

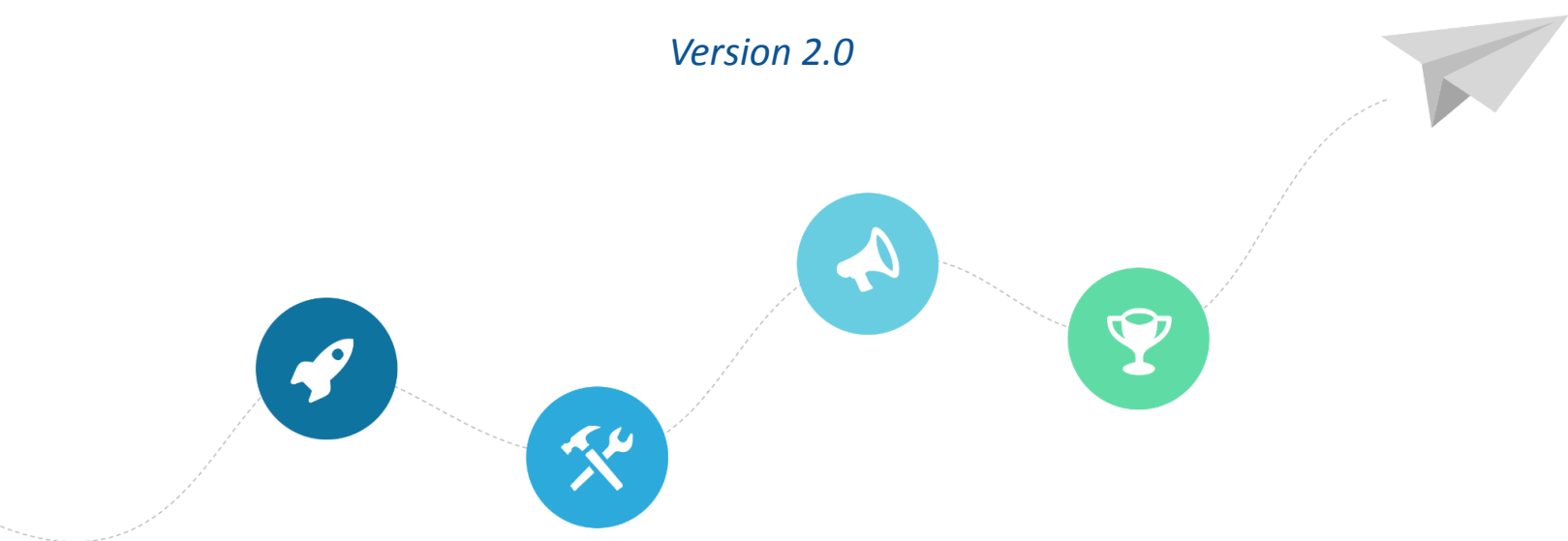
HUMAN GENOMES PLATFORM PROJECT

DAC Automation

DISCOVERY PHASE REPORT

National Community Needs

Version 2.0



DOI: 10.5281/zenodo.6644050

Table of Contents

Glossary	3
Authors	3
1. Introduction	4
2. Current State Findings	6
2.1 Roles	6
2.2 Researcher Data Application Process	8
2.3 System Use Cases	10
2.4 Data Access Requests and the Data Points Collected on DAR Forms	13
2.5 Volume and Frequency of Data Access Requests	18
2.6 Known Development	18
2.7 Stakeholders in the Community That are Not Involved in The Project	18
2.8 Early Findings - Issues with Current DAR and DAC Practices	19
3. User Stories and Survey	20
4. Requirements and Gap Analysis	23
5. Other Standards and Global Projects for Benchmarking	30
6. Conclusion	32
References and Links	35
DAC Automation Sub-Project Team Artefacts	35
Endnotes	35

Glossary

DAC	Data Access Control - a prefix to describe the roles (for example DAC Committee, DAC Coordinator), processes, systems and tools to assess and process Data Access Requests.
DAR	Data Access Request - the suite of forms sent by the Principal Applicant to the DAC Committee.
HGPP	Human Genomes Platform Project

Authors

in alphabetical order by surname

Carnuccio, Patrick

Cowley, Mark

₁Davies, Kylie

Druken, Kelsey

Holliday, Jessica

Kummerfeld, Sarah

Monro, David

Patterson, Andrew

Pearson, John

Pope, Bernie

Scullen, John

Shadbolt, Marion

Wong-Erasmus, Marie

Wood, Scott

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.

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. Figure 1 illustrates the interrelationships between the subcomponents of the Human Genomes Platform Project.

A major challenge to human genome data sharing is navigating restrictions on secondary use. Decisions on how and to whom to grant access to data require significant human effort by DAC Committees. This manual approach is slow and burdensome. The aims of the DAC Automation sub-project are to explore a new data access request and approval paradigm driven by automation for the national human genome research community. Once a researcher/clinician applies to the DAC Committee for access to relevant data from a participating holding organisation, DAC Committees will be able to quickly and easily determine whether access is permitted for the requested purpose. There can be hesitancy from Data Owners and DAC Committees around automation methods. Understandably this can include fears that automation may take away some of the important controls over data use. This needs to be taken into consideration as we progress. However, this new paradigm will improve a DAC Committee's evaluation of data access requests for any data set for a requested purpose. This sub-project will pilot a DAC Automation system with participating repositories and other sub-projects of the HGPP for national human genome community adoption.

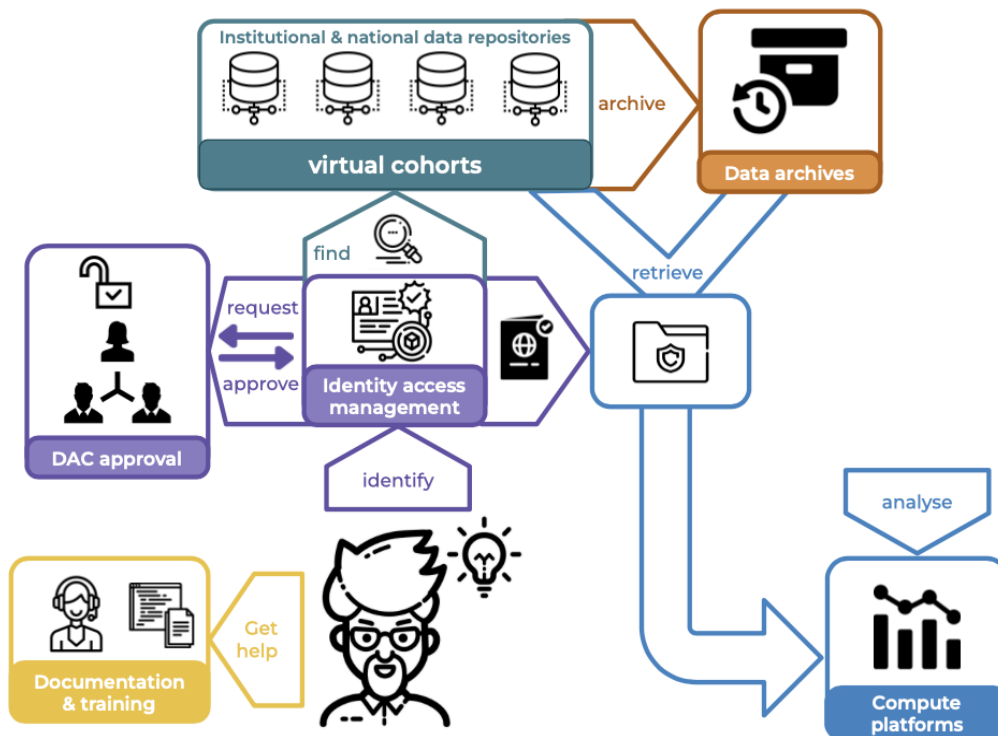


Figure 1 - The HGPP infrastructure ecosystem from the perspective of the research end user showing the key elements of the human genomics data sharing toolbox and data and information flows. (Drafted by Marion Shadbolt, Human genome data specialist, HGPP.¹)

The initial focus of the DAC Automation sub-project team (from here on referred to as “the team”) was a discovery and recording phase to define:

- the current state of data access requests and data sharing agreements within the community
- the set of problems that need to be addressed
- key sub-project areas and their (likely) requirements (Figure 1).

For an Australian genomic data sharing system to be successful, widespread adoption of the new processes and systems is necessary. Therefore, any proposed system must take this into consideration. To ensure the current environment and challenges are well understood, the team used several 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 participating in the project (MCRI/VCGS and CSIRO)
- a survey of human genome researchers to validate user stories recorded by the project team.

The team includes subject matter experts from 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)
- QIMR Berghofer Medical Research Institute (QIMRB)
- National Computational Infrastructure Australia (NCI).

The DAC Automation Discovery Phase Report (this document) records:

- the current state of processes and tools for data access requests and data sharing across the community
- national community needs
- gap analysis
- identification of international projects with potential solution components for piloting in later project stages.

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

2. Current State Findings

2.1 Roles

Role names used in this document to describe researcher data application and DAC approval process actors are described below in Tables 1a and 1b:

Table 1a - Organisational Roles in the Data Access Request and DAC Approval Process

Organisation	Description in the Context of Data Access Control
Holding Organisation ¹	An organisation responsible for holding and sharing data. A holding organisation assesses access requests for the data from a receiving organisation. In many cases the receiving and holding organisations may be the same with in-house researchers requesting access. Third-party organisations may perform the role of the DAC Committee. For ease of process description they are defined discretely within the DAC Automation Discovery Phase Report.
Receiving Organisation ¹	Prospective data recipient, requiring access to a holding organisation's data. The receiving organisation requests data and enters into a data sharing agreement with the holding organisation. Note that in many cases the receiving and holding organisation may be the same entity. Third-party organisations may perform the role of DAC Committee. For ease of process description they are defined discretely within the DAC Automation Discovery Phase Report.
Owning Organisation ¹	There are cases where organisations are the owners of the data, but they are not the holder. For example, the EGA (European Genome-Phenome Archive) holds data on behalf of many owning organisations. In this scenario, the owning organisation tells the holding organisation who is permitted to access the data.

¹ **Why use the terms Holding Organisation, Receiving Organisation and Owning Organisation rather than Holder, Host and Owner?** The intention is to disambiguate the organisational roles from the individual (people based) roles outlined in Table 1b.

Table 1b - Person Roles in the Data Access Request and DAC Approval Process

Role	Description in the Context of Data Access Control
Principal Applicant	Person who is the primary applicant for access to data and who in many cases leads a research project as the Principal Investigator (PI)
Co-applicant	A person invited, by the Principal Applicant, to join a research project, and who may need access to the same data in the data access request.
Data Owner	The representative of an institution that owns the data collection. The Data Owner may report on how often the data is used and what outputs result from its use.
Data Custodian	Person who is responsible for a data collection. Reports to a Data Owner. May report on how often the data is used and what outputs result from its use. May curate the data collection and perform technical tests to ensure its integrity.
DAC Coordinator	Person who receives and evaluates data access requests, sometimes in conjunction with a colleague or as a lead role on a DAC Committee.
DAC Committee Member	Person who reviews data access applications and provides recommendations on granting access as part of a committee led by a Data Access Coordinator (DAC Coordinator).
Authorised Officer	Person who is authorised to sign data sharing agreements with another organisation on behalf of their own organisation (e.g. board member, executive officer or CEO).
Data Distributor	Person who arranges for the provision of the approved data on a successful data access application with a signed data sharing agreement in place. This role may sit within a variety of teams in an organisation and data is provided using various methods.

2.2 Researcher Data Application Process

In the Australian human genome research community, the research data application process is typically structured in one of two ways:

1. Data requests are handled by an in-house Data Access Committee. These organisations contributed valuable DAC role based user stories.
2. An external partner operates the Data Access Committee on behalf of the data owners. Data requests from Principal Applicants in these organisations are typically “pre-approved” for access by an external DAC. In these cases DAC user stories were not collected.

The current researcher data application processes for several project partners were mapped using business process model notation (BPMN). The process steps common to the majority of partners are illustrated in Figure 2. The following notes outline variations across the community:

- Some of the project partners do not operate in-house DAC Committees and are not discussed further.
- In cases where project partners operate in-house DAC, the process starts with a Principal Applicant submitting a data access request (1).
- The DAC Committee reviews the application (2). For some data collections and organisations, a review is conducted by a DAC Coordinator for application completeness, before circulating the application to the DAC Committee (8). In other cases, the reviewer and approver is a single individual as the sole gatekeeper for the data collection (6).
- At any stage in the review process, the DAC Coordinator or the DAC Committee members may request additional information from the Principal Applicant (5).
- The Principal Applicant is informed of the outcome of their application (9).
- Integral to the approval process (7) is the execution of an appropriate data sharing agreement between the holding organisation (granting access to the data) and the receiving organisation (data recipient). If regular data sharing occurs between human genome research organisations, a master data sharing agreement may already exist to streamline data access. Otherwise, a new data sharing agreement may be created for each data request.
- Once the approval is granted, Data Distributors provision access for the Principal Applicant and their team (11). The data access request may include details of the data required and the Data Distributor uses these details to prepare the data for distribution. In several cases Data Distributors inform the Principal Applicant of the access details for the requested dataset.
- The Principal Applicant (or their delegate/team member) receives the details for access to the data (12).

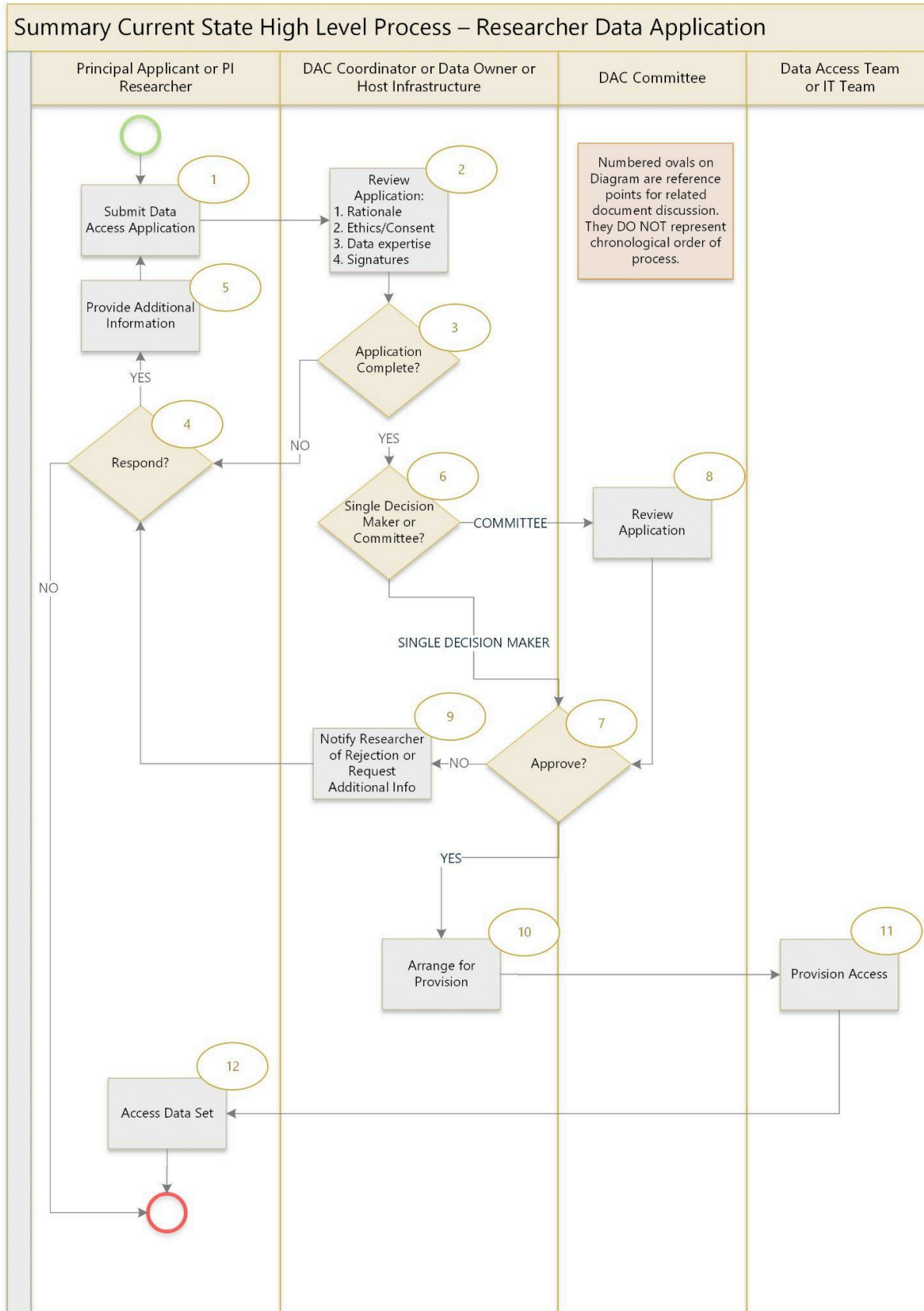


Figure 2 - Researcher Data Application Process Current State Summary.
DAC Automation Discovery Phase Report - Human Genomes Platform Project

2.3 System Use Cases

System use cases for roles participating in the data access application process are outlined in Figures 3 and 4. These diagrams illustrate the interaction between roles and the existing systems for data access application and approval. Many of these interactions are manual. Some organisations omit some of the steps described in the diagrams. Candidate solutions for the DAC Automation sub-project must include features to automate most of these interactions. The known exceptions, illustrated in the diagrams for completeness but out-of-scope for automation, are:

- **Sign Data Sharing Agreement** - executives at the data holding organisation (granting data) and the receiving organisation (data recipient) complete signing and countersigning of data sharing agreements. During requirements review, the sub-project team identified that it is highly unlikely that these executives will access a DAC Automation System. These signing executives expect a suite of supporting information is available and to physically sign a document. If automated systems such as REMS do not support a third party digital signature solution directly, an external digital signature service may be an option for some organisations.
- **Provision Data Access** - various data provisioning methods are employed. The integration of the data delivery methods is not in scope for DAC Automation. However, the provision of data access will have a dependency on information in the proposed DAC Automation solution, so this system use case is shown in the diagrams.
- **Remove Data Access** - the removal of data access is closely related to the method that provisions access. The removal of data access is also out of scope for the DAC Automation sub-project. However, the provision of data access will have a dependency on information in the proposed DAC Automation solution, so this system use case is shown in the diagrams.
- **Destroy All Copies of Data** - this is a manual process for the researchers who have the data on their own systems or storage/compute facilities or control of copies of received datasets.

System Use Cases | Submit Data Access Application and Approve, Reject, Sign, Provision

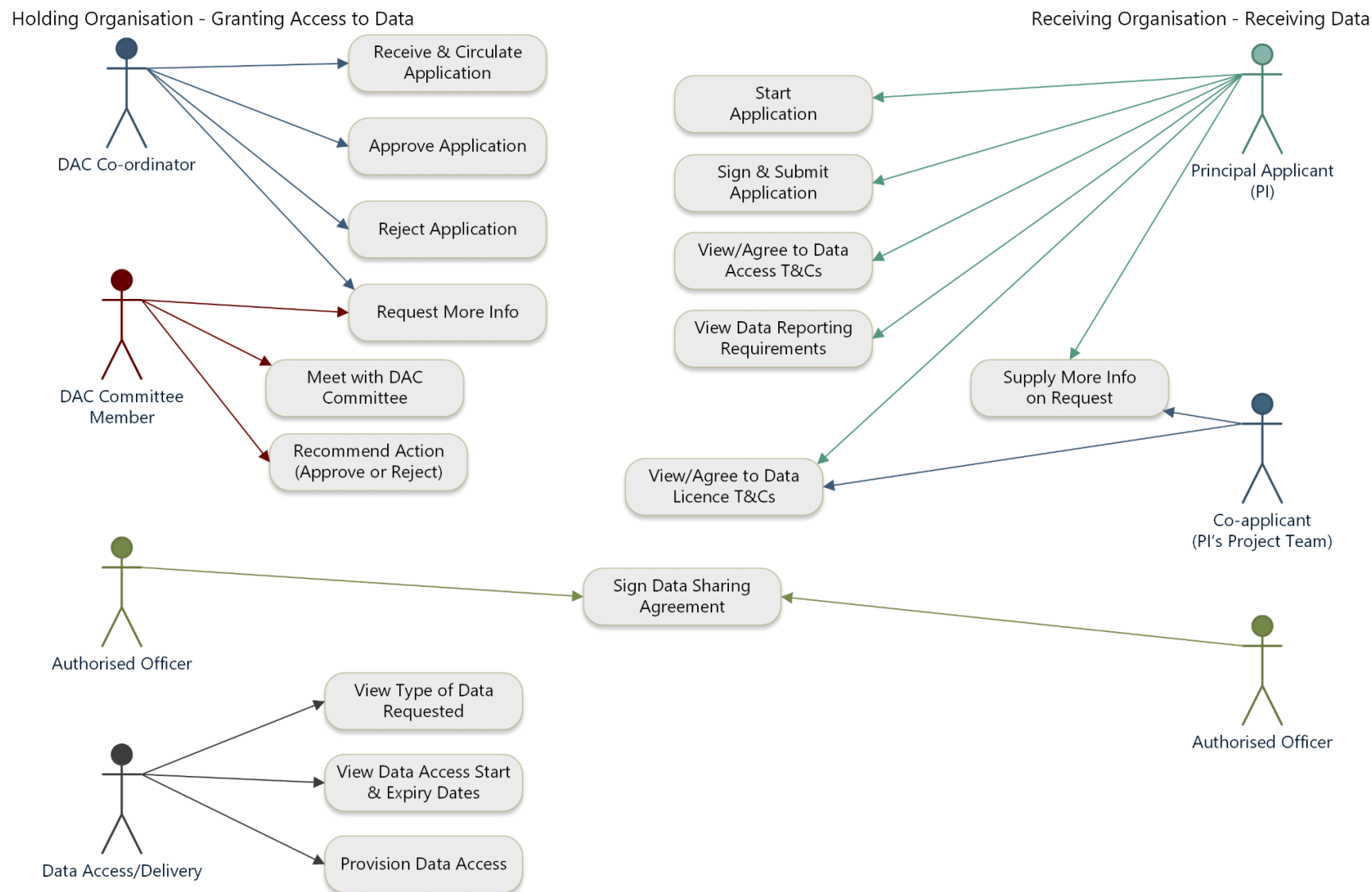


Figure 3 - System use cases for data application, review and approval processes

System Use Cases | Manage Project Lifecycle, Report, Review, Exit

Holding Organisation - Granting Access to Data

Hosting Organisation - Receiving Data

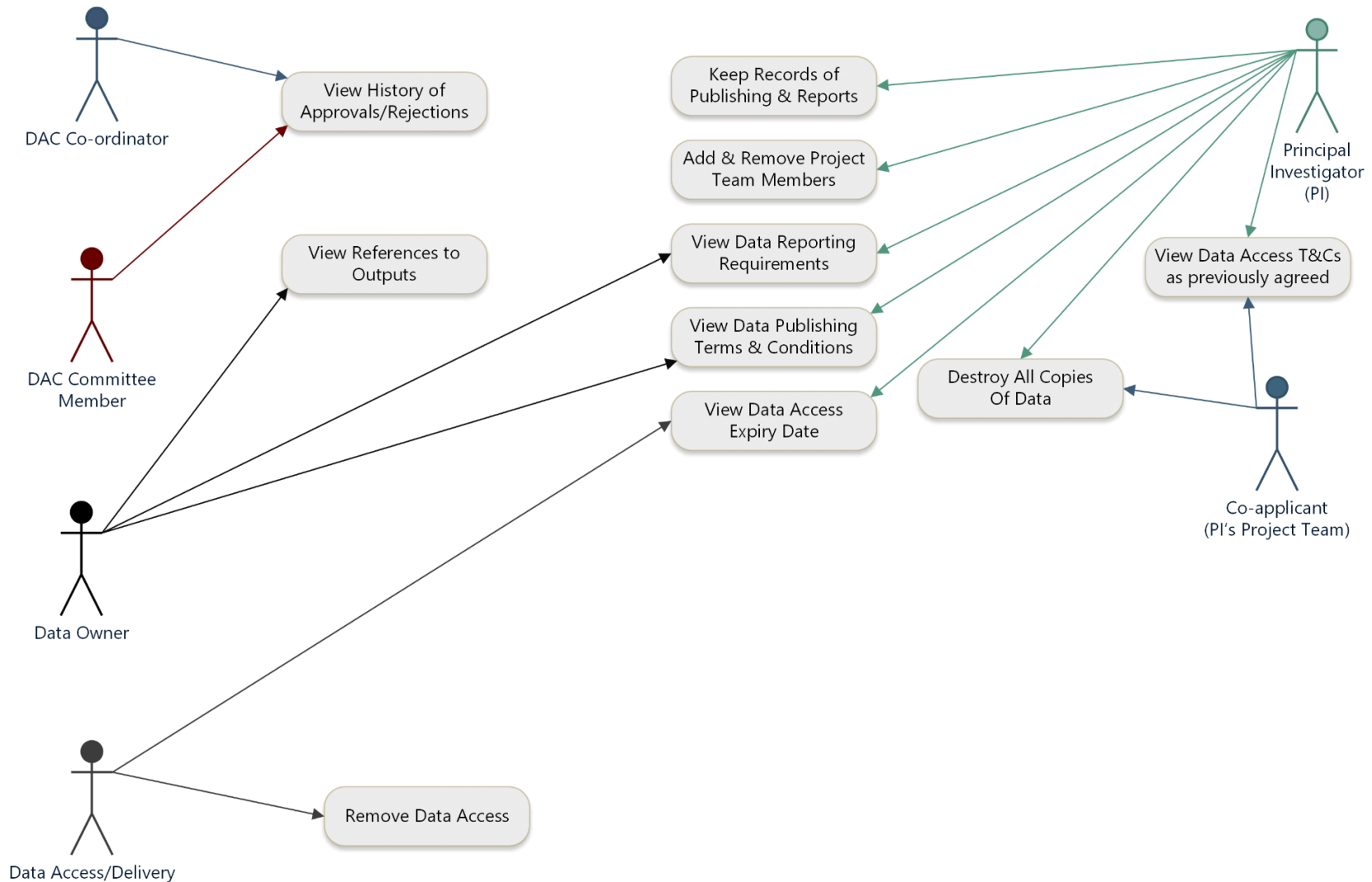


Figure 4 - Once the approval is complete and the project is underway, these system use cases show actions for managing the project life cycle to its conclusion

2.4 Data Access Requests and the Data Points Collected on DAR Forms

Data Access Request (DAR) forms are used to gather information and submit data access requests. The project partners' DAR forms were examined and found to contain similar data entities (groups of data points about one element). Figure 5 shows the data entities on DAR forms and the phases in which they are relevant throughout the research project lifecycle. Colour coding indicates who is responsible for providing the information for each data entity.

Table 5 lists the data points currently collected by project partners' DAR forms, grouped by process/interaction (see **Appendix: References and Links** for links to individual organisation DAR Forms).

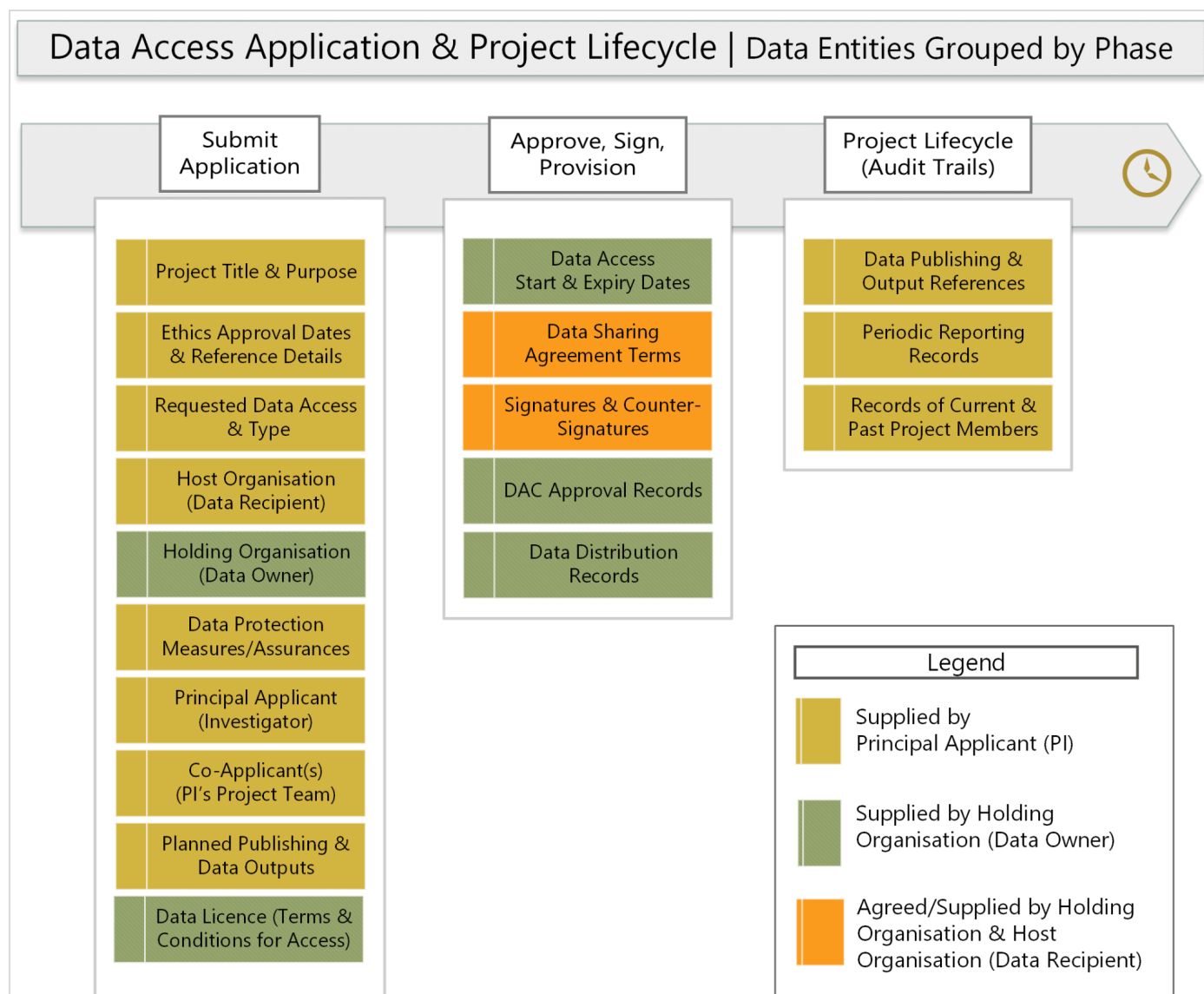


Figure 5 - Data Entities and the timing of their use through the research project lifecycle

Table 5 - Data Points Collected on DAR Forms by Medical Research Institutes

Datapoint	Data Type / Detail	Institutes Collecting DataPoint on DAR Form			
		Garvan (for MGRB)	UMCCR (per AGHA)	Zero CCI	QIMR Berghofer ²
Project Purpose					
Project Title	Text	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Ethics Approval Letter Attached	Yes/No checkbox	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Name of Human Research Committee (HREC) that approved	Text	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Date/Expiry Date of HREC Approval	Date/Time	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
HREC Number	Text / (Human Research Ethics Committee Number)	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Project Type (please select all that apply)	Checkbox / Selections are: 1. Research re-analyses 2. Method 3. Clinical re-analyses 4. Cohort research 5. Non-commercial 6. Commercial 7. Methods development	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Proposed Purpose Outline (including background, aim, analysis methods) ³	Paragraph	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Details of Requested Data Access					
Data file types (select all that apply)	Checkbox / Selections are unique to each org so not listed (examples include VCFs, FASTQ, BAMs, Germline WGS FASTQ)	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Flagships of interest	Checkbox / various	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Select Data Details/Where to Access	Checkbox / various	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Level of Data required ⁴	Checkbox / Selections are: 1. Controlled	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>

² QIMR Berghofer uses 3 different forms depending on researcher status, plus a data sharing agreement. Datapoints indicated are the combined total of all forms.

³ Length restrictions: UMCCR/AGHA max 1 page, Garvan MGRB max 500 words, Zero CCI max 500 words,

⁴ Excluded as Garvan/MGRB offers an option for Open Access but does not require submission of a data access request.

Datapoint	Data Type / Detail	Institutes Collecting DataPoint on DAR Form			
		Garvan (for MGRB)	UMCCR (per AGHA)	Zero CCI	QIMR Berghofer ²
	2. Restricted				
Host Institution (Data Recipient)					
Institution	Text	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Address	Text	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Phone	Text	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Data Protection Measures/Assurances/Justifications					
Provide detail as to how data will be stored by host institution and what security measures are in place to ensure conformity with the conditions stipulated in data sharing agreement(s) "section number..."	Paragraph	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Should FASTQ data be requested – provide appropriate justification as to why this format (unprocessed rather than processed) is necessary for the purpose of the proposed project.	Paragraph	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Principal Applicant/Investigator Details (ie the principal authorised user requiring access to data for this project under data sharing agreement)					
Title	Text / Honorific	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Full Name	Text	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Address	Text	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Role at Institution	Text	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Phone	Text	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Email address	Text	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

Datapoint	Data Type / Detail	Institutes Collecting DataPoint on DAR Form			
		Garvan (for MGRB)	UMCCR (per AGHA)	Zero CCI	QIMR Berghofer
Co-Applicants/Authorised Personnel⁵ (additional authorised users requiring access to the requested data for this project under the Data Access and Sharing Agreement)					
Full Name (incl honorific)	Text	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Department/Division	Text	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Email/Phone	Text	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
Role/Job Title at Institution	Text	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
NCI Australia Username (for Garvan/MGRB this is for those who will/have access at NCI)	Text	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Collaborative or Multi-institute Projects⁶ (optional)					
Corresponding data applications submitted for this project	Text	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Has this project been previously approved via another application form? ⁷	Text / details of approval	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Expected Outputs Arising from this Data Access⁸					
Journal Publications	Yes/No	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Conference Proceedings	Yes/No	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Other outputs (please specify)	Text	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

⁵ On UMCCR/AGHA form: Room for 5 co-applicants, On Garvan form: room for 4 co-applicants, On ZERO form - expandable. For QIMR Berghofer it *seems* like a separate agreement is signed by each person wanting access.

⁶ On UMCCR/AGHA form: For multi-institute projects, each institute must submit a separate DAR form. However this section notifies DAC Committee of corresponding applications for the same project being run across multiple institutes.

⁷ On UMCCR/AGHA form: "new data access requests for existing approved projects will normally be exempt from full Data Access Committee review."

⁸ On UMCCR/AGHA form: Condition specified on DAR form (no data point collected for this) "If results of potential clinical significance to the proband are found, this information must be returned to Australian Genomics. Australian Genomics must be notified of such results prior to publication of these findings." On Garvan form and Zero CCI forms publication acknowledgement policies are noted/referenced - the need to acknowledge is specified.

Datapoint	Data Type / Detail	Institutes Collecting DataPoint on DAR Form			
		Garvan (for MGRB)	UMCCR (per AGHA)	Zero CCI	QIMR Berghofer
<i>Data Access and Sharing Agreement</i>					
Is there a Data Access and Sharing Agreement in place between the Institution and the Administering Institution to govern this data access and sharing request if approved?	Yes/No	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
If yes to above insert the date of Data sharing agreement	Date	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Signature and Co-signature Panels ⁹	Signature of applicant Institution's duly authorised officer and Co-signature of Data-granting institution's duly authorised officer	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

⁹ ZERO CCI and QIMR Berghofer have a data sharing agreement as a separate document that includes signature panels. Garvan/MGRB and AGHA/UMCCR include signature panels in the DAR.

2.5 Volume and Frequency of Data Access Requests

The annual number of data access requests varies greatly across the national community:

- The file type and size can be a factor, with frequent requests for small volume datasets of a single genome/exome.
- Conference presentations or the publication of a paper will raise awareness of featured datasets and result in an increase in requests.
- Where there is an existing overarching data sharing agreement between organisations, data requests are more frequent because the trust frameworks and sharing systems between organisations have been established.

Interviews with HGPP partners and other Australia-based human genome research organisations revealed:

- Requests for small individual datasets, for example one exome/genome total of 20GB of data uploaded to the cloud or to HPC environments and shared by secure URLs can be fairly common, as much as several each week.
- Larger collaborations that encompass many study participants can result in transfers of 50TB of data. The requests are less frequent and pose an infrastructure challenge to achieve reliably.
- Knowledge of the existence of data is an issue. Increased value could be derived from data collections by raising researchers' awareness of the available datasets in combination with a process to securely apply for access that was efficient, transparent, timely, clear and consistent.

2.6 Known Development

- UMCCR and ZERO CCI stood up Gen3 instances in 2021 as a pilot project to demonstrate Gen3 integrated with CILogon to share selected data.
- Garvan's Medical Genome Reference Bank (MGRB), a valuable and sought after collection of healthy (disease depleted) elderly genomes, is hosted and managed by NCI within their hosted reference data collections.

2.7 Stakeholders in the Community That are Not Involved in The Project

The team interviewed representatives from MCRI and sought information from CSIRO. These organisations operate within the national human genome research community but are not participating in HGPP. These stakeholders face similar challenges as those identified by the project. MCRI is involved in a relevant project working with the GenoVic genomic data system managed by Melbourne Genomics. MCRI currently loads all of their human genome data into GenoVic. Though the data sharing capability is not yet available, MCRI expects to accomplish a significant amount of data sharing through this cloud platform in the future.

2.8 Early Findings - Issues with Current DAR and DAC Practices

Table 6 - DAC Automation - Current State Issues Identification

#	Issues Description	DAC Automation Solution Mitigation Strategies	In scope for DAC Automation?
1	Elapsed time from data access request to approval is often longer than necessary.	Workflow and tracking in the solution could reduce time from data access request to DAC approval.	Yes
2	The DAC Process can be opaque.	An automated process should increase transparency for all involved.	Yes
3	People requesting data may need to repeatedly enter their credentials and details for multiple requests.	An automated system can save these details and avoid re-entry.	Yes
4	Holding organisations and data owners need better visibility of researchers who have access to datasets or have received copies of data. Information about the value researchers may have gained from data (publications, discoveries, etc) could improve.	A persistent identifier can link a project to the data access granted. These identifiers could incorporate researcher IDs, digital object identifiers and project IDs (examples ORCID, ROR, RAID and DOI).	Yes
5	Legal negotiations “can be the death of data sharing agreements”. If we could standardise and automate more and find commonality through trust frameworks and a shared system, we could avoid repeating unnecessary legal work.	DAC Automation Policies agreed to by infrastructure members could lower legal hurdles. Including these policies in the DAC Automation system enables their presentation to users and the collection of user agreement. This ensures data requests meet a minimum baseline. If this reduces some of the legal negotiation workload then time and money can be saved.	Yes
6	Excessive effort is necessary from the DAC Coordinator to follow up requests using multiple communication pathways (emails etc) rather than the convenience of a shared system.	Solution could include automated workflows that store all relevant components and communications in one place.	Yes
7	Excessive effort for DAC Committee members to review and respond. Committee members cannot easily review historic approvals.	Solution could include an audit trail of approvals. This will streamline new requests while records of historic approvals can assist new committee members see the projects their predecessors approved/rejected.	Yes
8	Excessive effort for Data Distributors that lack a dashboard view of the researcher details, data type and access granted, and the effective dates.	DAC Automation could provide clear information to Data Distributors on identifiers, access requests, access duration, researcher experience and training, and a project’s ethics approval.	Yes

3. User Stories and Survey

Representatives of project partner organisations, with relevant experience, contributed role-based [User Stories](#) to describe their own data access request and DAC approval process needs. The team surveyed the wider human genome researcher community asking them to evaluate these user stories. Comments from the survey respondents provided additional information:

Researcher User Story Survey Feedback	
User Story	Community Feedback Received
As a Human Genomics researcher, I need to be able to find and review datasets electronically, so as to select the most appropriate data to interrogate as part of my project.	<i>The end of this sentence makes it sound like researchers can select any data set that is appropriate for their research. There is no reference to the role of caveats placed on secondary data use by data custodians (i.e. DTA) and participant consent clauses.</i>
	<i>The two most important but quite different criteria for search are: 1. search by scientific utility (phenotype, cohort, etc.) 2. search by data usage conditions (commercial use ok? collaboration required? restricted access time? restricted redistribution conditions?)</i>
As a Human Genomics researcher, I need to easily read and understand the terms and conditions stipulated by a Data Access Committee, so as to abide by the requirements of the people who provide the data.	<i>I'd prefer machine readable consent forms that can be automatically evaluated against the provided research proposal.</i>
	<i>Keep it simple.</i>
As a Human Genomics researcher, I need to be able to easily apply for access to a data cohort, so as to free up my time to begin valuable research.	<i>A single approval process for all deidentified Australian data would be of great value.</i>
	<i>While I agree with the sentiment that less of researchers' time should be spent on paperwork, the wording of this statement is elitist. It comes across that it should be easier because researchers' time is more important than all the reasons we do data governance. There are so many other ways to end this sentence, for example, "As a Human Genomics researcher, I need to be able to easily apply for access to a data cohort, to ensure consistency in processes and reduce the burden of learning new systems with each request."</i>

Principal Investigator User Story Survey Feedback

User Story	Community Feedback Received
As a Principal investigator, I need to be able to add researchers in my group to a single data access application, and add and remove team members through the life of the project so as to easily establish seamless access for all members of my team.	<p><i>This is critical and ideally would not require amendments as the team changes, or at most a simple notification.¹⁰</i></p> <p><i>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'.</i></p>
As a Principal investigator, I need to know the proposed use of each dataset (i.e. the project summary) so as to only use the shared data in accordance with the terms of the data sharing agreement.	<p><i>This is a fine point as there needs to be some ability to explore the data and try new approaches.</i></p> <p><i>Wouldn't the PI or their staff be proposing the use of each dataset? Are you suggesting the PI's staff are putting in requests without PI's knowledge, and they need to be able to see details or that data custodians provide information on the datasets? If the latter, remove "proposed use of each dataset" and change to "scope of secondary research allowed for the dataset under the data access approval."</i></p>

DAC Coordinator User Story Survey Feedback

User Story	Community Feedback Received
As a DAC Coordinator, I need to see all the details together to be able to review a data access request, and easily determine the current status of data requests at any time, so that I can efficiently process requests and don't waste time going back to the researcher for further information.	<i>Strongly agree - This is a key role that will "make or break" the system.</i>

DAC Committee Member User Story Survey Feedback

User Story	Community Feedback Received
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.	<i>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.</i>

¹⁰ Early indications from our team is that if project membership is represented in CILogon, this will be easy to implement.

Institutional Signing Officer User Story Survey Feedback	
User Story	Community Feedback Received
As an Institutional signing officer, I need to have access to all information on data access applications and data/material transfer agreements, so as to make only the appropriate commitments on behalf of my institution.	<i>Strongly agree - Don't need to be overwhelmed by detail, just the necessary information.</i>

4. Requirements and Gap Analysis

Human genome data is often sensitive, which elevates the need to qualify data access requests carefully. However, the team recognises current processes take longer than necessary to progress from the request stage through to approval. Team members contributed DAC Automation requirements based on the user stories, the current state, and technical, legislative and operational considerations. The team then reviewed and prioritised requirements. The team 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 7 to 15).

The mapping exercise produced a gap analysis, illustrated in Figure 6. Actions to address each gap will be undertaken to build the pilot system.

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

- User interface provides metadata (about dataset) and an electronic form to apply for access - [Table 7](#).
- DAC automation system supports workgroup managers to manage their team and their access to resources - [Table 8](#).
- DAC automation system provides all of the required verified information to enable DAC committee members to assess Data Access Requests - [Table 9](#).
- Users can use the system to easily access information that supports their obligation to comply with the terms and conditions and reporting requirements associated with the data access granted to them - [Table 10](#).
- The DAC Automation System leverages persistent identifiers (for example ORCID ID, ROR, and RAID) to link entities and to surface data to enhance decision making - [Table 11](#).
- Data distributors use the system to track the approvals granted and see the type of access granted. This assists data distributors to distribute only the datasets and types approved and to follow up where applicable egress billing or other charges arising (this has a dependency on data sharing agreements) - [Table 12](#).
- The DAC Coordinator uses the DAC Automation system to manage and circulate all Data Access Requests with automation and workflow built into the DAC Automation system- [Table 13](#).
- Sufficiently granular data access approval capability is available to respect individual consent and restriction of data types. Future dynamic consent capability is not blocked/precluded by the components selected for the Minimum Viable Product - [Table 14](#).
- The system has the necessary flows and notifications to keep users informed of the progress of Data Access Requests - [Table 15](#).

Gap Analysis



- User Interface provides metadata and electronic forms to apply for access.
- DAC Automation System supports PIs in managing their own workgroups.
- Users can check the agreement terms quickly during the project to support their compliance to T&Cs.
- DAC Automation System provides all the information to enable DAC Committee to verify the applicant and assess data access requests.
- DAC Coordinator uses DAC Automation System to circulate and workflow data access requests.
- Users can attach files of certain types where needed.
- Data Distributor uses DAC Automation System to track approvals granted & see the type of access and the file types required.
- Users view completed requests (history) in the system as a reference of what was agreed.
- Data access approval is granular enough to allow for individual consent and data file type restrictions.



THE GAP

DESIRED STATE

- Awareness of data and access to metadata is limited to those who are aware of data's existence & User Experience varies.
- PIs can find managing access for their own workgroups onerous and there is no visibility of who is a current workgroup member.
- It is hard to look back at what you agreed to which makes it hard to comply with the T&Cs.
- DAC Committee Members have to loop back for more information & look in different places.
- DAC Coordinators have an onerous task with many follow-ups and cannot check the status.
- Data Distributors are not supported because applicants don't understand their ethical obligations.
- Hard for data owners to show what value is being derived from data.
- Hard to track requests.
- Individual consent support is not automated.

CURRENT STATE



ACTION STEPS TO BE DEvised BY DAC AUTOMATION SUBPROJECT TEAM

Figure 6 - Summary of DAC Automation Gap Analysis

Table 7 - USER INTERFACE PROVIDES METADATA AND ELECTRONIC FORM TO APPLY FOR ACCESS.

HGPP REQ NUMBER	Related Reviewed & Scoped Requirement	Mandatory or Should Have	Current State
HGPPREQ-001	Dataset catalogue/summary details (metadata) should be viewable by all authorised users of the platform before commencing a data access application.	M	<p>Limited publicity of datasets constrains community awareness of data/metadata and access pathways.</p> <ul style="list-style-type: none"> Conference presentations raise awareness of the existence of a dataset, which then drives a spike in requests for that dataset. Other constraints include the lack of electronically viewable metadata from a central source. Data access requests are sent via email as PDFs (excluding requests for some datasets at Garvan and NCI). Terms and conditions are also PDF or paper based.
HGPPREQ-002	Researcher/PI can use the platform to apply for access to data.	M	
HGPPREQ-003	The platform presents the terms and conditions for access to data.	M	

Table 8 - DAC AUTOMATION SYSTEM SUPPORTS WORKGROUP MANAGERS TO MANAGE THEIR WORKGROUPS AND THEIR ACCESS TO RESOURCES.

HGPP REQ NUMBER	Related Reviewed & Scoped Requirement	Mandatory or Should Have	Current State
HGPPREQ-010	Principal investigator can use the platform to add researchers in their group to a single data access application, and add and remove team members through the life of the project so as to easily establish seamless access for all members of the project. This has a dependency on the data distribution model.	SH	<p>Principal Investigators can find workgroup management onerous since there is limited visibility of workgroup membership.</p> <ul style="list-style-type: none"> When data is not made available for download via an encrypted URL or on physical media, Principal Investigators often rely on technical teams to manage their team member's access to both data and services/resources. Visibility of who may have access to data is not always readily available. As projects often span multiple years, Principal Investigators may struggle to track their team members' access lifecycle over the life of a project. This opacity will affect any risk mitigation strategies for a project or the service provider.
HGPPREQ-011	The platform should allow appropriately authorised users to find out what datasets their research team have requested and been granted access to, which people have been granted access, and to know the duration of the agreement so as to effectively manage the holdings and mitigate any risks from holding sensitive genomic datasets.	M	

Table 9 - DAC AUTOMATION SYSTEM PROVIDES ALL OF THE REQUIRED VERIFIED INFORMATION TO ENABLE DAC COMMITTEE MEMBERS TO ASSESS DATA ACCESS REQUESTS.

HGPP REQ NUMBER	Related Reviewed & Scoped Requirement	Mandatory or Should Have	Current State
HGPPREQ-015	DAC committee member has access to all components of data access applications together, electronically, so as to efficiently, quickly and confidently complete their review.	M	<p>DAC Committee members may have scant knowledge of an applicant and may rely on reference information from different sources or request more information from the applicant during their assessment of a DAR.</p> <p>Apart from isolated cases, common scenarios include:</p> <ul style="list-style-type: none"> • Distribution of DARs to DAC Committees is via email or on paper. • DAC Committees request additional info from an applicant via email. • No central record of messages and no workflow. • DAC Committees may cross reference info from various sources in order to gather sufficient information to reach a decision on a DAR. • DAC Committees cannot easily search historical DAR decisions for details on applications and DAC membership. A search feature would aid the work of new DAC Committee members. • DAC Committees signal their decision via email or physical forms.
HGPPREQ-016	DAC Committee member has automated workflow for data access request, so as to be able to easily add remarks for the DAC Coordinator, approve, reject or request more information from the applicant.	M	
HGPPREQ-017	DAC Committee member has access to an electronic archive of their historic approvals dating back 5 years after research project ends , so as to cross reference new applications with previous approvals made. NEED TO CLARIFY WHETHER DAC COORDINATOR SHOULD SEE ALL HISTORY AND DAC COMMITTEE MEMBER JUST SEES THEIR OWN.	SH	
HGPPREQ-018	DAC Committee member can attach documents to an applicant's application so that all signatories have access to external supplementary information if needed.	M	
HGPPREQ-019	DAC Committee member can electronically indicate the approval or rejection of an application, so as to eliminate the need for printing, signing and scanning. The platform credential is reliable enough to support this feature.	M	
HGPPREQ-021	DAC Committee member can know the verified identity and institution/workgroup details of the applicant so as to avoid the need for double checking with institutions and to be assured that they are only granting access to appropriate applicants.	M	
HGPPREQ-051	The DAC Committee members should be well informed enough to be able to complete the cohort sign off respecting the consent/rule preferences of the underlying dataset/cohort (of which I may have not been a part of formulating i.e the data committee may not be exactly the same people who collected the data) (for example DUO CODEs should be visible to the DAC Committee so they are informed even if new).	M	

Table 10 - USERS CAN USE THE SYSTEM TO EASILY ACCESS INFORMATION THAT SUPPORTS THEIR OBLIGATION TO COMPLY WITH THE TERMS AND CONDITIONS AND REPORTING REQUIREMENTS ASSOCIATED WITH THE DATA ACCESS GRANTED TO THEM.

HGPP REQ NUMBER	Related Reviewed & Scoped Requirement	Mandatory or Should Have	Current State
HGPPREQ-012	After approval is granted, the platform should allow a user to cross reference their data access application with their interface to the data, to know their proposed use of each dataset so as to only use the shared data in accordance with their proposed use original application and the terms of the data sharing agreement. (We are imagining an icon on the screen where the data is, so we can click and see our application and know what we promised).	SH	<p>For projects that received approval in the past, it may be hard for researchers to confirm the scope of the DAC approval.</p> <ul style="list-style-type: none"> DAC approval information may be filed in different places. It is difficult for Principal Investigators and their research associates to retrieve information on their proposal to the DAC Committee – for example: purpose, terms and conditions, reporting obligations and time limits.
HGPPREQ-039	Executed data sharing agreements/contracts must be stored in the system and linked to the original data access request.	M	

Table 11 - ID USED IN PLATFORM IS LINKED TO IDs THAT IDENTIFY PUBLISHING RESEARCHERS (ORCID iD and ROR) and A PLACEHOLDER FOR POSSIBLE FUTURE RAID ID.

HGPP REQ NUMBER	Related Reviewed & Scoped Requirement	Mandatory or Should Have	Current State
HGPPREQ-020	The platform should have the facility to capture an open researcher ID and project/activity ID (e.g. ORCID and ROR) and a placeholder for future RAID.	M	<p>Data owners lack visibility on the value derived from shared datasets for reporting purposes.</p> <p>Usage information, when available, would help prioritise those hosting methods for data sharing and enable Data Owners/DAC Coordinators/DAC Committees to report on publications and discoveries arising from all data use granted.</p>

Table 12 - DATA DISTRIBUTORS USE THE SYSTEM TO TRACK THE APPROVALS GRANTED AND SEE THE TYPE OF ACCESS GRANTED, IN ORDER TO DISTRIBUTE ONLY THE DATASETS AND TYPES APPROVED AND TO FOLLOW UP ANY EGRESS BILLING OR OTHER CHARGES (dependency on data sharing agreements).

HGPP REQ NUMBER	Related Reviewed & Scoped Requirement	Mandatory or Should Have	Current State
HGPPREQ-037	Data distributors need a way to track the transfer of data (for example the number of complete downloads so as to enable billing of consumers for data egress if they download more than once.)	M	<p>Though data distribution methods vary between institutions, data distributors rely on their internal systems, to track the approvals granted and the access granted, to only distribute the datasets and types approved and to follow up any egress billing or other charges – charges are dependent on data sharing agreements.</p>

Table 13 - DAC COORDINATOR USES THE DAC AUTOMATION SYSTEM TO MANAGE AND CIRCULATE ALL DATA ACCESS REQUESTS WITH AUTOMATION AND WORKFLOW BUILT IN.

HGPP REQ NUMBER	Related Reviewed & Scoped Requirement	Mandatory or Should Have	Current State
HGPPREQ-022	DAC Coordinator can see all the details together to be able to review a data access request, and easily determine the current status of data requests at any time so as to efficiently process requests and not waste time going back to the researcher for further information.	M	<p>DAC Coordination is a complex task that relies on manual follow-up with DAR Applicants. DAC Coordinators:</p> <ul style="list-style-type: none"> typically receive applications via email may need to request more information from Applicants, often via email multiple time must distribute DARs to DAC Committee members via email and separately record messages without the support of a central message store or a workflow process may need to cross reference info from different sources to gather information for a DAR decision cannot easily search historic approvals because there is no audit trail of DARs and the decisions. This impacts new DAC Coordinators in their role report approval/rejection to Applicants via email or with paper forms.
HGPPREQ-023	DAC Coordinator should have a workflow and communication tool, so that they can easily communicate with the applicant, the DAC and data access team/data distributor: fielding questions, advising outcomes and coordinating data teams to provision access.	M	
HGPPREQ-025	DAC Coordinator can use the platform to track the progress of a data access application from receipt to transfer so as to efficiently monitor the progress of an application and finalise applications once datasets have been transferred.	M	
HGPPREQ-026	DAC Coordinator can record and track executed agreements so as to efficiently report on what datasets have been shared and to whom.	M	
HGPPREQ-027	DAC Coordinator can view and understand the ethical obligations linked to the data so as to ensure applicants have the necessary ethics to work with data that they have applied for. THIS MAY HAVE A DEPENDENCY ON HGPPREQ-034.	M	
HGPPREQ-030	DAC Coordinator can support the review and update process of data access request (DAR) forms, to ensure that the institute holding the data complies with current regulations and to ensure applicants have the right version of the form. NEEDS CLARIFICATION as to editing privileges (we are imagining an audit trail for any changes).	SH	
HGPPREQ-032	DAC Coordinator can record and follow-up on any periodical/publication/other reporting requirements via the platform. (Ability to track whether an annual report is due and has been done for data collection). Report attached to DAC Sharing Agreement/Application would be nice to have.	M	

Table 14 - SUFFICIENTLY GRANULAR DATA ACCESS APPROVAL CAPABILITY IN ORDER TO RESPECT INDIVIDUAL CONSENT AND RESTRICTION OF DATA TYPES. FUTURE DYNAMIC CONSENT IS NOT BLOCKED/PRECLUDED BY MVP.

HGPP REQ NUMBER	Related Reviewed & Scoped Requirement	Mandatory or Should Have	Current State
HGPPREQ-047	The DAC Committee/Coordinator should be able to restrict the permissions of an applicant to a subset of the complete cohort population. (for example DUO Codes to be applied at the record level not dataset level).	SH	Automation of individual consent is not implemented in Australia. <ul style="list-style-type: none"> Benchmarking against international exemplars also shows no operational support internationally. For example, a study participant may approve cancer research on their specimen, but not other diseases; or approve for not-for-profit research, but not approve for commercial research; or the study participant may withdraw partial or all consent. The study participant consent process is currently a manual process. Dynamic consent is a recent paradigm change and is not a feature of automation processes. Australian genomics has two projects exploring the implementation of dynamic consent.
HGPPREQ-048	The DAC Committee/Coordinator should be able to restrict access of an application to a subset of data types in the cohort (BAMs only, VCFs only, not fastqs etc).	SH	
HGPPREQ-049	The platform must be able to restrict access of an applicant to genomic regions that match the genomic regions of interest to the applicant (as opposed to the whole genome). This is the "avoid incidental findings" requirement.	SH	
HGPPREQ-050	The DAC Committee/Coordinator should be able to enable access to the data cohort respecting the individual consent preferences of all the participants in the cohort.	SH	

Table 15 - PLATFORM HAS NECESSARY FLOWS & NOTIFICATIONS TO KEEP USERS INFORMED OF PROGRESS OF DATA ACCESS REQUESTS.

HGPP REQ NUMBER	Related Reviewed & Scoped Requirement	Mandatory or Should Have	Current State
HGPPREQ-042	The platform should allow for automated notifications to be sent to users when request status changes (such as progress, approval, rejection and request for more information, expiry).	M	Managing DARs manually and communicating via email supplies little or no information on a DAR's current status nor the active role for any stage of a manual process.

5. Other Standards and Global Projects for Benchmarking

European and the United States genomic research communities are well advanced in their multi-year programs to improve secure sharing of human genome data. Table 16 contains examples of these efforts. The team reviewed the activities undertaken by these communities and organisations:

- to ensure compatibility with these communities by adopting international standards for the work of this sub-project.
- to gain familiarity and leverage existing work, to achieve maturity faster in the sub-project's solutions.

As part of the candidate solution analysis phase for the DAC Automation Sub-Project, the team will assess the suitability to the Australian human genome research community of selected work packages and products from the other projects. The selected packages, their source and the reason for canvassing them are summarised in Table 16.

Table 16 - Global Projects and Products with Candidate Solution Potential

Work Package/Product	Source	Why Canvass This Item?	Who, What and How?
Passport/Visa 2.0	GA4GH DURI Work Stream	Open-Source, widely used technology (JWT), holds details describing researcher attributes/qualifications in the form of a bearer token that is consumed by a technology solution (a Broker) 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 ¹¹ documentation is plentiful. The team reviewed this specification when drafting our sub-project requirements. Waiting on finalisation of Passport specification 2.0.
CILogon	CILogon	An as-a-Service implementation of the best practice AARC Blueprint Architecture pattern for research collaborations. Allows researchers simplicity in login but also supports workgroup management. Used by NIH. Proven locally with ZERO and UMCCR in their 2021 pilot.	CILogon has demonstrated their product. The team has set up a test instance of CILogon and will be testing through scenarios and personas.
REMS	GA4GH and CSC – IT Center for Science, Finland	Open source, widely used, integrates with Beacon, enhancements can be suggested to the project.	REMS is operational at Garvan (production use for MGRB) and by our team on AWS S3.
DUOS	Broad Institute	Open source (though not easy to spin up - based on early findings), in use by the Broad and for NIH, UI looks good on initial tour. Enhancements by request (at cost)	Team has met with Broad. The team led by John Pearson (QIRMB) had discussions with Jonathan Lawson (Broad Institute) late 2021, and Sarah Kummerfeld (Garvan) is progressing interaction.

¹¹ GA4GH GitHub <https://github.com/ga4gh>

6. Conclusion

The national human genome research community in Australia has well developed technical capabilities and systems for managing data. However, at present these are highly siloed and differ between institutions. Processes and supporting technologies vary considerably between different organisations, often relying on email trails for delivery. The discovery phase of the HGPP uncovered several contributors that fall within the remit of data access requests and data access approvals.

The absence of a trust framework and an inter-organisational DAC Automation system adversely impacts the national community's ability to leverage maximum value from the aggregation of data holdings. Reasons for this absence are articulated fully in the scoped requirements. High priority current-state issues include:

- Elapsed time from data access request to approval is often drawn out and is slowing down researchers' use of the data. The delay is also longer than operationally necessary.
- Applicants must re-enter their details, credentials and bona-fides anew for every data access application.
- Holding organisations and data owners need better visibility of researchers who have access to datasets or have received copies of data. Current information about the value researchers may have gained from data (what publications, what discoveries) could also be improved.
- Legal negotiations "can be the death of data sharing agreements". Standardisation and automation, finding commonality through trust frameworks and a shared system could reduce unnecessarily repetitive legal work.
- Excessive labour for DAC Coordinators, following up parties and using various communication pathways (emails etc) rather than having the convenience of a shared system.
- Excessive labour for DAC Committees to review and respond to requests and wait for more information. Committee members cannot easily review historic approvals - this would be useful to recall whether they had granted access for an aligned purpose or to the same group in the past.
- Excessive labour for data distributors since they cannot easily see the researcher details, data type and access granted, and the effective dates in one place.
- There can be hesitancy from Data Owners and DAC Committees around automation methods. Understandably this can include fears that automation may take away some of the important controls over data use.

The team 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:

- User interface provides metadata (to inform about dataset) and electronic form to apply for access.
- DAC automation system supports workgroup managers to manage their team and their access to resources.
- DAC automation system provides all of the required verified information to enable DAC committee members to assess Data Access Requests.
- Users can use the system to easily access information that supports their obligation to comply with the terms and conditions and reporting requirements associated with the data access granted to them.
- The DAC Automation System leverages persistent identifiers (for example ORCID ID, ROR, and RAID) to link entities and surface data to enhance decision making.
- Data distributors use the system to track the approvals granted and see the type of access granted. This assists in the distribution of only the datasets and types approved and to follow up where applicable egress billing or other charges arise (this has a dependency on data sharing agreements).
- The DAC Coordinator uses the DAC Automation system to manage and circulate all Data Access Requests with automation and workflow built into the DAC Automation system.
- Sufficiently granular data access approval capability is available to respect individual consent and restriction of data types. Future dynamic consent capability is not blocked/precluded by the components selected for the Minimum Viable Product.
- The system has the necessary flows and notifications to keep users informed of the progress of Data Access Requests.

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 the team will create work packages for DAC Automation 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.

Appendix A - Issues Referred to HGPP Project Reference Group

Table 17 - Issues of Note Identified as Out of Scope for DAC Automation Sub-project however due to value, referred to HGPP

#	Issues Description	Possible Mitigation Methods	In scope for DAC Automation in HGPP?
1	Multiple data distribution methods.	This is not covered by DAC Automation but has been referred to Project Reference Group	No
2	Timed encrypted URLs are commonly used for data access. Researchers frequently do not access on time so requests are repeated.	This is not covered by DAC Automation but has been referred to Project Reference Group	No
3	Cloud repositories are too expensive, egress charges, not enough bandwidth/capacity at the right price. Problem exacerbated when the data recipient tries to download multiple times.	This is not covered by DAC Automation but has been referred to Project Reference Group	No
4	Restrictions can at times kill long running processes (for example, where a security processes terminate multi-day executables).	This is not covered by DAC Automation but has been referred to Project Reference Group	No

References and Links

<< Some links have been redacted for internet publishing. For a copy of a cleansed artefact example please contact author Kylie Davies via Australian BioCommons.>>

AGHA Data Access Application Form (applicable for UMCCR) (obtained January 2021)

Garvan Data Access Application Form for Medical Genome Reference Bank (obtained October 2021)

NCI Data Collections and Publishing, NCI, viewed December 2021,
<https://opus.nci.org.au/display/NDP/NCI+Data+Collections+and+Publishing>

QIMR Berghofer Affiliate Agreement (obtained October 2021)

QIMR Berghofer Visiting Scientist Agreement (obtained October 2021)

QIMR Berghofer Volunteer Agreement (obtained October 2021)

QIMR Berghofer Data Access Agreement (redacted) (obtained October 2021)

REFEDS Assurance Framework, REFEDS, viewed 27 February 2022, <<https://refeds.org/assurance>>

ZERO CCI Data Access Application Form (obtained January 2021)

<<DAC Automation 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.)