SUPPLEMENTARY INFORMATION

for the

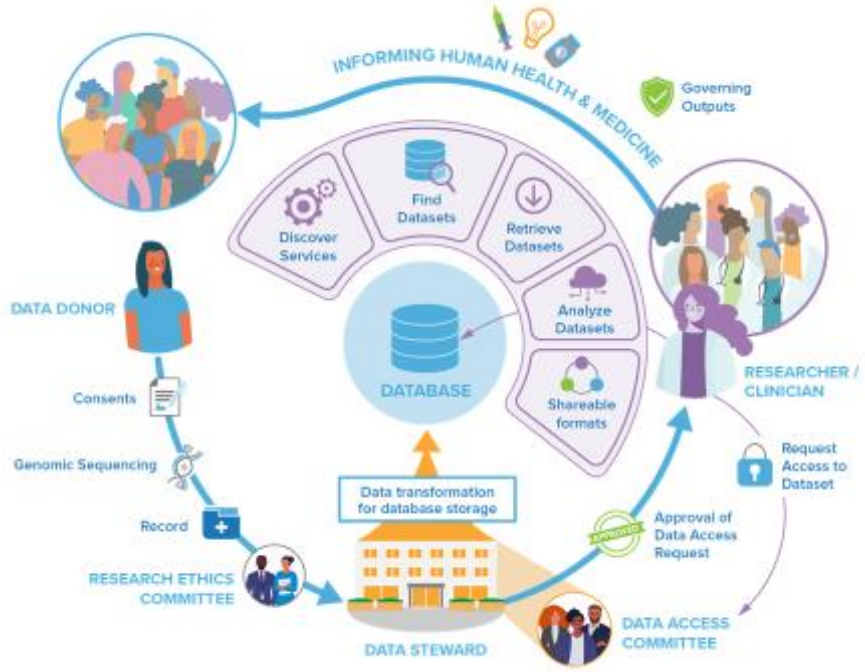
National Approach to Genomic Information Management (NAGIM)

Implementation Recommendations

Australian Genomics

TABLE OF CONTENTS

APPENDIX A – Global Alliance for Genomics and Health - Genomic Data Sharing Ecosystem



APPENDIX B – The NAGIM Blueprint

Key elements of the Blueprint for a National Approach to Genomic Information Management (NAGIM Blueprint) were summarised in the Australian Genomics Implementation Scoping Review, provided below. For full details please refer to the original NAGIM Blueprint.

Principles for a NAGIM

The NAGIM Blueprint principles are listed below.

Table B1. NAGIM Blueprint Principles

CONSUMERS & COMMUNITIES	ABORIGINAL AND TORRES STRAIT ISLANDER GENOMICS	GENOMICS RESEARCH	TRANSLATIONAL GENOMICS	GENOMIC MEDICINE	DATA MANAGEMENT
Person-centred focus	Collective & individual benefit	Rationalised repositories	Efficacy, utility & effectiveness	Maximise clinical benefit	Clinical vs research data
Trust	Authority to control	Ethical data & provenance	Leverage research flexibility	Support for data sharing	Multiple repository environment
Informed consent	Responsibility	Reuse with permission	Emerging & validated knowledge	Sharing across boundaries	Genomic data retention
Right to access	Ethics	Plan for change & scalability		Sustainable genomics	National & international collaboration
Use of data / portability				Frequently updated knowledge	Strong governance models
Equity of access				Genomic data is clinical data	Contemporary use & contribution
Benefit of use					Pragmatism
					Data quality

All 32 principles are detailed in the NAGIM Blueprint pages 11 – 34.

NAGIM Blueprint High level Requirements

The NAGIM Blueprint identified several high-level requirements in its "considerations for designing a framework".

These include:

- Standards based interoperability (between research and healthcare, and Australia and internationally);
- Interdependence between research and health delivery;
- Decentralisation of genomic data repositories (via standards and federated approaches);
- Researcher access to genomic data;
- Nationally coordinated approach to research capabilities;
- Addressing needs of Aboriginal and Torres Strait Islander peoples;
- Improvements in privacy, consent and security;
- Consent mechanisms; and
- Building for the future.

For a complete description of these high-level requirements see page 47, Chapter 4 of the NAGIM Blueprint.

Proposed Architecture for Genomic Research and Genomic Medicine in Australia (*Chapter 5)*

The Blueprint outlines a proposed logical architecture for a future national ecosystem supporting:
i) genomic medicine, ii) genomic research and iii) integrations between the two.

The current architecture in Australia

The current architecture is described in the Blueprint for:

i) Australian genomic medicine systems: including information systems for operating health services (Electronic Health Record (EHR) and laboratory information management systems (LIMS), clinical sequencing and bioinformatic analysis capabilities, and for exchanging clinical genomics knowledge;

ii) Australian genomic research systems: including information systems for managing research data (Research Data Management Systems (RDMS), Data Access Management systems (DAMS), research sequencing and bioinformatic analysis capabilities, and for exchanging genomic information;

iii) Both frameworks: the Blueprint notes there is a combination of local system-focused repositories (EHRs and LIMS, for genomic medicine; RDMS and DAMS, for genomic research), data staging systems on local and cloud infrastructure, and core genomic data stores in a variety of databases, repositories and formats.

Considering these architectures and based on their consultations, the NAGIM project identified:

- **Critical to national adoption**, the proposed future architecture will require standards-based interfaces (APIs) to other systems; and data orchestration between repositories and systems;
- Importance of data flow connections to external data repositories, that are both providers of critical information for genomic activities and recipients of data from our health service and research organisations;
- Increasing preference for cloud-based genomic data stores, amongst clinicians and researchers in Australia and internationally, who were consulted as part of the NAGIM project.

The NAGIM Blueprint listed 'patterns of interactions' that would support a mature national genomics ecosystem:

- Point-to-point requests for data (standard data custodian process);
- Synchronisation of datasets (agreed standards for storage, transport, access; federated national genome archives);
- Remote querying (data requestor defines query; data provider executes and returns results; supports FAIR data);
- Federated queries (remote query for multiple datasets);
- Self-describing repositories (capabilities with an interface that allows high-level queries by others).

Increasing Genomic Data Interoperability

The Blueprint noted that the current architecture in Australia is largely bespoke 'point-to-point' data flow, with inconsistent standards use. Potential next stages are described that would ensure progress to a mature genomic data ecosystem, by incrementally establishing:

- standards-based genomic data processes: adding standards-based processes to existing systems, across organisations; developing consistency to allow aligned solutions to emerge.
- standards-based integration: nationally agreed standards for federated queries across multiple data providers, nationally supported identity management, national approach to consent.
- standards-based interoperability-enabled systems: the 'highest level of maturity', progressing from integration to an interoperability-enabled standards-based system.

The NAGIM Blueprint identifies the need for **flexibility** in working towards a mature genomic data ecosystem, noting that not all jurisdictions or research organisations will have the same capabilities or priorities, and a mix of technologies and capabilities will likely remain over the short to mid-term, as those organisations with capability and capacity lead development. Standards-based APIs will allow us to add these organisations to the ecosystem as their capabilities improve.

Determining the correct approach

The NAGIM Blueprint notes that when considering the architectural approach for Australia, not everything can, or should, be centralised. But, where appropriate, the Blueprint recommends a central authority establish the aspects of the national genomic information network that need centralising.

The NAGIM Blueprint also acknowledges that growing capabilities of cloud technologies are changing perspectives about what 'national' or 'centralised' systems should look like, and urges the future governance group and national participants to be open to new approaches to federation.

The Blueprint encourages consideration of the following, for a federated interoperable system:

- Jurisdictional differences in data management requirements, policy, funding, priorities;
- Central management of some essential functions through a federated model, to lower barriers to uptake by less mature or 'resource poor'[1] jurisdictions;
- Federation would enable scale and cost levelling; and
- Centralised, distributed, large or small compute capabilities – either part of or outside the system – are all compatible with a federated model.

The Blueprint argues that **a hybrid solution employing advanced technology and best practice approaches could leverage both centralised and highly decentralised models.**

Genomic Data Governance (*Chapter 6*)

The NAGIM Blueprint asserts that a national approach to genomic information management will require a strong governance framework that can be applied with consistency, with an operational model that can support the **diverse requirements of both the clinical and research sectors**.

The Blueprint notes a future national approach will need to address **significant complexities around data sovereignty**, for national and international data sharing and data storage, when considering jurisdictional data, and data from Aboriginal and Torres Strait Islander peoples; as well as complexities with data ownership, permissions, and intellectual property rights, that are **unique to human genomic data** generated from healthcare and research.

The Blueprint defines five governance operational models (Decentralised, Network, Centralised, Federated, Hybrid)[2] and concludes **that the complexities of the Australian healthcare and research sectors suggest that a federated or hybrid model are most appropriate.**

Standards and interoperability (*Chapter 7*)

The NAGIM Blueprint defines interoperability as "… the ability of different information systems, devices and applications to access, exchange, integrate and cooperatively use data in a coordinated manner, within and across organisation, regional and national boundaries to provide timely and seamless portability of information and optimise the health of individuals and populations globally".

The Blueprint emphasises the importance of interoperability for Australian genomic data systems, further noting interoperability as one of the seven strategic priorities for the Australian health sector, by the National Digital Health Strategy, and the progression of the Australian Digital Health Agency's interoperability program and it's expected potential impact for genomic data standardisation.

The NAGIM Blueprint has concluded that to support interoperability in the health sector – **architectures, application interfaces (APIs) and standards are required** to enable data to be accessed and shared appropriately and securely, within all applicable settings and with the relevant stakeholders. These include HL7 standards for integration and exchange of electronic health information, Observational Health Data Sciences and Informatics (OHDSI), and metadata standards.

[1] As termed in NAGIM Blueprint

[2] NAGIM Blueprint Definitions: **Decentralised operating model:** Data management responsibilities are distributed across multiple functions with no single owner. This provides the simplest structure, but governance and decision-making are more difficult. **Network operating model:** More formalised than a decentralised model, a network model introduces defined relationships and accountabilities. The difficulty is in maintaining the defined relationships and expectations. **Centralised operating model:** The most formal and mature model but requires substantial organisational change to achieve and the separation of data management from the operational 'coal face' can lead to a lack of focus on the strategic outcomes. **Federated operating model:** A federated model provides a centralised strategy with decentralised execution. A centralised coordination process is required, and this can introduce complexity through the need to balance operational independence against the needs of the whole. **Hybrid operating model:** In a hybrid model, data management is coordinated through a centre of excellence working with more decentralised operating areas, supported with more tactical working groups.

In particular, the Blueprint specifically highlights the international standards and tools of **GA4GH**, as core standards considered important to achieving interoperability. These include APIs such as for authenticating researchers, issuing approvals for data access, discovering and querying data, extracting data, and running workflows in different computing environments.

NAGIM Blueprint Proposed Roadmap of Activities

The NAGIM Blueprint outlines three phases ('Horizons'), to achieving a future national approach to genomic information management, as an ecosystem that would support both genomic medicine and genomic research (*pp. 59-63*). The Horizons, or phases are defined around core themes of Governance, Medical Genomics, Genomic Research, and *Infrastructure.*

The proposed NAGIM '*Infrastructure*' activities are extracted below.

Table B2. Extract from the NAGIM Blueprint 'roadmap for implementation' - Infrastructure

NAGIM Horizon: Infrastructure	Infrastructural activities required to support the delivery of the medical genomics and genomic research
Horizon 1: Leverage and Plan	• Undertake **implementation studies** of the leading genomics systems in use across Australia to map against the logical model and establish baseline and learnings for future implementations. Such studies should examine existing research partnerships (ideally cross-jurisdictional) as well as existing and emerging jurisdictional solutions. A study of clinical/research partnerships would be beneficial. • Develop a **standards-based, interoperable approach to cloud adoption** to support storage and retrieval of genomic data in both medical and research domains. • Work with international groups (such as GA4GH) to agree **standards for self-describing repositories** that can identify their content and capabilities. • **Trial the establishment of a shared, cloud-based repository for genomic research** data across at least two jurisdictions to establish baseline and learnings to inform future implementations. • Establish **standards for federated query** across genomic data repositories. • Work with international groups to agree **standards for international research data sharing**.
Horizon 2: Build on Foundations	• Adopt **national interoperability for cloud** infrastructure. • **Trial federated query standards** across repositories to support a national genomic information network operational framework. • Expand a **shared, cloud-based repository for genomic research** data across at least two jurisdictions to establish baseline and learnings to inform future implementations. • Work with international groups to operationalise **international research data sharing**.
Horizon 3: Transition and Operate	• Continue roll out and standardisation of **national interoperability for cloud** infrastructure. • Leverage **federated query standards** across repositories to support a national genomic information network operational framework. • Expand a **shared, cloud-based repository for genomic research** data across all jurisdictions to complete the national genomic information network operational framework. • Monitor and leverage **international research data sharing**.

The NAGIM Horizon activities for *Infrastructure* show a clear focus on:

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

Further activities, under the NAGIM proposed roadmap, for additional areas of Governance, Medical Genomics and Genomics Research are listed below. These highlight the importance of progressing national data governance, clinical data sharing, phenotype capture, national consent mechanisms and national data sharing agreements, as part of an operational national genomics ecosystem.

Table B3. Extract from the NAGIM Blueprint 'roadmap for implementation'

Area	NAGIM Horizon 1 Proposed Roadmap Activities
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.

APPENDIX C – The National Health Genomics Policy Framework

The National Health Genomics Policy Framework **(NHGPF)** was published by the Federal Department of Health in 2018. This set the direction for a nationally coordinated approach to genomics in Australia.

The aim of the National Framework was to avoid duplication of effort and leverage current activities, to drive improvements in health outcomes for Australians and provide a pathway to personalised health care.

'Data: the responsible collection, storage and management of genomic data' - was one of five strategic priorities in the National Framework (Person-Centred Approach, Workforce, Financing, Services, Data).

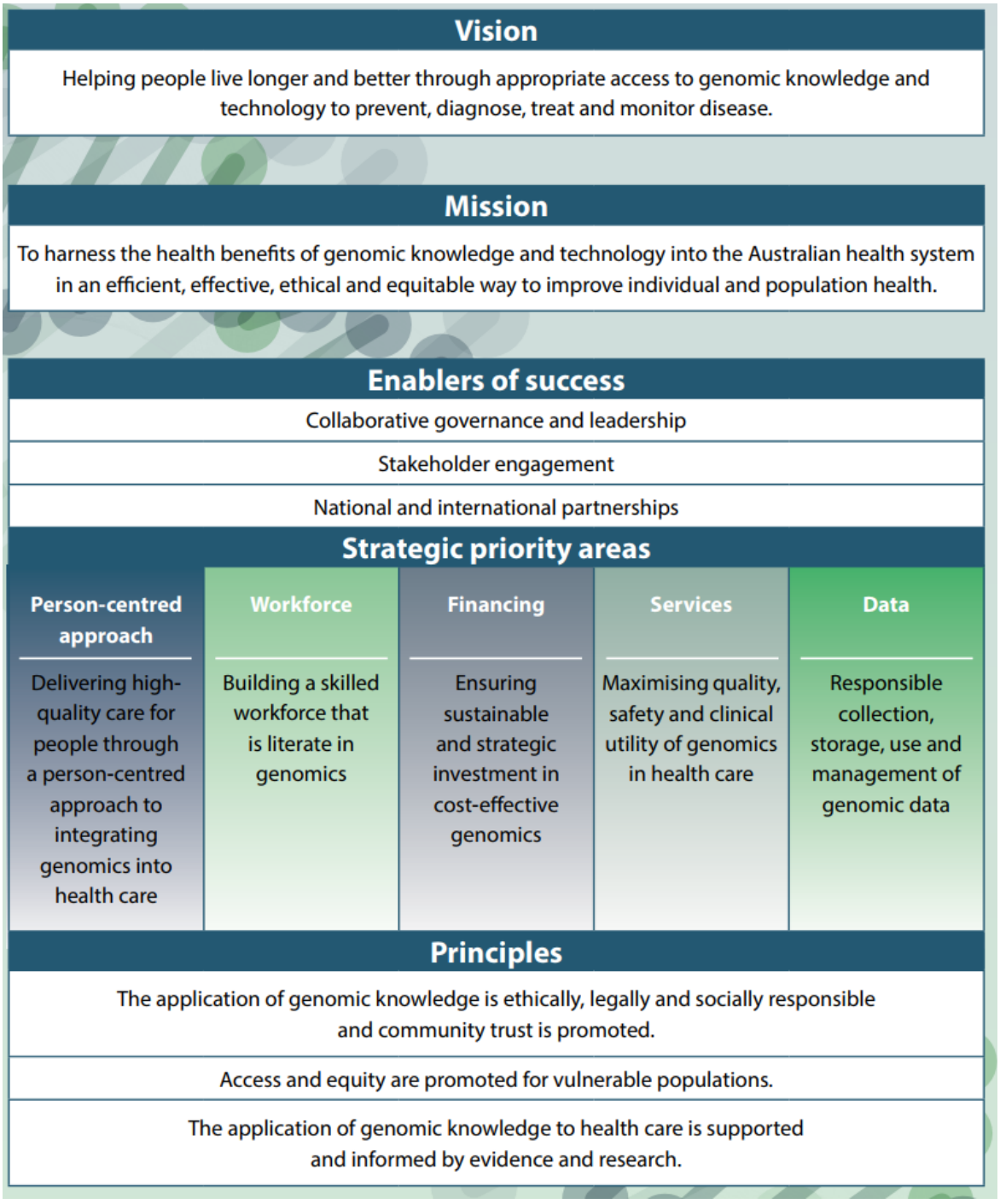


Figure C1. National Health Genomics Policy Framework

NHGPF Implementation Plan

Under the National Framework's associated Implementation Plan, a series of national implementation action items were published for the priority area of **'Data'**. These were:

Action 19: Develop a national genomic **data governance framework** that provides for appropriate decision-making for governments and aligns with international frameworks.

Action 20:
A: Adopt international best practice standards on **cybersecurity and privacy standards** for genomic data systems and data sharing, across all levels of the health system, including consideration of vulnerable populations.
B: Consider the national adoption of appropriate **international standards** on (but not limited to) phenotypes, disease classification systems, and pathogenic variants.

Action 21:
A: Leverage opportunities for integration of individual genomic information with **electronic health records** (including, but not limited to, My Health Record) in ways that maintain public trust.
B: Explore opportunities to capture and integrate **population genomic information** to inform health care decisions, research and policies.

Action 22: Through consultation and engagement, develop **information resources** tailored to the general population and vulnerable groups, in the community on the implications and benefits of genomic data sharing to build community trust in the delivery of health care and for secondary purposes such as research.

Action 23: Build on existing work to develop a **national proof of concept for data sharing** across IT systems in different health care and research settings (such as pathology laboratories, hospitals, registries and research institutions).

APPENDIX D – The Genomics Health Futures Mission Information ICT Recommendations

Summary

The Executive Advisory Committee to the Genomics Health Futures Mission (GHFM) developed its Operational Plan 2018 with recommendations on data provided by its Information and Communications Technology (ICT) subcommittee[3]. These GHFM ICT recommendations emphasised the need for standards-based approaches, interoperability, international standards, and improved cybersecurity such as "model to the data" approaches. An ICT survey for the GHFM identified cloud capabilities as a key priority in the genomics landscape. The GHFM ICT recommendations further described three key users, (patients, clinicians, researchers) whose requirements will need to be met by the future ICT infrastructure.

Vision for Genomics in Australia

In considering ICT requirements, the GHFM Operational Plan notes its vision for genomic data sharing in Australia as requiring a full data steward role with:

- Provision of a **single access point** for use and reuse of (potentially federated) genomic and phenotypic data
- Connecting to other **domestic and international datasets**
- Establishment of **standard protocols for collecting, storing and accessing**; security, privacy, and consumer data rights in accordance with Australian law and research ethics obligations.

Interoperability

As with the NAGIM Blueprint, the GHFM Operational Plan emphasises that ICT for genomics needs to consider Australia's **National Digital Health Strategy (NDHS) and a national interoperability framework** – coordinated data services, seamless access and control by citizens, establishing national digital infrastructure to support digital health initiatives, such as MHR. The Operational Plan indicates that the NDHS will be part of the ecosystem for any future genomics capability and leveraging national infrastructure will need to be considered.

The Operational Plan suggests that the GHFM objectives will likely be achieved with nationally accessible data that allows **flexible access**. This could be through dedicated physical infrastructure or cloud.

Standards

The GHFM Operational Plan for ICT also identifies standards and regulatory compliance as important, for confidence and speed of adoption, public trust, data portability, national/international ecosystem interoperability, efficiency, and maximising impact. Designing compatibility with international standards, with full compliance with Australian legislations and regulations, compliance with international regulations, and interoperability with other systems and international collaboration.

The GHFM Operational Plan noted its commitment to national interoperability standards under the National Digital Health Strategy, and that future ICT solutions for the GHFM need to interface with interoperability standards and frameworks. Further, it expects future proposed ICT solutions to detail how they will be compliant or enable:

- Industry and **international standards for genomic data storage** – particularly those of Global Alliance for Genomics and Health (GA4GH)
- Current Australian and **international standards for clinical and medical terminology**, including SNOMED, ICD11, HL7, and Australian Medicines terminology.

Key User Requirements

The GHFM Operational Plan for ICT identified three key user types whose needs and requirements will require consideration: **patients, clinicians and researchers**.

[3] Used with permission from the Australian Medical Research Advisory Board

Table D1. Key national infrastructure user requirements

Infrastructure Users	User Requirements*
Patients	• accurate and timely testing • clinical genomic data stored in accordance with regulatory requirements • research test data stored in accordance with research protocols • consent for data use • privacy, security, information about data usage
Clinicians	• clinical genomic testing information reported to aid decision making • data maintained and stored in accord with regulatory requirements • data reused only where consent obtained and clinically indicated
Researchers	• genotype and phenotype data that can be analysed, stored and shared • data accessible in an environment with sufficient compute that can support multiple analysis tools and techniques • genotype and phenotype data generated through multiple research projects (including clinical trials) and clinical activity made available for research • research can be executed collaboratively with multiple national and international researchers • data captured from multiple sources (including from new studies or samples, international and national research, clinical datasets) • data quality, interoperability, and standards compliance for optimum reuse • outcome: perform research using data across a range of applications – rare disease, oncology, genomic variation, etc.

*summarised

Cybersecurity

The Operational Plan indicates it expects emerging technologies for data security to be explored (including homomorphic encryption, secure multiparty computation, zero knowledge proofs, and secure enclaves) which enable "**sharing without access**" and "**model to data**" paradigms, which are changing the traditional ways in managing security and privacy risks for biomedical data sharing.

Priorities in the Australian genomics landscape

A survey conducted for the GHFM and reported in the Operational Plan, from current clinical and research genomics stakeholders in Australia, identified potential gaps and underlying priorities for change in genomics technology.

The GHFM survey (2018) reported that most ICT environments for genomics at the time were on premises, with less than 30% using cloud-based storage. However, the majority of respondents identified scope for change and improvement in general within ICT for genomics. A similar survey (conducted by Australian Genomics, 2020) found that 83% of the responding Australian genomic data infrastructures have either cloud or hybrid cloud/non-cloud infrastructure. As both surveys attempted to capture most significant genomic infrastructures in practice, it is likely this reflects an increased uptake of cloud services, from 2018 to 2020.

The identified priorities of respondents of the GHFM ICT survey included access to cloud-based computing, cloud-based storage, data sharing capacity, analytics and data curation (*Table D2*).

Table D2. GHFM Operational Plan ICT Priority Areas from survey respondents

Priority	Area	Definition
1	Storage, access and sharing capacity	Storage availability, technology to support data sharing, and to control access to shared data
2	Analysis and informatics	Availability of specific applications and capabilities to execute analysis on genomic data
3	Data curation	Applications and standard datasets to support data integration and management
4	Cloud computing resources	Cloud compute capacity for genomics data
5	ICT networking and data transfer	Enhanced WAN/LAN capacity
6	Cloud storage	Cloud storage capacity for genomics data
7	Sequencing devices	Capacity/additional sequencing devices
8	Local storage	Local/on premises storage for genomics data
9	On premises computing resources	Local/on premises computing (e.g. HPC) for genomics data

APPENDIX E – The Australian Genomics Infrastructure Capabilities Reports

Australian Genomics completed genomic data infrastructure reports based on Australian (domestic report) and global (international report) genomics initiatives and data infrastructures, based on surveys conducted in 2020.

Domestic Report.
Responses from 17 Australian organisations managing genomic research and clinical data infrastructures, including university and medical research institutes, diagnostic testing laboratories, translational research centres or programs, and data service providers.

International Report.
Responses from 17 large-scale genomic initiatives internationally, representing North and South America, Europe, Africa and Australasia. Infrastructures included large-scale national precision medicine initiatives, research cohorts, service-based platforms for data storage and data analysis, and variant databases.

Key findings from the Domestic and International reports include:

- **Federated infrastructures:** Transitioning to federated infrastructures is a key next theme for international initiatives;
- **Cloud:** A high proportion of domestic (83%) and international (54%) infrastructures currently use either cloud or hybrid cloud/non-cloud infrastructures;
- **Data Sharing:** Most of the international initiatives surveyed (81%) currently support external data sharing. In contrast, Australian infrastructures were not typically engaging in external data sharing (31%), citing governance challenges among key limitations;
- **Standards:** Most of the international initiatives surveyed were adopting one or more forms of standardised terminologies. Domestically, few infrastructures surveyed were collecting or storing clinical information in standardised terms.

Operational and resourcing comparisons

- **FTE allocations:**
  - Large-scale international initiatives are operating on 15-50 FTE;
  - All domestic infrastructures are operating with <14 FTE. Many operating with <3 FTE.
- **Operating costs:**
  - All international initiatives operating costs for all initiatives were funded by government and national research funds;
  - Domestic infrastructure operating costs covered by various means, including internal organisation operational funds, state health departments, cost recovery from data owners and grants.
- **Data capacities:**
  - International large-scale initiatives managing 5,000TB or more, with a quarter of international initiatives expecting an increase in 1000TB or more per year;
  - Domestic large-scale initiatives are managing 5,000TB or more, with no domestic infrastructures expecting more than 1000TB a year in increased data.

For further details, see the full International and Domestic Reports:

https://www.australiangenomics.org.au/wp-content/uploads/2021/06/NAGIM_Domestic-Data-Survey-Report_October-2020.pdf

https://www.australiangenomics.org.au/wp-content/uploads/2021/06/NAGIM_International-Data-Survey-Report_October-2020.pdf

APPENDIX F – NAGIM Prototype Program

NAGIM Prototyping

To inform the NAGIM implementation recommendations, prototypes for genomic data infrastructure were launched in 2021 to develop and encourage adoption of scalable, interoperable and extensible approaches to the collection, storage and use of genomic data in Australia.

Data infrastructure stakeholders nationally were invited to participate in prototype development, in an open call, leveraging existing capabilities/funding. The goal of the prototyping was to address priority infrastructure elements from the NAGIM Blueprint, to **identify the best combination of components to serve as the basis for long-term national research infrastructure**.

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 brief EOI process was undertaken for teams to submit a proposal and confirm ability to:

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

All teams submitting an EOI were accepted for participation: some unfunded, others requesting Australian Genomics funding to support involvement.

Prototype Requirements

The prototyping phase was conducted for research infrastructure, as the first phase, with clinically-focused activities to be progressed in a separate, but complementary phase.

Prototyping teams were expected to address:

- NAGIM Blueprint priority areas for infrastructure
- NAGIM Pilot Technical Guidelines

These elements are outlined below.

NAGIM Blueprint Priority Areas for Infrastructure

Prototypes were required to address the following priority areas for infrastructure identified by the NAGIM Blueprint:

- Federated frameworks
- Standards-based approaches
- Interoperability across systems
- Alignment with international data sharing initiatives
- Cloud-based or hybrid solutions

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

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 datasets 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 F1*) which outlined initial components required for a national genomics infrastructure with an emphasis on phase 1: the research environment. 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.

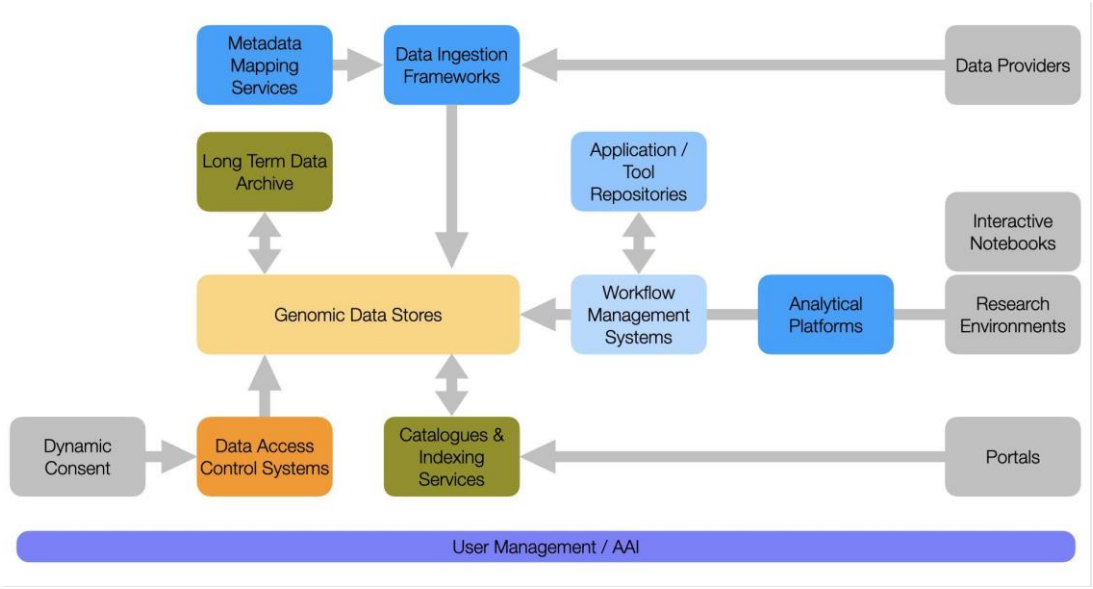


Figure F1. Components of a national genomics infrastructure

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 *(e.g., Cromwell, Toil, Airflow, etc.)*.
- **Tool Repositories:** Systems for storing and sharing reusable analytical tools or workflows used by the workflow management systems *(e.g., 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 harmonise metadata accompanying genomic data submissions (*e.g., OntoServer*) and proposed metadata definitions and schemas.
- **Genomic Data Stores:** Object stores *(e.g., 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.

Prototyping Timeline and Outcomes

Table F2. NAGIM Prototyping Timeline, Deliverables, and Responsible Parties

July 2021	***Open Call for Participation*** *(Australian Genomics)*
Aug 2021	***Prototype Development Commenced*** *(Prototype Teams)*
Dec 2021	***Prototype Infrastructure and Documentation Complete*** *(Prototype Teams)*
Jan 2022	***Evaluation of Prototypes and Future Recommendations Submitted*** *(International Expert Panel)*
Apr 2022	***Preliminary NAGIM Recommendations Delivered*** *(Australian Genomics)*

NAGIM Prototype Submissions

Twelve proposals were initially received, and a *NAGIM Prototype Community of Practice* was established with all 12 teams, that aimed to **build and integrate** components of the proposed NAGIM research ecosystem (Prototypes and lead organisations are summarised in *Table F3* below). This engagement occurred as both formal (monthly meetings ~40 attendees per meeting) and informal (Slack channel, with 54 members) interactions.

Nine prototypes were completed over five months (*Figure F4*).

An **international expert review panel** was convened to independently advise and evaluate the NAGIM Implementation prototypes, and establish the evaluation framework. International panel assessments of the NAGIM prototypes were completed in January 2022, together with their future recommendations for progressing NAGIM.

Prototype Technical Reports

Teams submitted the following items as their final prototype and documentation:

- Prototype methodology, including how the NAGIM infrastructure priorities were addressed;
- Description of interoperability – API documentation, standards used, integrations;
- Architectural diagrams;
- Demonstration videos;
- Gaps and limitations of the prototype; and
- Test environment for members of the international review panel to access (optional).

Prototype Team Reflections and Recommendations

In the final report, the prototyping teams were also given 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.

Final prototype submissions are available by request from Australian Genomics.

Table F3. Summary of NAGIM prototypes

No.	Name	Project Lead	Prototype Aim
End-to-End Genomic Data Platforms			
1	**Human Genomes Platform Project**	**Assoc. Prof Bernie Pope** Australian Biocommons	Prototypes 1 and 2 have merged into: A federated network of services using standards and a unified authentication / authorisation framework, demonstrating the flow of genomic data, clinical metadata and data access control.
2	**UMCCR Gen3 AWS Platform**	**Assoc. Prof Oliver Hofmann** University of Melbourne Centre for Cancer Research	
8	**NCI HPC Analytics**	**Dr Ben Evans** NCI	Analytics platform for large scale genomics data processing and analysis of managed data stores at the NCI. Covers analytical platforms, workflow management, and accessing genomic datasets.
10	**Garvan Terra Google Platform**	**Prof Daniel MacArthur** Centre for Population Genomics, Garvan Institute & MCRI **Assoc Prof Sarah Kummerfeld** Kinghorn Centre for Clinical Genomics, Garvan Institute	Massively scalable genomic analysis platform.
11	**Max Kelsen Terra Google Platform**	**Nicholas Therkelsen-Terry** Max Kelsen	Cloud native genomics & healthcare data platform, enabling inter and intra organization collaboration, data sharing and analysis.
Specific Tools			
3	**VariantSpark Analytics Platform**	**Dr Natalie Twine** Transformational Bioinformatics CSIRO	A cloud/HPC-based analytics module for genome-phenome associations, with capability to target large- scale datasets and integrate into Australia's genomics analytics platforms.
4	**Serverless Beacon**	**Dr Laurence Wilson** Transformational Bioinformatics CSIRO	An implementation of the GA4GH Beacon protocol, aiming to scalably and cost-effectively support queries and responses across future mega-biobanks.
5	**Trustless Data Access Control**	**Dr Denis Bauer** Transformational Bioinformatics CSIRO	Self Sovereign models of consent and data access control to deliver better engagement, security and trust for medical research participants, researchers and organisations.
6	**Standardised Clinical Phenotype Tools**	**Dr Alejandro Metke** Australian e-Health Research Centre CSIRO	Tools to standardise clinical phenotype data, which is of fundamental importance in research.
7	**Cloud-native WGS Tools**	**Mr John Pearson** Genome Informatics QIMR Berghofer	Modify bioinformatics tools to add support for directly accessing cloud-native data storage.
9	GeneTrustee	Dr Warren Kaplan - withdrawn Garvan Institute	Withdrawn
12	VariantGrid	David Lawrence	Withdrawn

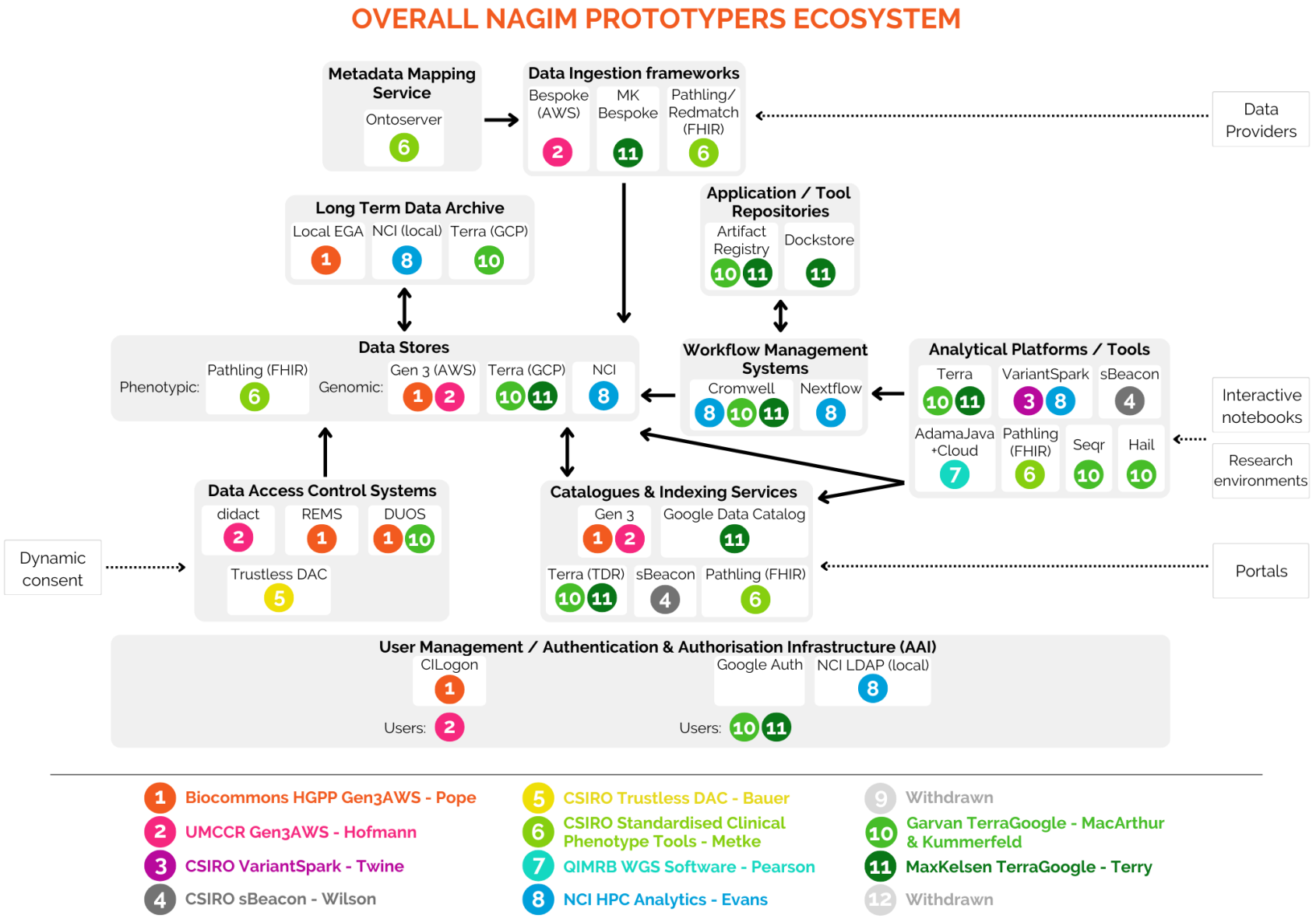


Figure F4. NAGIM Prototypers' Systems and Overall Ecosystem

NAGIM Prototype Evaluation

NAGIM International Panel

A group of eight international experts were assembled to advise and evaluate the NAGIM prototype phase. The panel had significant expertise from large-scale genomic sequencing initiatives, national precision medicine programs, international standards in clinical informatics and interoperability in healthcare.

International Panel for NAGIM Prototyping Reviews and Recommendations

1. Augusto Rendon — *Genomics England*
2. Christina Yung — *Indoc Research & Ontario Institute for Cancer Research*
3. Daryl Waggot — *Genome Canada*
4. Eric Banks — *Broad Institute of MIT & Harvard*
5. Grant Wood — *Intermountain Clinical Genetics & HL7*
6. Jonathan Dursi — *Canadian Distributed Infrastructure for Genomics (CanDig)*
7. Mar Gonzalez Porta — *Precision Health Research Singapore (PRECISE)*
8. Robert Freimuth — *Mayo Clinic*

Evaluation Framework

The international advisory panel convened three times to review the NAGIM prototyping phase, shape the evaluation framework and conduct interviews with the prototype teams.

The panel noted extreme diversity in the prototypes – from a single-purpose or single component tool, through to comprehensive, end-to-end genomic data platforms. The panel noted the challenges inherent in assessing prototypes with this level of diversity, for applying systematic metrics to all prototypes, and to infrastructures that are not yet mature or 'in production'. The evaluation framework was built in consideration of this diversity.

The key metrics assessed were those identified by the international panel as three '*critical pillars*' required for NAGIM-enabled infrastructure:

1. **Interoperability:** able to interact with other systems to seamlessly exchange and make use of information;
2. **Scalability**: able to accommodate growth and manage / process large-scale datasets; and
3. **Extensibility**: able to accommodate new additions and expand functionality.

Additional elements of the evaluation framework are described below.

Core Evaluation Elements

The technical evaluation, as determined by the NAGIM international panel, included:

Individual and ecosystem appraisal

i) Individual evaluations of prototypes
ii) Appraisal of the collective prototyping ecosystem, as a whole

General metrics

i) Significance
ii) Impact for NAGIM
iii) Innovation
iv) Consideration of barriers

Key metrics

i) Interoperability (standards use, and systems interoperability)
ii) Scalability
iii) Extensibility

Additional considerations

Including but not limited to:

i) Intersections with clinical systems and clinical data
ii) Maturity of prototype components
iii) Interaction of components *across different prototypes*
iv) User experience

Evaluation Outcomes and Panel Recommendations

The international evaluators 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."
>
> **NAGIM International Reviewer**

A summary of key findings and recommendations from the International NAGIM panel are provided below.

NAGIM Prototypes

Nine national prototype teams completed the Australian Genomics NAGIM prototyping phase which ran for five months (Aug-Dec 2021). Eight of the final prototypes were cloud based and one was based on HPC. Prototypes were classified as either 'End-to-End' (ETE) genomic data platforms or specific tools for analysis, queries, consent or clinical information.

Key Prototype Outcomes

The NAGIM reviewers' panel comprised eight leading international genomics experts. The panel members reviewed all prototypes and identified **the ETE platforms as the foundational initial priority for progressing NAGIM –** with the specific tools considered valuable components for factoring into a NAGIM platform(s) in a secondary phase**.**

Key criteria for evaluation focused on demonstrated interoperability (standards use; and external integrations), scalability, and extensibility of the infrastructure. The prototypes significance, overall impact for NAGIM, innovativeness, and consideration of barriers were also assessed.

The Panel's evaluations indicated that two ETE platform prototypes best satisfied the key NAGIM criteria, each with different features:

- Prototype 1/2: A demonstration of federated data processes using the University of Chicago opensource Gen3 data platform on Amazon Web Services, prototyped by the Australian BioCommons-UMCCR Human Genome Platform Project team - focused on federated services, standards and data access. It was assessed as offering a custom-developed, locally-made solution that can be tailored to Australian needs – at the cost of high development, operations and maintenance.
- Prototype 10: A demonstration of a scalable data analysis platform using the Broad Institute opensource Terra platform on Google Cloud Platform, prototyped by the Garvan Institute team – focused on massively scalable genomic analysis capability. It was assessed as offering a 'ready-to-go' platform that could be in production relatively quickly – at the cost of being currently tied to one cloud vendor.

NAGIM Prototypers Overall Ecosystem

Strengths of the overall prototyped ecosystem, identified by the Panel, included:

- ✓ Prototyping an Australian-wide authentication mechanism (HGPP prototype) – considered a significant step forward for NAGIM;
- ✓ Instances where integrations *across* multiple prototypes were achieved;
- ✓ Collaborative efforts across prototype teams;
- ✓ Use of mature components already in production internationally for national initiatives.

Limitations and gaps identified:

- Prototypes focused on technology – not data or processes, indicating a big challenge for progressing NAGIM in this ecosystem will be data harmonisation;
- Data ingestion was underexplored; Simple, tightly controlled processes are needed for high quality data;
- Role of HPC in the ecosystem was not clear;
- Potential gaps included: analytic tools, international integration, nongenomic data.

Future key considerations identified by the Panel, that were not in scope for this prototyping phase, but noted as critical for a future NAGIM ecosystem, included addressing: clinical systems, governance and ELSI (ethical, legal and social issues), management of Indigenous data, security, and nonhuman genomic data.

Summary of Panelists' General Recommendations

- **Prioritise building the foundation** of the genomics data ecosystem, in order of:
  1. Shared national services
  2. Multi-institute data flows and deployment of standards
  3. Bridging individual institutions into the ecosystem
  4. Innovation of new methods
- **Key parallel efforts for the next phase.**
  1. Collect specific use cases – to define requirements around types of data and analytics to support.
  2. Design the architectural backbone of NAGIM: use the prototypes, knowledge of similar efforts elsewhere, and the defined principles (standardisation, interoperability, scalability, extensibility).
  3. Identify and analyse the systems that will serve as data sources.
  4. Identify the data and messaging standards for each type of data that will be exchanged.
  5. Assemble a group to provide review and oversight.
- **Timelines.** Determine when it is intended to "go live" for NAGIM to be able support many users and large amounts of data. A short timeline favours [Terra], a longer more staged timeline supports [HGP].
- **Data collection.** Assess the level of support in healthcare or research institutes for depositing their data into a NAGIM platform; And the nature of existing relationships of platform teams with data custodians.
- **Science/Engineering Balance.** Identify what investment would be usefully spent on technological development and capacity building, and developing Australian leadership there, in addition to on the medical science from the genomics itself.
- **'Best of breed' approach.** Use solutions with components that are mature and "production ready" or already in production use by other national initiatives. Consider prioritising platforms or components that don't require assembling multiple teams.
- **Functional requirements.** Contextualise the functional requirements around the *larger issues*, such as public cloud provider strategy, long term funding, additional legislation required to facilitate interoperability.
- **Operational management.** Identify who will maintain these platforms operationally.
- **Innovation.** Fund innovation for new methods separately from research data management within NAGIM.
- **Landscape and Prototype Mapping.**
  - Create a national landscape map of the current health data infrastructure, i.e., including region, roles, services, users, etc. to provide high-level context for the diverse prototypes in NAGIM and how they contribute to the overall Australian genomics data ecosystem. This will provide a roadmap of next steps.
  - Classify the prototypes into themes relative to i) their role in the ecosystem i.e. shared national services, vs bridges to institutional research programs, vs innovative methods; and ii) their specific requirements i.e. for funding, governance, sustainability; and iii) to their role in building the foundations of the ecosystem.
- **Real world piloting.** Use the current prototypes to deliver real-world projects – this is the best way to learn.
- **Industry partners.** Leverage industry partners, with experience in operational efficiency and in navigating the regulatory landscape. This will be key when transitioning to more clinical applications.
- **Private sector:** Separate potential contributions of (and provide opportunities for) private sector vs academic research groups to participate in the NAGIM ecosystem e.g. by separating innovation and infrastructure, with oversight and contracts coordinated by national organisations.
- **Governance structure.** Define the principles for a governance structure from the outset, that is inclusive (i.e. national organisations, partners, institutes, researchers), driven by the needs of users and facilitates progress and adoption. I.e. defining the direction and nature of the NAGIM infrastructure and shared services.
- **Governance and policies.** Ensure governance and policies keep up with and do not impede the technical infrastructure, as is often the case. Ensure strong support from the social, privacy and legal perspectives.

- **Standards and harmonisation.** Establish the tools, standards, and a governance process that will be required to support the harmonisation of data collected nationwide by many different people, workflows, and systems. This includes use of data and/or messaging standards (e.g., FHIR) as well as efforts to ensure those standards are implemented and used consistently and correctly, with conformance to a common standard to enable accurate retrieval and analysis within research studies.
- **Data Sharing.** Ensure data sharing is standardised and well-structured to ensure high level of data completeness, quality, and interoperable format.
- **Clinical.** Ensure consideration of genomic medicine. Many datasets originate in the clinic, and are part of a wider collection of clinical data points around individuals.
- **Non-human genomic data.** In the longer term, partner with other genomic stakeholders to develop shared infrastructure for genomics research, to avoid duplication or parallel efforts.

The full evaluation report is available from Australian Genomics.

APPENDIX G – NAGIM Infrastructure Ecosystems

Proposed conceptual federated NAGIM ecosystem maps are presented below, for both research and clinical NAGIM ecosystems. These provide a map of aligned, interoperable and integrated components for progressing towards a federated approach.

The NAGIM research infrastructure ecosystem (*Figure G1*) was developed to guide the NAGIM prototyping phase (*Appendix F*), and align infrastructure developers towards common, interoperable and scalable components, and encouraging the integration of systems across a community of practice for research data.

An equivalent NAGIM clinical infrastructure ecosystem has also been developed as an exemplar (*Figure G2*), that represents elements of a local clinical genomics system, jurisdictional healthcare systems, and national data infrastructure. This clinical ecosystem below, may also be used to guide the development and integration of nationally aligned, interoperable clinical genomics systems and integrated national components. However, it is intended as an initial exemplar only – to guide commencement of cross-organisational, jurisdictional, and federal health system discussions. The NAGIM clinical ecosystem that is to be progressed, should be developed and consolidated with clinical infrastructure stakeholders, at the initiation of the NAGIM Clinical infrastructure workstream (WS3).

Schematic of a Federated NAGIM Research Ecosystem

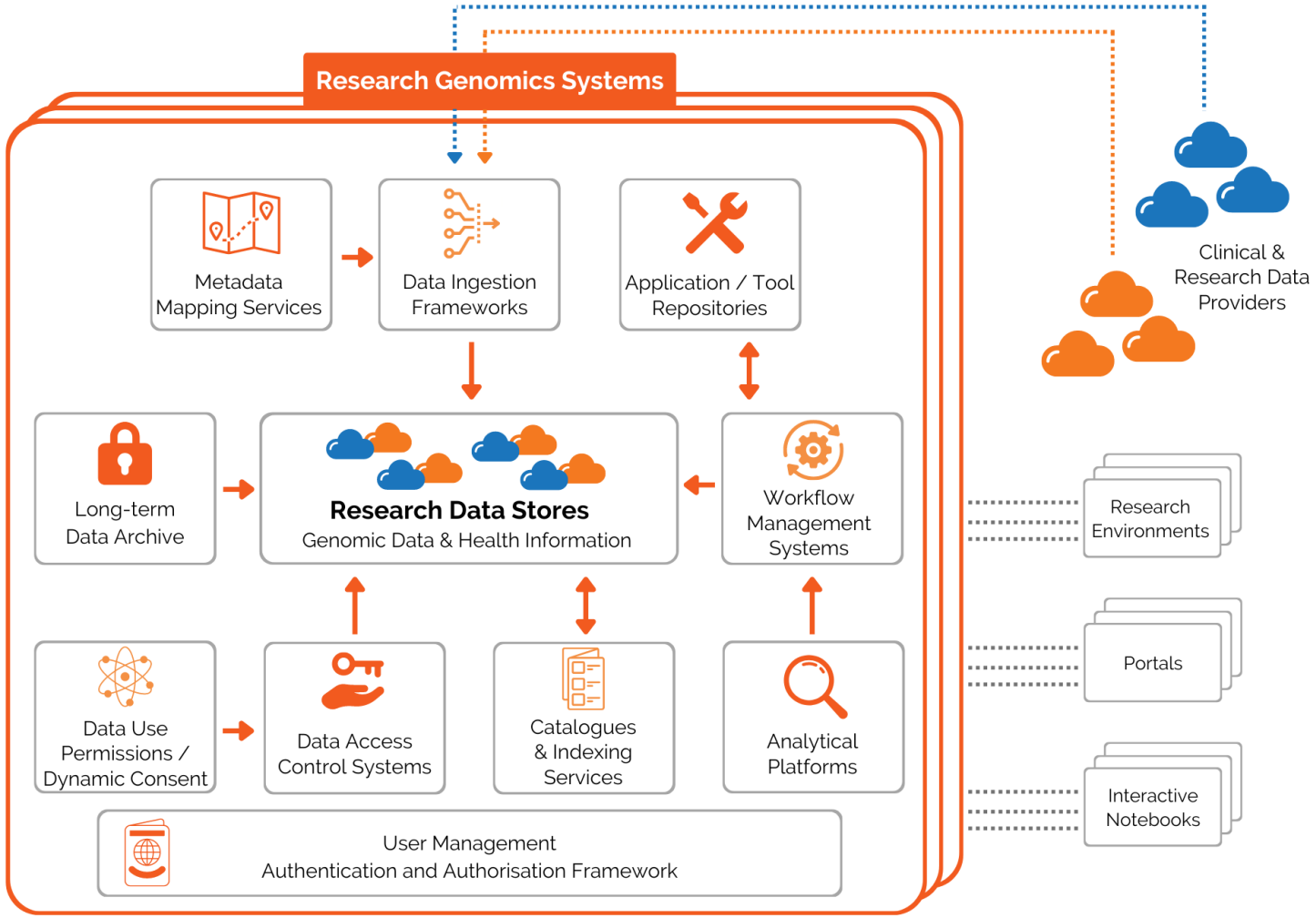


Figure G1. Schematic of components for a proposed federated NAGIM research ecosystem

Schematic of a Federated NAGIM Clinical Ecosystem

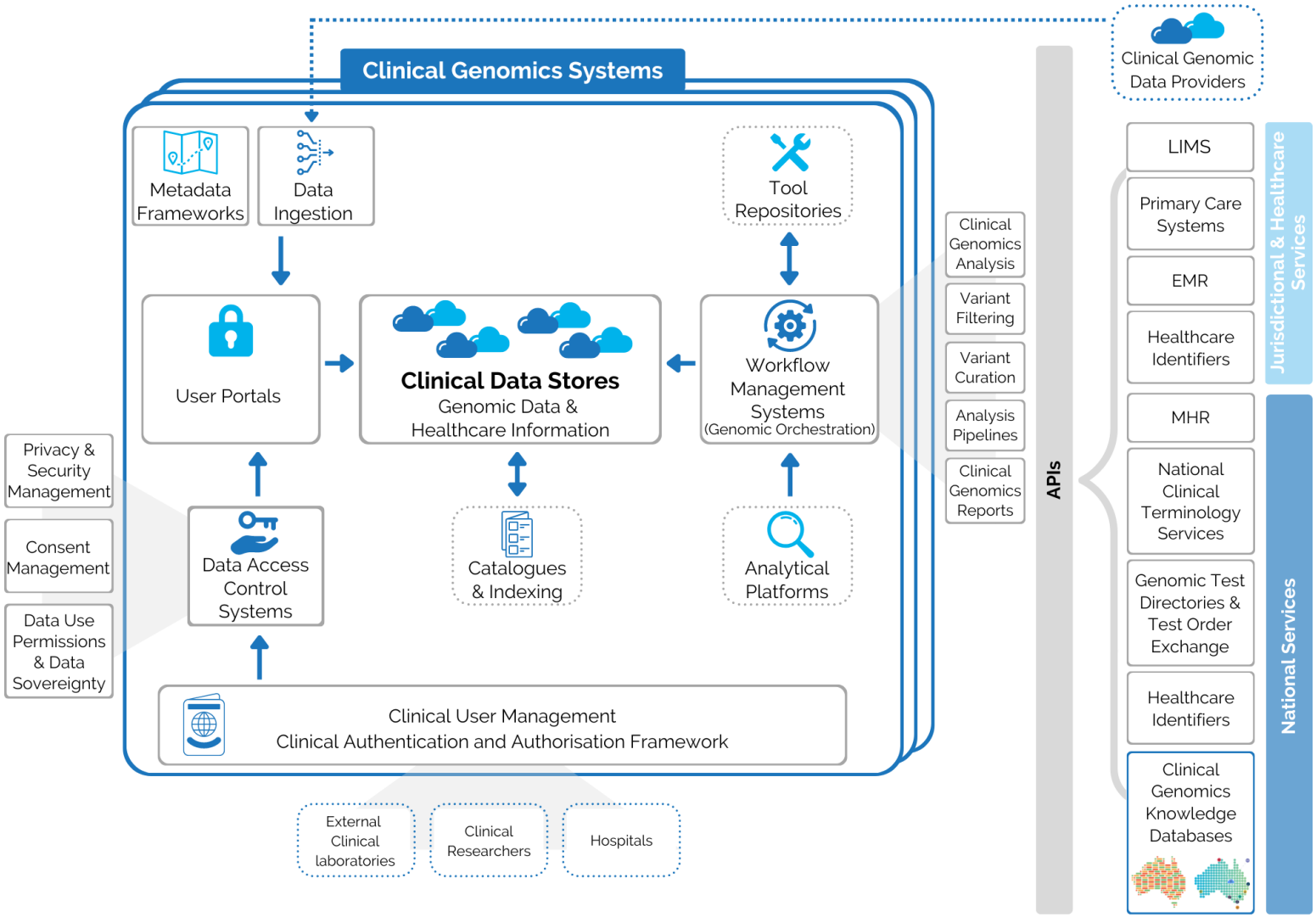


Figure G2. Schematic of components for a proposed federated NAGIM Clinical Ecosystem

APPENDIX H – National Consultations on NAGIM Implementation

Preliminary Recommendations Survey 2022

National consultation on the initial preliminary Implementation Recommendations were undertaken from May to June 2022 via a public online survey. The preliminary recommendations and consultation form were announced publicly via the Australian Genomics newsletter and website and circulated broadly to genomics stakeholders, including data, research, clinical, government, policy, indigenous, industry and community groups. Representative organisations and peak bodies were requested to distribute the consultation invitation throughout their respective networks. Entities or representatives that submitted a response to the national consultation as listed below.

2022 Respondents to the National Consultation Online Survey

Federal Government and Associated Agencies

- Australian Digital Health Agency (ADHA) **
- Australian Government Department of Health **

Clinical and Healthcare Delivery

- Genetic Health Queensland
- NSW Health Genomics IT & Infrastructure Committee **
- Royal Brisbane Women's Hospital
- Sonic Healthcare
- Victorian Clinical Genetics Services **

Medical Alliances and Associations

- HGSA Education, Ethics and Social Issues Committee **
- Melbourne Genomics Health Alliance **
- Royal Australasian College of Physicians (RACP) **

Indigenous Health

- Aboriginal Health Council of SA **
- CONNECT & National Indigenous Genomics Network **
- National Centre for Indigenous Genomics (NCIG) **

Industry

- Illumina **
- Industry Genomics Network Alliance (InGeNA)
- Lifebit Biotech Limited **
- TrakGene
- 23 Strands **

Community

- QLD Genomics Community Advisory Group

Research

- ANU College of Health and Medicine **
- Garvan Institute **
- Murdoch Children's Research Institute
- QIMR Berghofer **
- University of Adelaide
- University of Sydney
- University of Melbourne **

Computing and Infrastructure

- Australian BioCommons **
- National Computational Infrastructure Australia (NCI) **

** *denotes submissions received on behalf of the organisation and / or an aggregated submission following internal consultations by the submitting organisation.*

Additional Consultations and Engagement

- ARDC, Australian Research Data Commons
- NPAAC, National Pathology Accreditation Advisory Council
- RCPA, Royal College of Pathologists of Australasia
- Australian Commission on Safety & Quality in Healthcare
- NAGIM Blueprint Leadership (Health Translation QLD, Clinical Excellence QLD)
- Australian Genomics Community Advisory Group
- Australian Genomics National Networks:
    - Policy
    - Clinical, Diagnostic & Research
    - Data
- Genomics Technology summits
- State health and digital health representatives

Diagnostic Laboratories Engagement

Diagnostic laboratories participating in the Australian Genomics Shariant platform, were briefed on the NAGIM implementation project, to 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.

Shariant

The Australian Genomics Shariant platform facilitates sharing of clinical variant classifications and curation evidence into a centralised platform via a standardised data collection process. Shariant represents a successful, functional exemplar of a national approach to clinical NAGIM implementations.

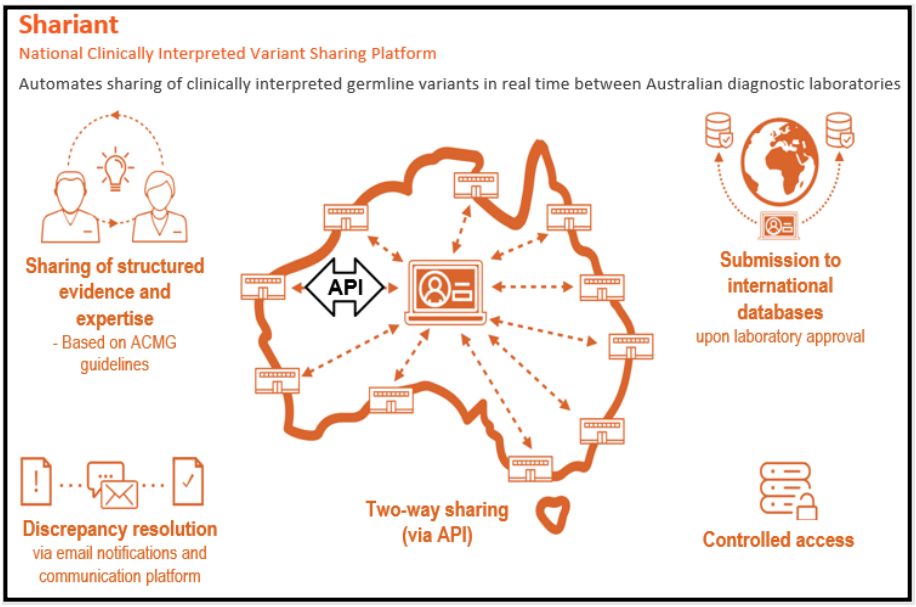


Figure H1. The Australian Genomics Shariant Platform

At the time of engagement for progressing NAGIM (October 2021), the Shariant users' network, included ~60 members across 19 laboratories nationally. This network continues to expand with the inclusion of additional laboratories nationally, including private labs.

Table H2. *Shariant User's Network in 2021 at the time of consultation*

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

Shariant Laboratories Consultation 2021

In addition to broad engagement with the Shariant User group (above), a survey addressing priority areas from the NAGIM Blueprint 'Horizons' roadmap, for ***Genomic Medicine*** and ***Infrastructure***, was distributed to Shariant users' network in early October 2021. 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.

Formal survey responses were received from representatives across eight organisations:

- Children's Hospital Westmead (CHW), NSW
- Melbourne Genomics (GenoVic), VIC
- NSW Health Pathology, NSW
- Pathology Queensland (PQ), QLD
- Royal Prince Alfred Hospital (RPA), NSW
- Royal Melbourne Hospital (RMH), VIC
- SA Pathology, SA
- Victorian Clinical Genetics Service (VCGS), VIC

Shariant Consultation Themes

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

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** (including phenotype, pedigree, metadata, consent)
- Progress infrastructure and processes that would enable diagnostic labs to **access data nationally for** their primary diagnostic activities:
    - Accessing additional (uncurated) variants beyond Shariant;
    - Genotype-phenotype data from individual patients to aid interpretation; and
    - 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

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.

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

- 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;

Data and Infrastructure Pilots

- 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.

These priorities, opportunities and barriers informed the development and content of the preliminary recommendations.

Stakeholder Feedback on NAGIM Preliminary Recommendations

Stakeholders' Key Considerations for Progressing NAGIM

Theme 1

PARTICIPATION & ENGAGEMENT

Equity and Breadth of Participation

- Equity of access across states and territories
- Equity of resourcing/funding across states and territories
- National participation
- Open and broad participation in workstreams
- Geographical reach
- Industry representation and participation
- Private and public healthcare sector inclusion

Indigenous Community and Expertise

- Sufficient and respectful timeframes and adequate budget allocation for Indigenous consultation
- Support for Indigenous leadership
- Indigenous participation across workstreams
- Processes for prioritising and triaging key governance and sovereignty issues
- Co-development with international models

Community

- Trust
- Transparency
- Meeting public expectations
- Co-design by consumers and personnel; Involvement of independent health consumers
- Communication with consumers and stakeholders
- PPIE approach (Patient Public Involvement and Engagement)
- Diversity of perspectives captured for community engagement
- Early community consultation
- Data donor control over their data and communication about how their data is being used
- Involvement in ELSI considerations
- Consideration of best interests of patients and public

Industry

- Ensuring engagement and partnership with industry
- Leveraging industry experience for:
  - Transitioning from academia led projects to national implementation
  - Complex and time consuming challenges e.g. large-scale data harmonisation
- Engagement with the primary commercial cloud vendors in pre-production planning
- Engagement with pharmaceutical industry, in relation to clinical trial needs
- Inclusion of frameworks for non-government funded (commercial) research

Stakeholders' Key Considerations for Implementation of NAGIM

Theme 2

WORKSTREAM AREAS

Clinical and Digital Health Infrastructures

- Intersections and distinctions between clinical and research infrastructure
- Role of clinical testing and funding on management of federated cloud infrastructure and data access
- Separate needs for data provision vs data use
- Clinical terminologies (SNOMED, LOINC), standards (ISO) and peak bodies
- Leverage, re-use and align with national digital health initiatives (e.g. MHR, national repositories and catalogues, digital health standards, government/ healthcare identifiers)
- Specific consideration of consent, privacy and data encryption for clinical services
- Health technology assessments and transitioning to routine healthcare

Workforce and Education

- Skilling the workforce on data management
- Workforce challenges to scalability
- Workforce education of guidelines for implementing NAGIM
- Communication and adoption of best practice approaches

National User Services

- Promoting user adoption
- Foster / develop Australian human genomics stakeholder community
- Communications with users
- Documentation, training, education and outreach
- Incentives for participation or adoption/use of NAGIM infrastructure

Innovation

- Business development
- Data monitoring, reporting
- Artificial intelligence
- Digital Trust solutions
- Distributed analysis
- Open ecosystem approach for innovation
- Scaling operations

Stakeholders' Key Considerations for Implementation of NAGIM

Theme 3

DATA & INFRASTRUCTURE

Security and Privacy

- Earlier and deeper consideration of security
- Security, privacy, choice and access control
- Progress Secure Research Environments
- Deploy security and privacy solutions in Data Custodians' own environment (vs outsourcing) e.g. Privacy-as-a-Service (PaaS-based solutions)
- Privacy and privacy preserving analytics, queries travelling to the data
- Security considerations & risks of open source software
- Implement the "Five Safes" for data security
- Security audits
- Data de-identification
- Data encryption

Data and Data Standards

- Addressing data quality / ensuring high quality data
- Support for making existing datasets 'FAIR' (Findable, Accessible, Interoperable, Reusable)
- Data standardisation challenges from upstream testing that is difficult to standardise
- Adopt and contribute to international standards, support global interoperability, alignment with international best practice
- Application of industry standards and vocabularies
- Consideration for the management and accessibility of different data types (raw, aggregated, indexing)
- Genomics data standards that are supported by policy
- Data management as key (hardware, data models, IP, costs, networks)

Data Access and Governance

- Consent management
- Data ownership
- Data-donor centricity and control
- Identity and access frameworks that include industry, clinical infrastructures/users and international collaborations
- Inclusion of ELSI and governance principles in the recommended architectural principles
- Accountability and responsibility
- Transparency
- Accessibility of differentially funded clinical data – for research use
- Establishing and maintaining governance over multiple systems

Architectural

- Considerations with federating:
  - Applying separately to access and analytics
  - Specific nature of federated approach in IT context (as agreed models)
  - Distributability of the infrastructure
- ETE system that considers both research and clinical
- Preference for cloud-based infrastructure in the clinical sector/low predicted future reliance on on-premises infrastructure in clinical
- Delivery of a data infrastructure in the pre-production phase that can be used for research/research data
- Differential needs of smaller infrastructures
- Explore Australian strengths in infrastructure development
- Software considerations (as key enabler of the infrastructure; open sourcing)
- Open ecosystem approaches to ensure user customisation

Stakeholders' Key Considerations for Implementation of NAGIM

Theme 4

OPERATIONS & STRATEGY

Strategic

- Requirement for strong leadership and management from a central group
- Program Governance & Coordination (WS1) is critical to the success of the program
- Importance of a competitive tendering process: open competitive funding processes with peer-review
- Importance of transparency and stakeholder engagement
- Funding model for federated data management for genomic tests funded by different mechanisms
- Align to Australian context (regulatory, legal, cultural)
- Engagement with international communities
- Need a business case for national scaling
- Long-term sustainability of the overall infrastructure(s)

Operational

- Independent / external program responsible for auditing auditing and evaluation
- Cost-benefit analysis; consideration of value for money
- Building on and leveraging existing services and infrastructure to reduce costs and mitigate project risk (including research infrastructure, clinical genomics infrastructure, and digital health infrastructure)
- Determine the feasibility and strategy for governance of federated and distributed infrastructures
- Demonstrated experience with federated infrastructure as a tender requirement
- Accountabilities and responsibilities – of data custodians; around breaches; legal agreements etc.
- Determine entity responsible for oversight of the ecosystem and management of contributing infrastructures

General

- Extensibility to nonhuman genomic data
- Consider population genomics
- Consider international suppliers or service providers
- Comprehensive consideration of data governance, ELSI* and best practice principles
- Consideration of competitive pricing models / affordability to support adoption and ongoing usage
- Needs compelling proposition and/or regulatory requirement to attract all stakeholders

Resourcing

- Dedicated funding required for community engagement and Indigenous engagement
- Budget allocations for data providers for integration activities
- Consideration of specific funding for clinical – research infrastructure intersections
- Develop early cost estimations for all workstreams

** ELSI – ethical, legal and social issues*

APPENDIX I – Key Policies, Frameworks and Strategies

It will be important for NAGIM implementation strategies and activities to consider alignment and integration of key jurisdictional, national and international policies and frameworks.

These include, but are not limited to:

National Healthcare and Digital Health Strategies

- National Healthcare Interoperability Plan, Australian Digital Health Agency
- National Digital Health Strategy and Framework for Action, Australian Digital Health Agency

Clinical Standards and Guidelines

- National Pathology Accreditation Advisory Council (NPAAC)
- National Association of Testing Authorities (NATA)
- Standards Australia (e.g., ISO/IEC 23092-1:2020 to ISO/IEC 23092-5:2020; ISO/TS 20428:2017; ISO/TS22692:2020)
- Health Level 7 International (HL7) standards, such as FHIR
- Observational Medical Outcomes Partnership (OMOP) Common Data Model
- Clinical terminologies, including SNOMED Clinical Terms (SNOMED CT), Human Phenotype Ontology (HPO), Logical Observation Identifiers Names and Codes (LOINC), Mondo disease ontology, and International Classification of Diseases (ICD)
- Royal College of Pathologists Australia (RCPA) Standardised Pathology Informatics in Australia (SPIA) Guidelines, information models and terminology reference sets.

State/Territory and National Healthcare Genomics Strategies

- State/Territory genomics health strategies (WA Genomics Strategy 2022-2032; NSW Health Genomics Strategy 2021-2025; SA Clinical Genomics Plan 2022; Genetic and Genomic Healthcare for Victoria 2021; Statewide Genetic Health Queensland Service Plan 2017-2022)
- National Health Genomics Policy Framework and Implementation Plan 2018 - 2021, Australian Government Department of Health and Aged Care

Indigenous Principles and Frameworks

- CARE Principles for Indigenous Data Governance
- Lowitja Institute Indigenous Data Sovereignty Tools and Frameworks
- Guiding Principles: Ensuring Culturally Safe Health Genomics in Partnership with Aboriginal and Torres Strait Islander Peoples, Aboriginal and Torres Strait Islander Advisory Group on Health Genomics (In Development)
- AIATSIS Code of Ethics for Aboriginal and Torres Strait Islander Research & AIATSIS Access and Use Policy

Research

- National Research Infrastructure Roadmap 2021, Australian Government Department of Education
- Management of Data and Information in Research, National Health and Medical Research Council (NHMRC)
- Medical Research Future Fund (MRFF) guidelines and policies
- Australian Research Data Commons (ARDC) national frameworks and strategies, including those of the Health Studies Australian National Data Asset (HeSANDA), and the Research Data Management Framework (in development)

National Data Strategies

- Data Availability and Transparency Act Scheme, Office of the National Data Commissioner (ONDC), Australian Government
- Australian Data Strategy, Australian Government Department of Prime Minister and Cabinet

Data Principles

- Australian Privacy Principles, Australian Government Office of the Australian Information Commissioner
- Five Safes Framework Data Confidentiality Guide, Australian Bureau of Statistics
- FAIR Guiding Principles for scientific data management and reuse, 2016

International Frameworks and Standards*

- Global Alliance for Genomics and Health (GA4GH) frameworks, policies and standards, including the GA4GH Framework for Responsible Sharing of Genomic and Health-related Data, and GA4GH open data standards – including those for consent, data retrieval and file formats
- ISO Standards (e.g., the work in development by ISO/TC 215 SC1 – the Genomics Informatics Steering Committee)

Reports and White Papers

- Genomic data in Australia, Industry Genomics Network Alliance (InGeNA)
- Essentially Ours: Assessing the Regulation of the Collection and Use of Health-related Genomic Information, McWhirter R et al, Centre for Law and Genetics 2021
- Genomic Data Policy Framework and Ethical Tensions, World Economic Forum (WEF)

** Complete list of standards available in the NAGIM Blueprint (Ch 7)*

APPENDIX J – NAGIM Implementation Stakeholders

STAKEHOLDERS	
Government Departments & Agencies	• Australian Government and State/Territory Departments of Health • Commonwealth portfolio agencies • Australian Institute of Health and Welfare (AIHW) • Australian Commission on Safety and Quality in Healthcare (ACSQH) • Services Australia • Office of the National Data Commissioner (ONDC)
Digital Health	• Australian Digital Health Agency (ADHA) • Australian Institute of Digital Health (AIDH)
Medical Associations & Peak Clinical Bodies	• Royal College of Pathologists of Australasia (RCPA) • Royal Australasian College of Physicians (RACP) • Human Genetics Society of Australasia (HGSA) • Australasian Society of Diagnostic Genomics (ASDG)
National Research Funding Agencies	• National Health and Medical Research Council (NHMRC) • Medical Research Future Fund (MRFF)
Standards Bodies – National & International	• Global Alliance for Genomics and Health (GA4GH) • Standards Australia • National Pathology Accreditation Advisory Council (NPAAC) • International Organization for Standardization (ISO) • Health Level Seven (HL7)
National Research Organisations	• Australian Research Data Commons (ARDC) • Commonwealth Scientific and Industrial Research Organisation (CSIRO) • Association of Australian Medical Research Institutes (AAMRI)
Research	• Universities and medical research institutes • Public and private research groups • Human research ethics committees and research governance offices
Community	• Data donors • Research participants, patients, public
ELSI & Governance	• Data custodians and data governance officers • Policy makers, legislators • Bioethics collaborators • GHFM 2021 ELSI of Data Governance Consortium (LINEAGE)
Indigenous Collaborations & Organisations	• Aboriginal Community Controlled Health Organisations (ACCHOs) • National Indigenous Genomics Network (NIGN) • National Centre for Indigenous Genomics (NCIG) • Australian Institute of Aboriginal and Torres Strait Islander Studies (AIATSIS) • CONNECT Research consortium • GHFM 2021 Indigenous Genomics Consortium
Data & Infrastructure	• NAGIM Prototypers Community of Practice (Australian BioCommons, Australian Access Federation (AAF), CSIRO, Garvan Institute, Max Kelsen, National Computational Infrastructure (NCI), QIMR Berghofer, Sezoo, University of Melbourne Centre for Cancer Research (UMCCR)) • Data generators and users – Research and Clinical • Health data infrastructure providers – Research and Clinical (e.g. Melbourne Genomics) • National infrastructure providers (e.g. NCRIS, AARnet, AAF, ADHA) • Information Technology Service (ITS) divisions for universities, medical research institutes, and healthcare organisations
Clinical & Healthcare Delivery	• Accredited public and private pathology providers • State/Territory entities: genetic services, digital health units
Industry	• Industry partners and Industry Representatives • Industry Genomics Network Alliance (InGeNA) • Commercial cloud providers • Technology developers and providers
International	• International partners (e.g. European Genome-Phenome Archive, Genomics England, Genome Canada)