November 2021

A National Approach to Genomic Information Management

Australian Genomics
Implementation Recommendations
Progress Report

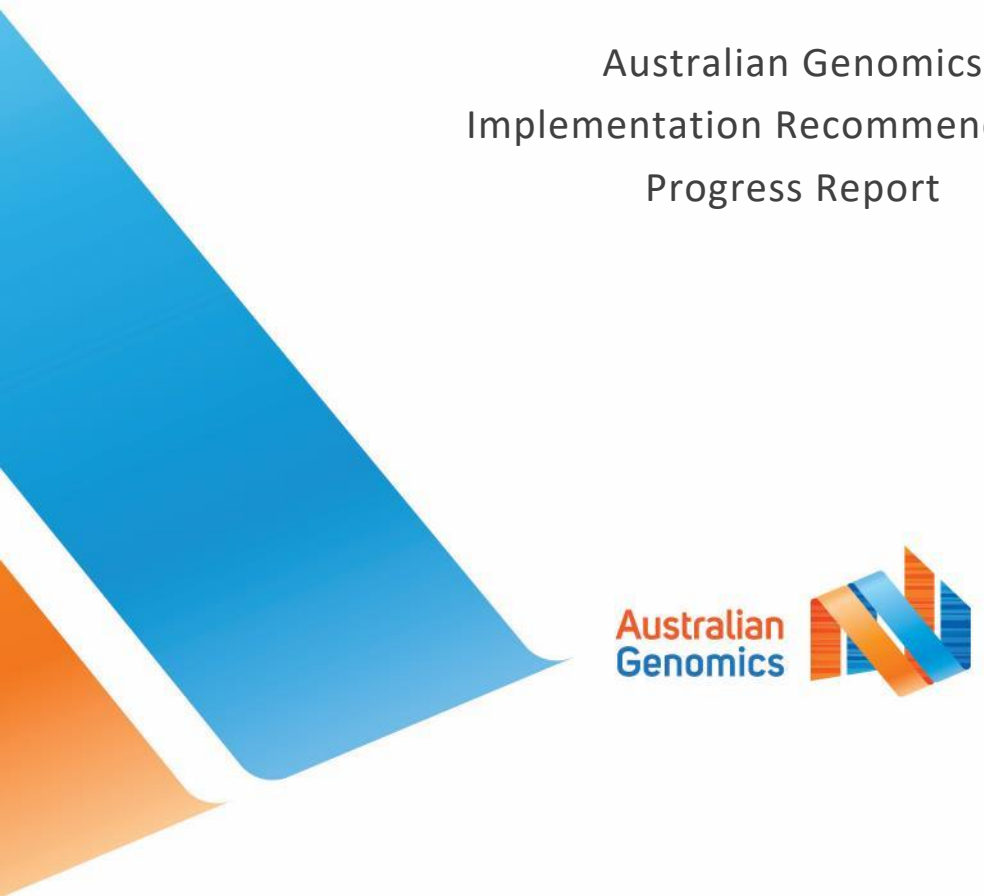

Table of Contents

Executive Summary ***2***

Progress Report ***3***

1 A National Approach to Genomic Information Management (NAGIM) 4

2 Infrastructure Prototyping Phase 5

3 Prototype Evaluation 13

4 Diagnostic Laboratories Implementation Priorities 15

5 Next Steps 18

Appendices ***19***

6 Appendix 1 – NAGIM Blueprint Summary 20

7 Appendix 2 – NAGIM Blueprint Roadmap – Proposed Priority Areas 21

8 Appendix 3 – Implementation Recommendations: Proposed Approach 23

9 Appendix 4 – NAGIM Implementation Survey for Diagnostic Laboratories in Shariant 24

10 Appendix 5 – Australian Genomics Priority Projects 25

Executive Summary

Australian Genomics was tasked by the Federal Government in 2021 to develop recommendations for implementing the Blueprint for a National Approach to Genomic Information Management (NAGIM).

The NAGIM Blueprint provided a proposed roadmap of activities for Infrastructure with a focus on:

i. Federated frameworks
ii. Standards-based processes
iii. Interoperability across systems using existing / accepted standards
iv. Alignment and interoperability with international data sharing initiatives
v. Cloud-based or hybrid solutions
vi. Basing decisions on national infrastructure on working pilot repositories

To progress implementation recommendations for infrastructure that will enable these elements of a NAGIM ecosystem for Australia, a **prototyping phase** is being undertaken, commencing with infrastructure for **research data**.

Twelve proposals were received, involving 7 different organisations, and a community of practice has been established. We are actively working to progress collaborative interactions across teams and are working on interoperability of systems.

A group of 8 international experts, from large-scale genome-sequencing initiatives, national precision health programs, and international bodies in interoperability, has been assembled to advise, and evaluate the NAGIM Blueprint Implementation Pilot. The group has met twice to shape the evaluation framework, and rather than just considering the individual prototypes or modules, the emerging emphasis is on the ecosystem as a whole: **interoperability, scalability and extensibility** are key.

The international evaluators have emphasised the global significance of this project.

Clinical activities are being progressed in parallel as a separate but complementary phase, in partnership with the Shariant users' network, which currently includes ~60 members across 19 laboratories nationally. A survey has been distributed to this group (October 2022) to explore the priorities of clinical and diagnostic laboratory systems. We will further engage with these clinical organisations to ensure that the research data pilots don't diverge from clinical information management evolution moving forward.

Despite the **ambitious timeline for the project,** it is meeting milestones:

- A scoping review was completed in June 2021, informed by prior work by the Genomics Health Futures Mission ICT recommendations; national and international data surveys of Australian Genomics; the data management ecosystem of the Targeted Call for Research 2016-2020; and the NAGIM Blueprint, to inform the design of the implementation strategy;
- Pilot implementations commenced in August, and are reporting back in December 2021;
- International assessors will evaluate the ecosystem over January 2022;

Draft recommendations to Governments will be framed February 2022, which will include key elements; early requirements; future development/infrastructure priorities; key stakeholders; ethical and policy considerations; and an indication of investment envelope and timelines.

Progress Report

1 A National Approach to Genomic Information Management (NAGIM)

1.1 The NAGIM Blueprint

Established following extensive national consultations, the Blueprint for a National Approach to Genomics Information Management (NAGIM; Queensland Genomics, Oct 2020) delivered a set of guiding principles for an **integrated genomics ecosystem** across healthcare and research.

From evaluating the current jurisdictional, operational, and technical landscape in Australia, the NAGIM Blueprint concluded that a **standards-based approach, using a federated (or hybrid) model,** would be the most appropriate strategy for a national approach to genomic information management in Australia (see ***Appendix 1*** – NAGIM Blueprint Summary).

The Blueprint also included a proposed roadmap of activities, as relevant to infrastructure, genomic medicine, governance, and genomic research (see ***Appendix 2*** – NAGIM Blueprint Roadmap – Proposed Priority Areas).

1.2 Australian Genomics NAGIM Implementation Recommendations

Australian Genomics was subsequently tasked by the Federal Government to develop recommendations for implementing the NAGIM Blueprint and progressing a national approach to genomic information management for Australia.

A scoping review was conducted (Australian Genomics NAGIM Briefing report, June 2021), identifying key implementation areas from the NAGIM Blueprint and prior relevant work; This included aspects of the Genomics Health Futures Mission ICT recommendations, national and international data surveys of Australian Genomics, and the Australian Genomics data management ecosystem from the National Health and Medical Research Council's Targeted Call for Research 2016-2020. These collectively aligned with progressing pilots of federated frameworks and standards-based approaches for cloud or hybrid systems, and the outcomes of this scoping review were used to inform the design of the implementation strategy.

The Australian Genomics strategy for developing implementation recommendations is leveraging infrastructure prototyping, research and clinical stakeholder engagement, and phased outcome and recommendation reports (see ***Appendix 3*** – Implementation Recommendations: Proposed Approach).

National infrastructure stakeholders have been invited to participate in a prototype construction phase, with the goal of **identifying the best combination of components that could serve as the basis for long-term national infrastructure.**

This is being progressed for research data in the first phase, with clinical stakeholders being engaged in parallel.

2 Infrastructure Prototyping Phase

2.1 Call for Participation

An open call was issued in July 2021 to invite participation in an infrastructure prototyping phase, to pilot components that would align to a future NAGIM ecosystem, and that could inform the recommendations being delivered to Governments in 2022.

A two-week EOI window was available for pilot teams to submit a proposal and confirm participation. Participation required teams that would be able to:

- Commence immediately in August
- Commit to a 5-month prototyping phase
- Build or extend an existing prototype using independent leveraged funds and resources
- Complete testable infrastructure and accompanying documentation, aligned to the NAGIM requirements (below), delivered by December

The prototyping phase was conducted for research infrastructure, as the first phase, with clinical activities to be progressed in parallel as a separate but complementary phase. This will include mapping intersections and gaps for clinical systems (see ***2.6***), identifying implementation priorities for diagnostic testing laboratories (***Diagnostic Laboratories*** Implementation Priorities), and stakeholder engagement with leaders from the jurisdictions and digital health sectors (***Next Steps***).

2.2 NAGIM Prototyping Phase Requirements

To progress implementation recommendations for infrastructure that will enable key identified elements of a NAGIM ecosystem for Australia, teams were notified during the call for participation, that a prototyping phase was being undertaken, commencing with *infrastructure for research data (phase 1)*.
Prototyping teams were expected to address ***NAGIM Roadmap priorities for Infrastructure***, and a set of associated ***NAGIM Pilot Technical Guidelines*** (below).

2.2.1 NAGIM Blueprint Priority Roadmap Activities for Infrastructure

The NAGIM Blueprint provided a proposed roadmap of activities for *Infrastructure* with a focus on:

i. Federated frameworks
ii. Standards-based processes
iii. Interoperability across systems using existing / accepted standards
iv. Alignment and interoperability with international data sharing initiatives
v. Cloud-based or hybrid solutions
vi. Basing decisions on national infrastructure on working pilot repositories

National infrastructure stakeholders were invited to participate in prototype construction, addressing these priority areas, with the goal of **identifying the best combination of components that can serve as the basis for long-term national infrastructure.**

This was emphasised as being an open process, that would seek to facilitate communication and collaboration between groups involved in developing prototypes.

2.2.2 NAGIM Pilot Technical Guidelines

Prototyping teams were provided with the NAGIM technical guidelines below.

"A National Genomics Infrastructure will ultimately have to handle millions of data objects. While prototypes are not expected to demonstrate the ability to handle data sets of this volume, the individual components should be designed and implemented with this scale in mind.

We expect almost all components to require interoperability with components up- and down-stream along the flow of genomic information. Interactions between components are expected to exceed the shown interactions; in general, prototypes should assume federation and provide APIs for data and information access to other components.

To facilitate this level of interaction the components should provide well-documented APIs based on existing, international standards where possible."

Prototype teams were provided with the diagram below (***Figure 1***) which outlined initial components required for a national genomics infrastructure with an emphasis on phase 1 (research data). Components shown in grey denote interfaces between the information architecture and end users and were not expected to be part of the phase 1 evaluation process.

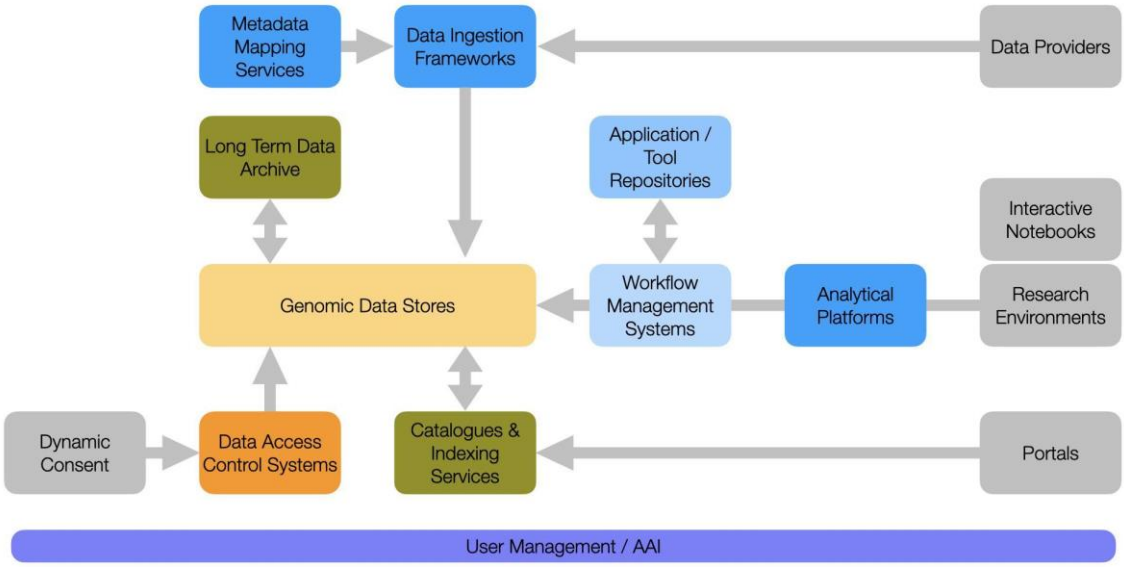

Figure 1. *Components of a national genomics infrastructure*

Component Descriptions:

- **Analytical Platforms:** Systems to design workflows, support workflow automation, monitor progress; enable external workflows to be run on locally stored genomic information without having to move/egress data; provide access points for secure research environments such as interactive notebooks, research UIs, portals (*e.g., Nextflow Tower, Terra, Hail)*.
- **Workflow Management Systems:** Frameworks to deploy, distribute and run computational pipelines written in standard domain-specific DSLs *(e.g., CWL, WDL, Nextflow)*, handle spare compute capacities *(Cromwell, Toil, Airflow, etc.)*.
- **Tool Repositories:** Systems for storing and sharing reusable analytical tools or workflows used by the workflow management systems *(Dockstore, EC2 Container Service, etc.)*
- **Data Ingestion Frameworks:** Data staging environments with processes to support data submission, check for data consistency, metadata availability, completeness prior to moving incoming data to an appropriate genomic data store *(e.g., Overture Song/Score, Human Cell Atlas-DCP).*

- **Metadata Mapping Services:** Ontology / vocabulary services to harmonize metadata accompanying genomic data submissions (*e.g., OntoServer*) and proposed metadata definitions and schemas.
- **Genomic Data Stores:** Object stores *(Gen3, Terra Data Repository, Arvados Keep, Overture)*
- **Data Access Control Systems:** Data Access Control frameworks and management systems that support machine-readable consent (*e.g., DUOS or REMS via GA4GH DUO*). Will need to support dynamic consent systems such as CTRL in phase 2.
- **Catalogues & Indexing Services:** Object-store indexing and querying capabilities to provide summaries of federated genomic data *(e.g., Gen 3 Object Indexing, Arvados Keep).*
- **User Management / AAI:** Infrastructure to underpin a national research identity network *(e.g., ELIXIR AAI, NIH Researcher Auth Service Initiative, AAF/CILogon)* with future support for GA4GH Passports.

It was noted that prototypes should be re-usable and ideally support deployment by interested parties in their own on-premises or cloud environment to test for interoperability with other components.

2.3 Prototyping Timeline and Deliverables

The following prototyping timeline (***Table 1***) was established, and all teams committed to completing the agreed deliverables by the end of year.

Table 1. *NAGIM Prototyping Timeline and Deliverables, and Responsible Parties*

Dec 20th 2021	***Pilot Infrastructure Completed*** *(Pilot Teams)*
Dec 20th 2021	***Pilot Documentation*** *(Pilot Teams)*
Jan 31st 2022	***Evaluation of Pilot Infrastructure*** *(External Assessors)*
Feb 28th 2022	***High-Level Recommendations Submitted*** *(Australian Genomics)*
May 2022	***Detailed Recommendations*** *(Australian Genomics)*

2.4 Prototyping Teams and the NAGIM Prototype Ecosystem

Twelve proposals were submitted, and a community of practice has been established, both formal (monthly meetings ~40 attendees per meeting) and informal (Slack channel, with 54 members).

The proposals received were:

1. **Human Genome Platform Project (*Australian BioCommons*)**
   Investigating best practice global technologies to support human genome data sharing, and provision of a services toolbox for Australian researchers.

   Lead: **Assoc. Prof Bernie Pope**
   Assoc. Director of Human Genome Informatics, Australian Biocommons

   Team: A. Lonie (Australian Biocommons), O. Hofmann (UMCCR),
   M. Cowley (Children's Cancer Institute), W. Kaplan (Garvan Institute),
   J. Pearson (QIMRB), H. Marks (AAF), B. Evans (NCI)

2. **Gen3 Data Commons on AWS (*UMCCR*)**
Piloting a genomics data store based on Gen3, a cloud-based platform for managing, harmonising, and sharing large datasets and its integration with other components (analytical workspaces, FHIR servers, etc.)

   Lead: **Assoc. Prof Oliver Hofmann**
   Head of Bioinformatics, University of Melbourne Centre for Cancer Research

   Team: F. Reisinger (UMCCR), V. San Kho Li (UMCCR), A. Patterson (UMCCR), S. Watts (UMCCR)

3. **VariantSpark (*CSIRO*)**
A cloud-based analytics platform for genome-phenome associations, with capability to target large-scale datasets.

   Lead: **Dr Natalie Twine**
   Team Lead, Genome Insights, Transformational Bioinformatics, CSIRO

   Team: L. Sng (CSIRO), P. Ramarao-Milne (CSIRO)

4. **sBeacon (*CSIRO*)**
An implementation of the GA4GH Beacon protocol, aiming to scalably and cost-effectively support queries and responses across future mega-biobanks.

   Lead: **Dr Laurence Wilson**
   Team Lead, Digital Genome Engineering, Transformational Bioinformatics, CSIRO

   Team: B. Hosking (CSIRO), Y. Yain (CSIRO)

5. **Trustless Data Access Control (*CSIRO*)**
A future-ready prototype to enable granular participant control over their data, using trustless principles such as self-sovereign identity and blockchain technology.

   Lead: **Dr Denis Bauer**
   Group Leader, Transformational Bioinformatics, CSIRO

   Team: J. Phillips (460 degrees), J. Spencer (460 degrees), C. Were (Verida), N. Lothian (Verida)

6. **Standardised Phenotyping Tools (*CSIRO*)**
Tools to enable sharing and querying clinical data in a federated way, using FHIR, Ontoserver, and FHIR-based tools for data queries (Pathling) and standardised data transformation (Redmatch).

   Lead: **Dr Alejandro Metke**
   Team Lead, Health Data Interoperability, Australian e-Health Research Centre, CSIRO

   Team: J. Grimes (CSIRO)

7. **Cloud-native bioinformatics tools (*QIMRB*)**
Modifying existing bioinformatics tools for whole genome sequencing analytical pipelines, to be cloud-native and interoperable with different on-premises and cloud environments.

   Lead: **Mr John Pearson**
   Lead, Genome Informatics, QIMR Berghofer

   Team: N. Waddell (QIMRB), C. Leopard (QIMRB), O. Holmes (QIMRB), S. Wood (QIMRB), C. Xu (QIMRB), M. Vidgen (QIMRB), R. Koufariotis (QIMRB)

8. **NCI-Garvan data stores and workflow (*NCI, Garvan*)**
   Genomic information management system across two on-premises environments: NCI and Garvan.

   Lead: **Dr Ben Evans**
   Deputy Director for HPC and Data Innovation, NCI

   Dr Warren Kaplan ***– [withdrawn]***
   Chief of Informatics, Kinghorn Centre for Clinical Genomics, Garvan Institute

   Team: M. Downton (NCI), K. Druken (NCI), A. Williams (NCI), A. Bayat (Garvan), J. Copty (Garvan), S. Ravishankar (Garvan), T. Nguyen (Garvan), M. Hobbs (Garvan)

9. **GeneTrustee data custodian framework (*Garvan*)** ***– [withdrawn]***
   Framework for managing genomic and clinical data access for research and clinical care

   Lead: **Dr Warren Kaplan**
   Chief of Informatics, Kinghorn Centre for Clinical Genomics, Garvan Institute

   Team: L. Burnett (Garvan), D. Degrave (Garvan), A. Palmer (Garvan), A. Hermanto (Garvan)

10. **Garvan Terra Platform on GCP (*Garvan, MCRI*)**
    Cloud-based data store and genomics workflow system demonstrating scalable processing capabilities.

    Lead: **Prof Daniel MacArthur**
    Director, Centre for Population Genomics, Garvan Institute and MCRI

    Assoc. Prof Sarah Kummerfeld
    Scientific Head, Kinghorn Centre for Clinical Genomics, Garvan Institute

    Team: L. Gruenschloss (CPG), D. Esposito (CPG, D. Reti (CPG), M. Franklin (CPG), V. Savelyev (CPG), D. Pavlic (CPG), J. Copty (KCCG), S. Ravishankar (KCCG), L. Goldstein (KCCG), A. Trinh KCCG), J. Torpy (KCCG)

11. **Max Kelsen Terra Platform on GCP (*Max Kelsen*)**
    Cloud-based platform for genomics information management, capable of supporting both research and clinical activities.

    Lead: **Mr Nicholas Therkelsen-Terry**
    CEO, Max Kelsen

    Team: C. Bean (Max Kelsen), N. Toosi Saidy (Max Kelsen), J. O'Farrell (Max Kelsen), A. Parker (Max Kelsen), L. Swan (Max Kelsen)

12. **VariantGrid (*UniSA/SA Pathology*)** ***– [withdrawn]***
    Variant database and analysis platform

    Lead: **Mr David Lawrence**
    Software Engineer

    Team: J. Andrew (UniSA/SA Path), S. King-Smith (UniSA/SA Path), J. Feng (UniSA/SA Path), H. Scott (UniSA/SA Path), A. Schreiber (UniSA/SA Path),

The GeneTrustee (#9) and VariantGrid (#12) submissions later withdrew. The corresponding proposed components are presented in ***Figure 2***, below.

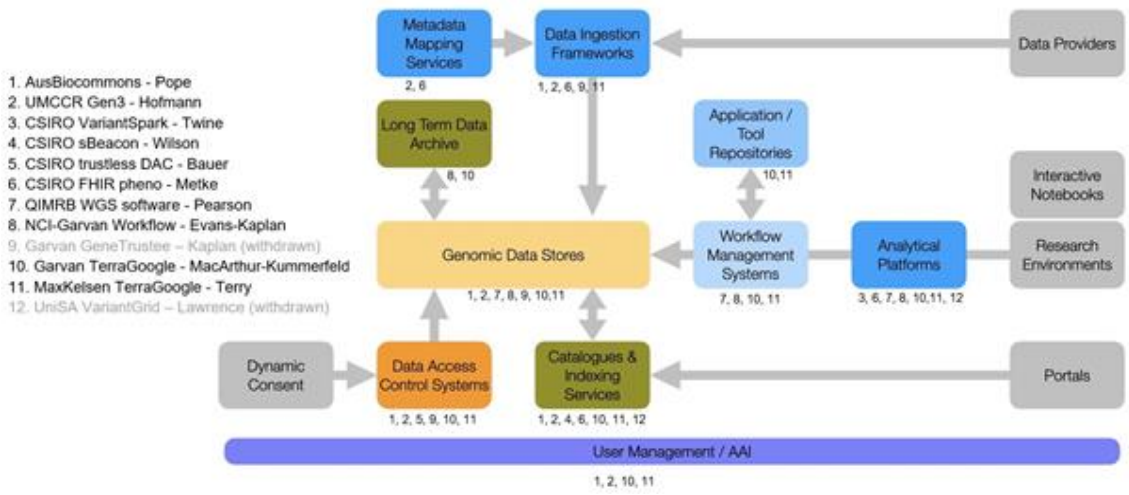

Figure 2. *NAGIM Prototyping Team Components*

2.5 Prototype Team Progress and Intersections

We are actively working to progress collaborative interactions across teams and are working on interoperability across the ecosystem (Figure 3).

After the opening kick-off meeting, teams were encouraged to self-organise their activities and collaborations. This was an organic process where each team progressed their own technology stack - but with regular cross team feedback, leading to opportunities to collaborate on integration across components.

Due to time constraints of the prototyping phase, the participants opted for different implementation strategies - ranging from standalone components to integrated proof-of-principle systems. Teams also outlined where future integration between defined components are technically feasible but only implemented the integration where the timeline allowed.

Not all components have been deployed on Australian-based infrastructure. This was due to resource constraints which affected the feasibility of deploying a unifying authentication/authorisation framework at this stage of the prototyping phase.

Some components have been able to take advantage of the CILogon infrastructure deployed and managed by the Australian Access Federation (AAF) in support of the NAGIM project. It allows seamless sign-on with institute credentials across the federation. Within this CILogon federation there has been opportunities to show interoperability and exchange using various standards.

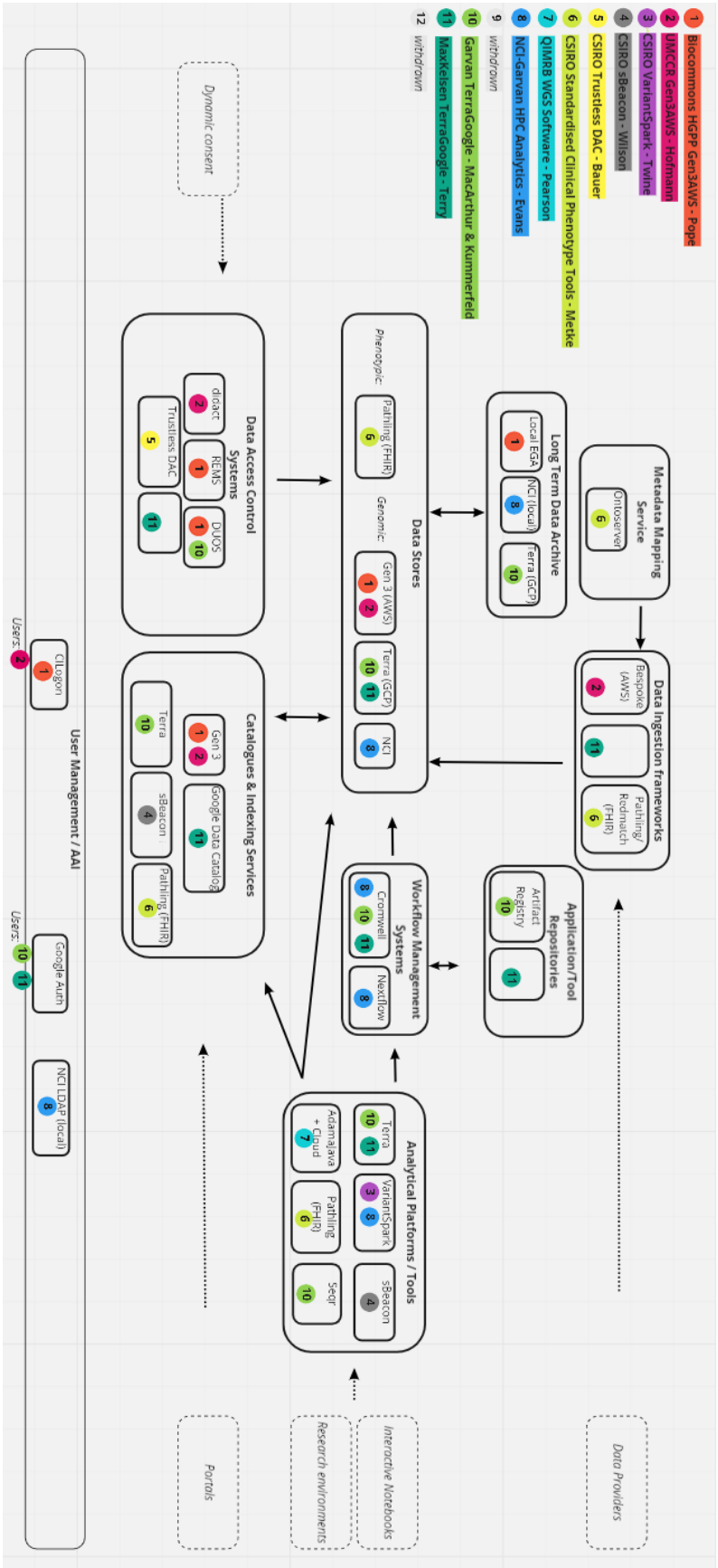

Figure 3. *NAGIM Prototypers' Systems and Overall Ecosystem*

2.6 Intersections with the Clinical Sector

Following the upcoming conclusion of the NAGIM prototyping development phase (December) and the prototype evaluations (January), review of the research infrastructure ecosystem in the context of clinical-grade systems will be pursued, including:

- Connection points for the healthcare system
- Leveraging AAF-provided authentication/authorisation frameworks in a clinical setting
- Expansion of FHIR support and expansion to EHR systems
- Adoption of unified identifiers
- Current laboratory processes for accessing data
- Representation and workflow processes around consent information; consent flow throughout the proposed components
- Clinically relevant pilot projects that may be leveraged

A range of these have additionally been identified as priorities from the diagnostic laboratories, for progressing the implementation of NAGIM implementation, based on the Shariant user group diagnostic lab survey (***Diagnostic Laboratories*** Implementation Priorities ***Appendix 4*** – NAGIM Implementation Survey for Diagnostic Laboratories in Shariant)

3 Prototype Evaluation

3.1 International Advisory Panel

A group of 8 international experts has been assembled to advise and evaluate the NAGIM Blueprint Implementation Pilot. The group brings significant expertise from large-scale genomic sequencing initiatives, national precision medicine programs, international standards in clinical informatics and interoperability in healthcare.

The international experts include:

- **Jonathon Dursi –** *Canadian Distributed Infrastructure for Genomics (CanDig)*
- **Augusto Rendon –** *Genomics England*
- **Daryl Waggot –** *Genome Canada*
- **Mar Gonzalez-Porta –** *Precision Health Research Singapore (PRECISE)*
- **Eric Banks –** *Broad Institute of MIT & Harvard*
- **Christina Yung –** *Ontario Institute for Cancer Research, Indoc Research*
- **Bob Freimuth –** *Mayo Clinic*
- **Grant Wood –** *Intermountain Clinical Genetics, HL7*

3.2 Evaluation Framework

The international advisory panel has convened twice to date, to review the NAGIM prototyping phase, and to shape the evaluation framework. The international evaluators have emphasised the global significance of this project:

"*This is an impressive activity and Australia should be proud to lead in this way and provide downstream benefits to its citizens. I can already say with certainty that the outcomes of NAGIM, including the assessment framework, will be used by organizations (and countries) in the future to guide their own work*." Bob Freimuth

The panel identified that given the extreme diversity in the prototypes - from a single-purpose or single component tool, through to comprehensive, end-to-end genomic data platforms, that establishing systematic metrics applicable to all prototypes, and applicable to infrastructure that is not yet mature or 'in production' is not likely to be feasible or optimal.

For the technical evaluation, the NAGIM advisory panel has recommended:

- Evaluating the NAGIM prototyping ecosystem *collectively,* as a whole, as well as *individually* for each prototype and module;
- Key emphasis to be on demonstrating **interoperability, scalability and extensibility**
- Include considerations of user experience, and intersections with clinical systems.
- Conveying status and maturity of prototype components and the status of interaction between components *contributed by different teams*

3.3 Final Prototype Submissions

3.3.1 Prototype Report

Prototype teams will be expected to submit as their final prototype and documentation:

- Report: prototype methodology, and how the NAGIM infrastructure priorities have been addressed
- Demonstration videos
- Description of interoperability – API documentation, standards used, integration map
- Test environment for members of the advisory panel to access (optional)
- Gaps and limitations of the prototype
- Intersection and capabilities for clinical systems

3.3.2 Prototype Team Recommendations

In the final report, the prototyping teams will have the opportunity to provide their comments and recommendations, regarding the future progression of an ecosystem, aligned to a national approach to genomic information management, based on their experience in the prototyping exercise and associated activities.

4 Diagnostic Laboratories Implementation Priorities

4.1 Diagnostic Laboratory Engagement

Activities are underway to brief diagnostic laboratories on the NAGIM implementation project, and identify implementation priorities, for a national approach to genomic information management, from the perspective of Australian diagnostic laboratories who are currently conducting genomic testing and generating clinical genomic sequence data.

This engagement is being progressed in partnership with the Shariant users' network, which currently includes ~60 members across 19 laboratories nationally (***Table 4***).

Table 2. *Shariant User's Network*

Laboratory/Service	State
Pathology Queensland	QLD
Cancer Genetics Diagnostic Laboratory, Kolling Institute	NSW
Children's Hospital Westmead	NSW
NSWHP – General	NSW
John Hunter Hospital	NSW
Randwick Hospital Campus Laboratory	NSW
Royal Prince Alfred Hospital	NSW
Concord Hospital	NSW
SA Pathology - General	SA
Frome Road Laboratory - Familial Cancer Lab	SA
Frome Road Laboratory - Molecular Genetics	SA
Women's and Children's Hospital - Molecular Genetics	SA
Women's and Children's Hospital - National Referral Lab	SA
Flinders Medical Centre - Molecular Genetics	SA
Royal Melbourne Hospital	VIC
Peter MacCallum Cancer Centre	VIC
Victorian Clinical Genetics Services	VIC
Melbourne Genomics - GenoVic	VIC
PathWest, Department of Diagnostic Genetics	WA
PathWest, Cardiovascular Genetics Lab	WA

4.2 Diagnostic Laboratories Survey

A survey addressing priority areas identified in the NAGIM Blueprint 'Horizons' roadmap, for ***Genomic Medicine*** and ***Infrastructure***, was distributed to Shariant users' network in early Oct (Survey provided in ***24***).

In considering these proposed priority areas from the Blueprint, the survey sought feedback from the laboratories in terms of their corresponding priorities, opportunities, and challenges for clinical and diagnostic information systems, in progressing a national approach to genomic information management.

4.2.1 Diagnostic Laboratory Responses

Formal survey responses were received from representatives from eight organisations:

- Melbourne Genomics/GenoVic, VIC
- NSW Health Pathology, NSW
- Children's Hospital Westmead (CHW), NSW
- Royal Prince Alfred (RPA), NSW

- Royal Melbourne Hospital (RMH), VIC
- SA Pathology, SA
- Victorian Clinical Genetic Services (VCGS), VIC
- Pathology Queensland (PQ), QLD

4.2.2 Key Themes

Key common themes, identified across the survey responses are summarised below.

i. Priority implementation areas of key value to laboratories:
   - Establish **national regulatory framework** to support data sharing - including national agreements
   - Establish **national standards and/or minimum requirements for clinical data capture** (phenotype, pedigree, metadata, consent)
   - Progress infrastructure and processes that would enable diagnostic labs to **access data nationally for** their primary diagnostic activities:
     - uncurated variants beyond Shariant
     - genotype-phenotype data from individual patients to aid interpretation
     - aggregated genomic data or control data for clinical pipelines e.g. CNV data, population-specific data
   - Establish **security and privacy** standards to safeguard data and systems, in pursuing interoperable infrastructure and data accessibility

ii. Barriers to progressing a national approach:
   - **Lack of interoperability within systems and data linkage challenges:** Difficulties, or inability, to access and link different data types (for a given patient) within the lab, and across healthcare settings within a jurisdiction;
   - **Governance:** existing processes are a barrier to data sharing;
   - **Resourcing and capacity**: labs need to focus on delivery of primary clinical activities; lack of resourcing as a key barrier for labs to progress many identified priorities areas for a national approach to genomic information;
   - **Local expertise:** limited governance and ITS expertise locally;
   - **Infrastructure maturity:** some lab infrastructures are not sufficiently mature to feasibly address identified activities or progress NAGIM priorities.

iii. Opportunities for progressing a national approach:
   - **Progressing cloud-based infrastructure**:
     - Some labs have already migrated to cloud, and others are in the process of doing so, or actively looking to do so;
     - Shared learnings or communications, and development of cloud/hybrid solutions that can be adopted by labs who are looking to migrate in the future, or who are without extensive local ICTS expertise to do so, was identified as valuable.
   - **Communication, engagement, piloting**:
     - Labs sought increased clarity and communication around progressing NAGIM, including several who are establishing future strategic priorities and IT roadmaps;
     - Select labs/organisations are particularly motivated to commence cross-jurisdictional engagement;
     - Potential opportunity to address **common standards, establish data sharing or infrastructure pilots**, and/or convene a **focused national working group** around clinical data infrastructure, with labs that have capacity and interest to do so.

4.2.3 Subsequent Engagement

Engagement will now progress around these priorities, to identify next steps and recommendations.

Additional engagement with these groups around the research data pilots (***Infrastructure Prototyping*** Phase) will also be progressed, to ensure the research infrastructure pilots and associated recommendations don't diverge from clinical information management evolution moving forward.

5 Next Steps

5.1 Additional Stakeholder Engagement and Contributions to the Recommendations

Australian Genomics will now be ensuring that the outcomes of the NAGIM briefing documents, research prototyping, and diagnostic lab engagement are communicated to the NAGIM stakeholders.

This will involve continued engagement with the diagnostic laboratories, and additional engagement with:

- The Australian Digital Health Agency
- The Australian Government Department of Health Digital Health Branch
- Leaders in Aboriginal and Torres Strait Islander genomics and data sovereignty
- Jurisdictional and healthcare representatives, including those of the Healthcare CEOs Forum, and the Australian Genomics National Implementation Committee

Australian Genomics will also leverage its Priority Project activities (***Appendix 5*** – Australian Genomics Priority Projects) and to engage with our project leads and working groups, for relevant projects. This includes the priority projects around health systems consistency, standardised phenotypes, data governance, consent, and RCPA Quality Assurance.

Collectively, these stakeholders will be consulted to inform the development of recommendations to be submitted in February.

5.2 Progression Timeline

Table 3. *NAGIM Project Progression Timeline*

Completed	**Oct 2020**	*NAGIM Blueprint completed by Queensland Genomics*
	Apr 2021	*Australian Genomics requested to develop implementation recommendations*
	Jun 2021	*Australian Genomics scoping document on implementing NAGIM*
	Jul 2021	*Infrastructure Prototyping phase defined and AG call for participation*
	Aug 2021	*Participating teams confirmed and prototype phase initiated for research data*
	Sep 2021	*International advisory panel assembled and briefed*
Ongoing	**Oct 2021 – Jan 2022**	*Parallel stakeholder engagement (diagnostic labs, digital health, jurisdictions, indigenous health leaders, Australian Genomics network)*
	Nov 2021	*Prototype evaluation framework developed*
	Dec 2021	*Completion of prototyping*
Upcoming	**Jan 2022**	*Prototypes and ecosystem evaluated by international assessors panel*
	Feb 2022	*Submission of high-level recommendations to government*
	May 2022	*Submission of comprehensive recommendations to government*

Appendices

6 Appendix 1 – NAGIM Blueprint Summary

6.1 Extract from Australian Genomics NAGIM Summary Briefing (p.9, June 2021)

Established following national consultations, the Blueprint for a National Approach to Genomics Information Management (NAGIM; Oct 2020) delivered a proposal for an **integrated genomics ecosystem** across healthcare and research organisations, to support genomic research and genomic medicine, simultaneously.

From evaluating the current jurisdictional, operational, and technical landscape in Australia, the NAGIM Blueprint concluded that a **standards-based approach, using a federated or hybrid model,** is likely to be the most appropriate strategy for a national approach to genomic information management in Australia.

The NAGIM Blueprint also reported an increasing preference, in clinical and research settings, for progressing **cloud-based solutions**, and identified the following factors as crucial for a future national genomics infrastructure:

- **Interoperability** across systems, with standards-compliant application programming interfaces (APIs) to support seamless access to information, within and across healthcare services and research organisations;
- Adoption of **international standards** and alignment with large-scale genomic initiatives, to enable international data sharing.

6.1.1 High-Level Requirements

The NAGIM Blueprint identified several high-level requirements in its "Considerations for designing a framework" (*Chapter 4*).

These include:

1. **Interoperability** of systems between i) research and healthcare systems nationally, and ii) between Australian and international systems;
2. Federated approaches to **genomic medicine** data repositories using standards-based development to support the different jurisdictions that have different data management solutions, legislation and regulation;
3. **Access** to genomic, clinical and phenotype data from healthcare, for Australian researchers
4. Nationally coordinated approach to **genomic research** capabilities, with multiple standards-based repositories, to address the inability for all research repositories to be able to be combined;
5. Address specific needs of **Aboriginal and Torres Strait Islander** communities and initiatives;
6. Improvements in **privacy, consent and security**

7 Appendix 2 – NAGIM Blueprint Roadmap – Proposed Priority Areas

Table 4. *Summarised Extract from NAGIM Blueprint (Queensland Genomics, October 2020)*

Priority Area	Proposed Activities
Infrastructure	• **Implementation studies** of the leading genomics systems in use across Australia • map against the logical model • establish baseline and learnings for future implementations • examine existing & emerging (cross-jurisdictional) research partnerships • study clinical/research partnerships • A **standards-based, interoperable approach to cloud adoption** for storing and retrieving genomic data (medical and research) • **Standards for self-describing repositories** that identify their content and capabilities -Work with international groups (e.g. GA4GH) • **Trial a shared, cloud-based repository for genomic research data across jurisdictions** to establish baseline and inform future implementations • **Standards for federated query** across genomic data repositories. • **Standards for international research data sharing** - Work with international groups
Governance	• Establish or leverage a **national governance group** comprising clinicians, researchers, policy makers, funders, consumers, and Aboriginal and Torres Strait Islander people to coordinate activities over the three horizons. The governance group should be informed by focused working parties and be inclusive of industry players acting in partnerships. • Developing a **robust data governance framework** that ensures that relevant protections are in place to protect the genomic information of individuals and groups should be a priority first action of the national governance group. • Consideration should be given to whether a **national or jurisdictional Data Custodian/Steward** is required to provide oversight of how data is managed, accessed, and shared. • **Confirm or amend the roadmap elements** of this national approach to genomic information management. • **Identify an organisation/group with the capabilities** to operate a national genomic information network or build a federated structure for all jurisdictions to participate equally. • Establish a **national consumer engagement group** to ensure that genomic data activities meet community expectations for addressing risks and benefits. This group should include representation of Aboriginal and Torres Strait Islander people and other groups with specific needs (such as Culturally and Linguistically Diverse (CALD) communities). • Agree/adopt national standards for **genomic data storage formats, genomic data exchange methods, computable consent and cybersecurity** policies, guides and standards informed by existing national and international standards. • Agree on an **interoperability capability model** that allows for organisational self-assessment in support of planning and funding decisions. • Agree on **national data retention policies** for all classes of genomic data that consider both clinical, diagnostic service and research requirements.
Medical Genomics	• Promote **collaboration and share learnings between the jurisdictions** undertaking activities, those planning such activities and other interested parties. • Establish a cross-jurisdictional working group to **standardise access to familial and pedigree data** for clinical purposes. • Establish **national agreements for genomic data sharing for clinical purposes**, leveraging existing clinical data sharing agreements working with private and public providers. • Establish an **agreed approach to capture or mapping of phenotype data** within clinical systems to support genomic diagnosis, predictions, and research. • Support ongoing operation and expansion of **variant curation repositories and tools** (e.g. Shariant) to support genomic medicine.

Genomics Research	• Establish national agreements for genomic data sharing for research, leveraging existing data sharing agreements. • Establish a national research consent mechanism for genomic data utilising strong credentialing for participants with dynamic approaches to ongoing engagement • Continue trials of research data sharing with leading clinical groups, leveraging existing genomic programs, to establish baselines and learnings for later implementations. • Establish national arrangements to consider Australia's access to and use of global genomics data assets, our dependencies and role on the world stage.

8 Appendix 3 – Implementation Recommendations: Proposed Approach

8.1 Extract from Australian Genomics NAGIM Summary Briefing (pp15-16; June 2021)

The materials reviewed suggest consensus on several key points relevant to implementation:

- It is unlikely that a single fully centralised system for genomic data storage and analysis can serve the needs of both research and clinical users nationally;
- A data model based on **federation across multiple repositories** using standardised interfaces will likely be necessary, especially for clinical data; however, a **single unified repository for research data** may be tractable and would be highly preferable;
- There is a growing preference for **cloud-based models** for international genomic data solutions, rather than models based on centralised HPC;
- **Decisions on potential platforms should be made based on the evaluation of working prototypes managing data across multiple jurisdictions.**

Australian Genomics will develop **a series of phased recommendations** that will be based on:

- Infrastructure prototyping of components that meet NAGIM recommendations;
- Outcomes from clinical and research stakeholder engagement;
- International evaluation and best practice

National infrastructure stakeholders will be invited to participate in prototype construction, definition of technical specifications, and infrastructure evaluation. This will be designed as an open process to facilitate communication and collaboration between groups involved in developing prototypes, with the goal of **identifying the best combination of components that can serve as the basis for long-term national infrastructure.**

Clinical and research stakeholders will be engaged in parallel to identify and progress additional core elements of the NAGIM ecosystem, including priorities for clinical data and data governance.

Table 5. *Australian Genomics Strategy for Developing Recommendations*

Approach	Components
Stakeholder Engagement	• Engage key stakeholders and infrastructure implementers • Define the intended NAGIM implementation ecosystem • Identify what can be centralised • Determine priorities for clinical and governance
Infrastructure Prototyping and Evaluation	• Define required components • Define critical specifications and evaluation process • Compare systems and technologies: • Computational infrastructure providers • Data warehousing platforms • User authentication and access control systems • Scalable platforms for data processing and analysis • User interfaces for core analyses • Other components identified by stakeholders • Assess the respective role of HPC and cloud computing platforms • Assess capability to support federation across states
Phased Recommendations for National Implementation	• Initial recommendations strategy • Prototype-informed recommendations • Research and clinical approaches

9 Appendix 4 – NAGIM Implementation Survey for Diagnostic Laboratories in Shariant

9.1 Implementation Recommendations for a National Approach to Genomic Information Management (NAGIM): Questions for Diagnostic Laboratories

Overall NAGIM goal: Standardisation of processes/data, and data sharing between clinical laboratories using a federated model (i.e. data remains in your state/jurisdiction but is standardised and accessible via remote querying)

Scope: Human genomic data (FASTQs, BAMS, VCFs), **associated metadata** (e.g. sequencing and methodology information), **clinical information** (phenotypes, condition, pedigree information, curated variants, test reports), **consent and any other data/information** useful for clinical research activities.

Below are the NAGIM Blueprint's Roadmap priorities relating to 'Medical Genomics' and 'Infrastructure'.

The NAGIM Blueprint Roadmap identifies priorities in **Medical Genomics** as:

- Collaborations and shared learnings between jurisdictions
- Standardised access and format of pedigree information
- National agreements for clinical data sharing
- Standardised approach to clinical phenotype data capture
- Expansion of variant curation repositories (e.g. Shariant)

The NAGIM Blueprint Roadmap identifies priorities in **Infrastructure** as:

- Supporting federated querying and frameworks
- Standards-based processes
- Interoperability across systems
- Alignment and interoperability with international data sharing initiatives
- Cloud-based or hybrid (HPC and cloud) solutions
- Basing decisions on national infrastructure on working pilot repositories

Australian Genomics is now seeking feedback from diagnostic laboratories on their recommendations for **implementing aspects of NAGIM**. Please return responses by Friday November 5th to:

MJ Brion (Australian Genomics Data Manager): Marie-Jo.Brion@qimrberghofer.edu.au

Laboratory: _____

Contributors (*Names and role at the lab*): _____

1. How much of a priority are these for your lab and are there other higher priority aspects that would progress NAGIM?

2. Including additional priorities identified in question 1, what NAGIM-relevant activities would you start with and how?

3. What are the challenges and opportunities for your lab, for progressing these NAGIM-relevant activities?

4. Any other comments?

10 Appendix 5 – Australian Genomics Priority Projects

The below activities have been identified by the Australian Government and Australian Genomics Investigators as priority projects and activities to be undertaken as part of the Australian Genomics Grant Program 2021-2023.

Table 6. *Australian Genomics Priority Projects*

Project	Activities
Support of Government-funded genomic research with established research networks and capabilities	• Support of a national network of coordinators and genetic counsellors • Access to established research capabilities, infrastructure, and resources • Systems for genomic research data management and standardization • Support genomic research outputs – transition, translation, uptake, sustainability
Multidisciplinary Clinical, Diagnostic, and Research Network & Specialist Working Groups	• Convening a national multidisciplinary Clinical, Diagnostic, and Research Network • Specialist genomic working groups to share approaches to genomic priority areas • Establishment of a national expert working group for future diagnostic technologies
RCPA QAP genomic interpretation module	• RCPA QAP interpretive module feasibility assessment, development, and maintenance
Progress the National Approach to Genomic Information Management (NAGIM)	• Continues development of national approaches to clinical genomic consent • Establishment of a multidisciplinary collaborative data group • Progress the NAGIM Blueprint Implementation and develop recommendations • Support uptake of research capabilities and optimized national standards • Standardise/converge approaches to genomics information management • Progress the optimal collection, standardisation, and sharing of data • Formulate recommendations for data aggregation and quality control • Collaborate on the development and sharing of open-sourced genomic tools • Develop a sustainability model for ongoing data/system management
Pilot Data Implementations	• Build a modern data commons (Gen3), deploy a platform for collaborative analysis (Seqr) • Identify, collaborate with, and evaluate other pilot data implementations
Data sharing infrastructure/practices	• Incentivise and support sharing of research data • Promote governance policies and procedures in support of secondary research data use • Implementation data access control systems to streamline data sharing

Phenotype data capture and sharing	• Review national clinical phenotype data collection • Progress standardised, detailed clinical data collection • Identify limitations and potential solutions for clinical data sharing • Evaluation of Electronic Request Forms – national activity and approaches
Shariant	• Progress the development, implementation, and evaluation of Shariant
PanelApp Australia	• Continue to enhance the content and support the uptake of PanelApp Australia
Policy Network	• Convene Policy Network Meetings to progress genomic discussions and implementation
Health Economics	• Evaluate economic impact of genomic research to inform policy and system uptake • Design a minimum health resource utilisation instrument for health economic data
Implementation Science	• Apply implementation science methodologies and engage health services
Indigenous Priorities in Genomics	• Indigenous Genomics Priorities: laying the foundations to a long-term vision • Community engagement; data sovereignty; integrated genomic healthcare
Bioethics research	• Establish a national collaborative ELSI network • Progress standard approaches to secondary/incidental findings nationally • Return of results and recontact; evaluating issues of genomic equity and reach
Reports, consultation responses	• Develop reports; papers; summaries; responses to Government/Industry consultations • Progress evidence-based policy development
MSAC, PBAC, HTA Assessments	• Support health technology assessment of genomic applications • Undertake evaluations of the implementation of new genomic MSAC items
Evaluations of jurisdictional policy practice, testing demand and capacity, genomic implementation	• Develop broadly consultative, evidence-based recommendations for genomic practice • Progress priorities identified by Government genomic policy frameworks • Evaluate the integration of genomics targeted to the primary health care • Review State genomics plans/strategies, translational genomics framework (NSW) • Evaluate testing demand, volume, and capacity • Achieve consistency between Health Systems including clinical and laboratory settings • Explore activity-based funding model evidence requirements, with IHPA & Governments • Develop guidelines around reanalysis of clinical data
Nationally-consistent genomic terminology	• Progress development and uptake of nationally consistent genomic terminology

Genomic test directory	• Explore the feasibility of and demand for an Australian genomic test directory
Workforce education, conferences, symposia	• Progress efforts to enhance genomic literacy for the broader health workforce • Develop materials and support continuing professional development • Convene meetings, workshops, and symposia
Genomic traineeships and internships	• Coordinate workforce development and training activities, including internships
Incubator/implementation projects	• Genomic implementation/incubator projects • Pharmacogenomics & Polygenic Risk Score Incubator Projects • Evaluate implementation/incubator projects to inform investment, policy, and practice
Involve Australia	• Develop policies and standards to improve public involvement in genomic research • Facilitate adoption of policies and evaluate effectiveness and impact • Genomicsinfo.org.au – improve genomic literacy of the public
SING Australia – Summer Internship for Indigenous Peoples in Genomics	• Support the coordination if the annual SING Australia • Review other potential projects to progress Indigenous genomic priorities
International Engagement – GA4GH, GHIF (International Advisory Group)	• International engagement and collaboration – GA4GH, GHIF, direct collaborations • Hardware/software engineers dedicated to international genomic data collaborations

Broad collaboration, across disciplines and jurisdictions, will be critical to progress these activities. If your organisation is active in these areas, or if you'd like to contribute to these efforts, please contact: info@australiangenomics.org.au