integrates with both the Data Discovery Portal and Beacon, and supports tiered access control, with different levels of access (public, restricted, sensitive) tied to user authentication.
R28	BioCommons will setup an instance of REMS (on behalf of CCI), which integrates with the AAI solution to support access request workflows and authorisation decisions
R29	The AAI solution enables new users (particularly non-AAF) to self-register and complete onboarding without invitations.
R30	The AAI solution login flow supports user terms & conditions acceptance.
R31	The AAI solution supports assigning entitlements to users once data access requests are approved.
R32	The AAI solution supports additional identity validation steps (e.g. or institutional vetting) as part of the data access workflow.
R33	The AAI solution enables CCI internal users to continue to be managed via Azure EntraID, while AAI will federate external users.
R34	The AAI solution has the functionality for admin to lock individuals out of accessing GUARDIANS services.

QIMRB	
Requirement Number	**Requirement Description**
R35	The AAI solution integrates with REMS to support federated authentication for users initiating data access requests
R36	The AAI solution supports authentication for external users from recognised institutional, academic, or commercial organisations (no access using Gmail)
R37	The AAI solution supports access for trusted users outside the AAF, through integration with approved external identity providers
R38	The AAI solution supports validation of non-AAF identities by consuming assurance attributes from upstream systems
R39	The AAI solution enables identity-based access control by consuming attributes (e.g., roles, entitlements) linked to validated access requests
R40	The AAI solution supports linking user identities to public/private encryption keys for secure data sharing.
R41	The AAI solution supports attribute-based access control (ABAC), enabling enforcement decisions based on attributes issued by REMS
NCI	
Requirement Number	**Requirement Description**
R42	The AAI solution enables users to register to GUARDIANS services and supports SSO login
R43	The AAI platform will delegate service-specific authorisation, enabling access to be determined at the application or login-node level, rather than centrally.
R44	The AAI solution supports Keycloak as an identity provider
R45	AAI supports manual and automated identity proofing workflows, including whitelist-based onboarding and manual vetting for users from non-trusted institutions
R46	Policy support is required to govern identity continuity when users move between institutions, including clearly assigned scheme manager responsibilities for access lifecycle management

BioCommons/UoM	
Requirement Number	**Requirement Description**
R47	The AAI solution provides an onboarding pathway for users outside of the AAF, including those from hospitals, MRIs, and potentially unaffiliated individuals.
R48	The AAI solution supports role-based access control and group management, with the ability to propagate access entitlements programmatically across integrated services.
R49	The AAI solution supports institutional login for most users, particularly those affiliated with Australian universities and research institutes
R50	Once data access is approved (REMS), the AAI solution should enable automatic assignment of users to relevant groups or roles, allowing programmatic access to authorised datasets.

Key Considerations, Challenges and Constraints

Outlined below are considerations, constraints, and open challenges identified during discovery that may influence the next phases of design and implementation.

- Identity assurance requirements vary by partner and user type (e.g., patients, clinicians, researchers), requiring a flexible model that supports both lightweight and high-assurance onboarding pathways.
- Managing users who change institutional affiliations is a common challenge. GUARDIANS and partners perceptions of how this should be managed is of interest to the AAF
- Supporting non-AAF users at scale introduces complexity around identity validation, attribute handling, and trust management, particularly for international or commercial collaborators.
- Authorisation remains decentralised, which preserves service-level control but increases the need for interoperability and consistent access models across services. This does not preclude services from assigning permissions based on data (group membership, roles, entitlements) supplied by the AAI.
- Technical environments differ across partners, with varying authentication platforms (e.g., Okta, Keycloak, LDAP) and service architectures, requiring careful integration planning.
- Some partners, such as NCI, do not anticipate integrating AAI components for at least 2 years. This impacts rollout sequencing, prioritisation of technical engagement, and interoperability.
- As of this document's preparation, the formal contract for this project stream is still in the process of execution. While this has not significantly impacted partner engagement, it may influence the timing of downstream activities
- An NCI identity assurance incubator led by AAF is running in parallel and is expected to inform and influence assurance models adopted by the GUARDIANS program. Findings from this incubator may shape how identity proofing, onboarding, and federation trust decisions are implemented across partner services.
- A key consideration emerging from partner discussions is whether GUARDIANS should prioritise the use of persistent institutional identifiers (e.g., AAF-issued ePPN) or support identity portability across institutions. This decision will influence how user continuity is handled when individuals change affiliations, particularly for researchers outside the AAF. While most partners want to maintain a consistent of view of the user, even when they move to a new organisation, several services explicitly prohibit this behaviour. They need to view a user coming from a different organisation as new and different identity, even if it is the same person.

Summary of Key Themes

This discovery has provided critical insights into the current identity, authentication, and authorisation practices across partner institutions participating in the GUARDIANS program. While each partner brings unique technical environments, service models, and governance considerations, several consistent themes have emerged that will directly inform the design of the Trust & Identity solution under Project 2B.

- Strong support for federated identity via AAF, with clear appetite for integration across internal and external services.
    - Need for seamless onboarding of non-AAF users, including researchers using Gmail and other external identity providers.
- Service-level authorisation will be maintained, with the AAI solution expected to integrate with and support existing local controls.
- Identity assurance and validation requirements vary, indicating a need for a flexible but consistent model that accommodates tiered assurance levels.
- Manual onboarding processes are common, presenting opportunities for streamlined workflows and improved user experience.
- Trust establishment with partner organisation and non-AAF/external users is a must before users can access to services provided under GUARDIANS. How these external users are managed needs to be determined.

Next Steps & Strategic Priorities

Progress Pilot Partner Engagement

- Prioritise pilot partner integration to demonstrate value, refine implementation patterns, and inform broader rollout.
- Establish a joint weekly working sessions with pilot partners (CCI, UMCCR, Garvan) to co-design the technical architecture, validate integration points, and confirm onboarding workflows.
- Maintain monthly check-ins with non-pilot partners to ensure ongoing alignment and visibility.

Finalise Remaining Discovery

- Complete targeted engagements with the BioCloud team.
- Initiate policy discussions to define MVP policy requirements following the onboarding of the Policy Analyst.

Develop Technical Design & Refine Requirements

- Begin development of the AAI technical architecture using discovery insights.
- Continue to iterate on requirements integration models, develop use cases and stub integration test cases in collaboration with partners.
- Validate non-pilot partner requirements.

Resolve Key Identity Management Decisions

Determine a path forward for:

- vetting and onboarding of non-AAF users
- a baseline identity assurance framework that accommodates varying partner needs
- operational ownership of shared partner services (e.g., BioCloud's REMS instance for CCI)
- select a primary user identifier strategy for GUARDIANS and connected services (refer to appendix)
- cross-institutional access remains a key consideration, particularly regarding which users from each partner organisation (e.g., Garvan, UMCCR) should be permitted to access services or datasets provided by others (e.g., NCI, QIMRB). This must account for varying access requirements and expectations across GUARDIANS.

Appendix

Selecting Primary User Identifiers for Research Communities

The use of common, persistent user identifiers such as eduPersonPrincipalName (ePPN) or samlSubjectID supports consistent user identification across services and is resilient to changes in email, affiliation, and other user attributes.

Persistent user identifiers are typically issued per user, per home organisation, and are intended to consistently identify the same user over time. For instance, the user can reacquire the same persistent identifier values after returning from an extended sabbatical.

Any identifier that links a user to an organisation naturally disables the portability of the value – the user cannot take the identifier value with them when they change organisation. A user's new home organisation will issue new persistent identifier values that have no correlation to the user's previous values or affiliation. A simple analogy is email addresses, which, although redirectable, are not transferable between organisations.

Persistent identifiers are excellent for identifying users across disparate services. Identifiers like ePPN and samlSubjectID are also human friendly. Though they appear similar to email address they are not deliverable email addresses.

The AAF creates persistent identifiers for home organisations by combining an organisation's internal user value with its domain name, e.g., user@vho.aaf.edu.au. The AAF releases persistent values to services connected to the federation and if enabled, to services in eduGAIN.

The AAI platform proposed for a research community is connected to eduGAIN and consumes and transmits AAF persistent identifiers to downstream integrated services.

The proposed AAI platform supports authentication from the following non-AAF identify providers:

- Google
- Microsoft
- GitHub
- ORCID.

These identity providers have several limitations, some of which can be overcome with other processes. For the proposed AAI platform, the following limitations are observable with these identity providers:

- not all support MFA signalling to services outside their ecosystem - for example Google and Microsoft Business accounts do, but personal accounts do not
- do not mint suitable, human readable persistent identifiers
- do not release organisational information, member affiliation or other AAF mandatory attributes
- do not support the same level of identity validation available from trusted AAF subscribing organisations.
    - A research community must establish an MOU with each of the organisations it connects through these identity providers as a source of users. The MOU should ensure that the users' organisation undertakes suitable and adequate person identity validation at onboarding when credentials are issued to their members.
    - This identity validation process should be documented, adhered to and published to the research community to demonstrate commitment to the process.

In support of these non-AAF identity providers, the AAI platform can mint persistent identifiers which are equivalent to and compatible with AAF mandatory attributes where necessary. The non-AAF users and their AAI platform minted persistent values will not have access (or be available) to services outside the research community's ecosystem. If a non-AAF user moves to an AAF subscriber organisation, their community issued persistent researcher identifiers will not be available to the user at the new home organisation. This constraint is similar to an AAF user moving between AAF subscriber organisations or to a non-AAF organisation participating in a research community.

This constraint on the migration of persistent user identifiers reflects the existing constraints on the migration of research projects, their funding and the associated data ownership between organisations. There may be exceptions, but this is not common.

User Identifier portability

If a research community considers user identifier portability critical to support user movement between home organisations, then the user's home organisation's persistent identifiers cannot be reused within the research community's AAI or the connected/integrated services. The research community must mint an internal persistent identifier for a user and map this to the user's AAF issued persistent identifier(s). This approach must be supported by an "identity linking" process, which allows a user to "link" or "connect" an AAF user identity for their new home organisation to an existing community user identity. The proposed AAI platform supports self-service "identity linking" as a default feature that can be enabled on request for all users.

If research communities adopt user identifier portability as a requirement, all user access to community resources (data and services) will not be available through integrations external to the community. To illustrate, if a resource is integrated with both a community AAI and the AAF, a user will have different primary identifiers on the same resource from each of the integrations. If a service wants to link these two identities for the same user, the service must develop and maintain a process to support this identity (account) linking.

Final sign-off

Document approval from both GUARDIANS representatives and project collaborators.

- BioCommons Project Manager:

Signature: _____ Date: _____

- AAF Collaborator Lead:

Signature: John Scullen Date: 12/05/2025

- AAF Senior Solution Analyst*:*

Signature: Patrick Carnuccio Date:13/05/2025

- AAF Project Manager:

Signature: Poornima Mani Date:13/05/2025