Clinical NAGIM Project

June 2025

Australian Genomics

Acknowledgement of Country

In the spirit of reconciliation Australian Genomics acknowledges the Traditional Custodians of country throughout Australia and their connections to land, sea, and community.

We pay our respect to their elders past and present and extend that respect to all Aboriginal and Torres Strait Islander peoples today.

Artwork by Yorta Yorta artist, Alkina Edwards, for Australian Genomics.

Clinical NAGIM

Author Acknowledgement

Project Coordinator: Marie-Jo Brion (AG)

Laboratory Working Group Members: Hamish Scott (SA Pathology), Karin Kassahn (SA Pathology), David Lawrence (SA Pathology), Ben Lundie (Pathology Queensland), Meg Jeppessen (Pathology Queensland), Wishva Herath (Pathology Queensland), Julian Soubrier (Pathology Queensland), Eva Chan (NSW Health Pathology), Sebastian Lunke (VCGS), Zornitza Stark (VCGS), Simon Sadedin (VCGS)

Australian Genomics (AG) Project Team: Andrew Patterson (AG/UMCCR), Oliver Hofmann (AG/UMCCR), Miranda Vidgen (AG), Sarah Casauria (AG), Marie-Jo Brion (AG), Tiffany Boughtwood

Project Partners, Project Collaborations and Advisors: State/Territory and Commonwealth Departments of Health, LINEAGE Consortium (Mark Taylor), Shariant Project Team (Emma Tudini), Clinical Consent Project Team (Keri Finlay), Genomics Acceptability Project Team (Keri Finlay), Australian Genomics Community Advisory Group.

Executive Summary

The National Approach to Genomics Information Management (NAGIM*)* outlines a vision for coordinated national genomic data and infrastructure in Australia.

In the clinical setting, national data access is needed to maximise public health benefits and improve diagnostic yield, accuracy, efficiency, and equity. However, due to Australia's federated healthcare system, pathology providers face significant challenges accessing genomic data generated by other services—hindering timely and accurate diagnosis.

The Clinical NAGIM project, led by pathology services, identified two priorities:
1) Aggregated genomic data access, to improve efficiency for prioritising variants; and
2) Patient-level genomic data and clinical phenotype access for increasing diagnoses.

Services across four jurisdictions participated in developing and piloting a technical platform for sharing aggregated genomic data. Three services successfully shared baseline aggregated data. The project **highlights opportunities to build on the aggregrated model** and, critically, the need to **move towards patient-level data access to fully realise clinical benefits**

Scaling nationally co-ordinated clinical data access requires addressing both technical and non-technical considerations. These include legal and governance frameworks, patient consent processes, Indigenous data sovereignty, and public trust in platform infrastructure and data use.

Strong public support was identified for using genomics data in diagnostic care and a clear need for cross-jurisdictional access across pathology services.

Recommendations are provided around data and infrastructure, legal and governance, consent, Indigenous data, public acceptability and diagnostic research.

Several infrastructures may be progressed from here, including:

- Centralised or federated platforms for aggregated data (as piloted)
- Extending existing centralised knowledgebases (such as Shariant)
- A centralised trusted data environment for patient-level data, or
- A federated platform for patient-level data

Patient-level data offers the greatest clinical utility, though with greater governance barriers. While a centralised model is technically simpler and easier to govern, a federated approach may be necessary where data cannot leave a jurisdiction due to legal or policy constraints.

To progress, the suggested next steps are:

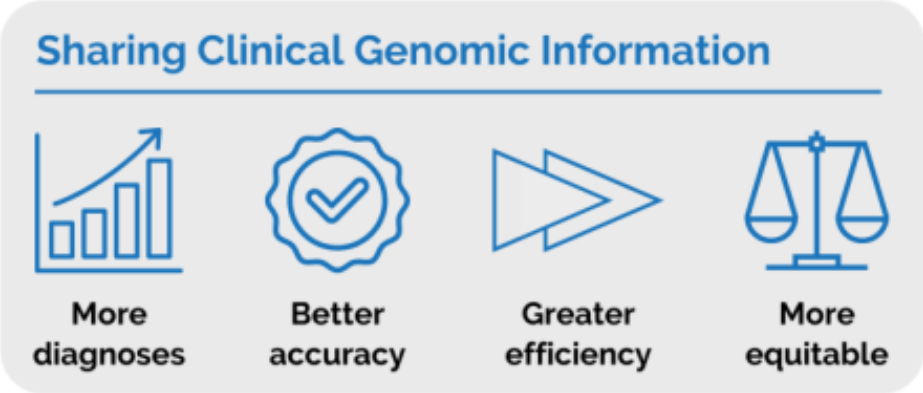
- **Advance to patient-level diagnostic data access**, building on the foundations established by the Clinical NAGIM project
- **Commit funding or incentives** to support data sharing infrastructure, clarify patient consent for clinical versus research use, and strengthen data access policies and mandates

- **Establish Executive Agreements** across jurisdictions and services to formally endorse clinical data sharing for diagnostics (as recommended by Genome Canada)
- **Prioritise implementation** of this report's recommendations, through the National Health Genomics Policy Framework (NHGPF) and Genomics Australia

After more than a decade of stakeholder engagement with limited national investment or progress, urgent coordinated action is required to avoid further fragmentation across jurisdictions and to preserve stakeholder support and momentum**.**

National approaches to clinical data access is fundamental to ensuring genomic data within diagnostic services can be used to maximise public health benefit - for patients and health services.

Background



Access to large genomic datasets is essential for accurate genomic data interpretation.[1] When a patient's genome is sequenced, pathology services must compare their genetic variants with those from other patients to identify potential causes of disease. Information about known or suspected disease-causing variants is often found in restricted-access platforms like Shariant or public databases such as ClinVar.

However, many disease-causing variants remain uncharacterised or are missing from these resources—particularly for patients with ultra-rare diseases or from underrepresented communities, where few genomes have been sequenced. In such cases, pathology services assess the frequency of a variant within a population to determine if it warrants further investigation. This is often done by querying the laboratories internal database of sequenced patients—the larger the cohort, the more powerful the analysis.

Enabling national access to clinical genomic data would significantly improve diagnostic processes. This is already standard practice in several international healthcare systems, including NHS England.

The Clinical NAGIM project seeks to expand the pool of variants available to diagnostic laboratories by building systems that allow authorised cross-jurisdictional access to patient variant data across Australia.

The expected clinical benefits include:

- Higher diagnostic rates with improved accuracy, safety and efficiency
- Reduced time spent ruling out non-relevant variants
- Greater public health benefit from better use of existing data for diagnosing other patients

Although a vast amount of clinical genomic data exists within Australia's healthcare system, much of it is inaccessible for routine diagnostic use due to the fragmented, federated healthcare structure. Providing secure access to de-identified data for pathology providers would enhance diagnostic accuracy, efficiency, and equity—especially for patients from minority backgrounds who are currently underrepresented in genomic datasets.

Aims

- Consider cross-jurisdictional genomic data sharing and access for diagnostic healthcare use
- Assess the feasibility of patient-level variant data access for diagnostic use: legal, governance, operational and technical feasibility.

Objectives

- Convene diagnostic laboratories representatives to establish cross-jurisdictional data access needs
- Design, develop and deploy a pilot platform to demonstrate technical and administrative feasibility of cross-jurisdictional data sharing.

Project Data Discovery

Consultation and representatives

Four genomics pathology services participanted in the Clinical NAGIM project: SA Pathology, Pathology Queensland, NSW Health Pathology and Victorian Clinical Genetics Services (VCGS).

Consultation with other Australian Genomics projects and collaborators: Shariant team, Genomics Public Acceptibility, Clinical Consent data use audit, Australian Genomics Community Advisory Group and legal academics from the LINEAGE consortium.

Data Access Needs

Representatives from genomics pathology services identified two initial priorities for advancing data access for diagnostic use in clinical genomic testing:

1. **Aggregated genomic data**: Summarised genomic data such as variant frequencies or counts. Primarily useful for excluding 'normal' or non-disease causing variants. While it has moderate diagnostic value, lower than accessing patient-level data, it represents low-risk, anonymised information with low expected governance requirements.

2. **Patient-level genomic data and clinical phenotypes:** Individual-level genomic data linked to patient clinical and phenotypic information, made available via queries for de-identified information and diagnostically-relevant insights. Higher sensitivity and more complex governance expected for sharing across services. However, has high diagnostic value, representing the core data needed by diagnostic laboratories and a key end goal for data accessibility for diagnosis.

Clinical Genomic Data

Clinical genomic testing by pathology services involves several critical data types:

- Genomic data and variant information
- Phenotype and clinical data
- Consent documentation
- Ethnicity information (important, but not always routinely available)

Many Australian laboratories capture these data in unstandardised formats and use bespoke data systems that lack interoperability. Standardised and machine-readable formats for these data types, aligned to national and international standards, will be important for data to be harmonised across services and for national data access processes to be scalable.

Exisiting Genomic Data Systems

Australian pathology services typically have a mix of on-premise and cloud-based systems, with cloud based infrastructure being increasingly adopted for large scale genomic data.

Unlike international initiatives using centralised solutions (e.g., Clinical Variant Ark in England) or common platforms (e.g., Emedgene adopted in eight Canadian provinces), Australian laboratories each use different genomic software and infrastructure.

Data AccessModels

Data access models considered include:

Centralised patient-level data model houses variant data and associated clinical and phenotypic data supplied by pathology services in a single infrastructure. It is accessed by pathology users through a restricted access platform.

Federated patient-level model keeps variant data and associated clinical and phenotypic data on the pathology service's infrastructure. Users send queries through a restricted access platform and receive summary data from multiple pathology services as a query output.

Aggregate data model is a low risk approach and would suit a centralised system, without a need for extensive governance due to the lack of sensitive data (patient-level variants and clinical phenotypes) present in the system.

Technical Pilot Outcomes

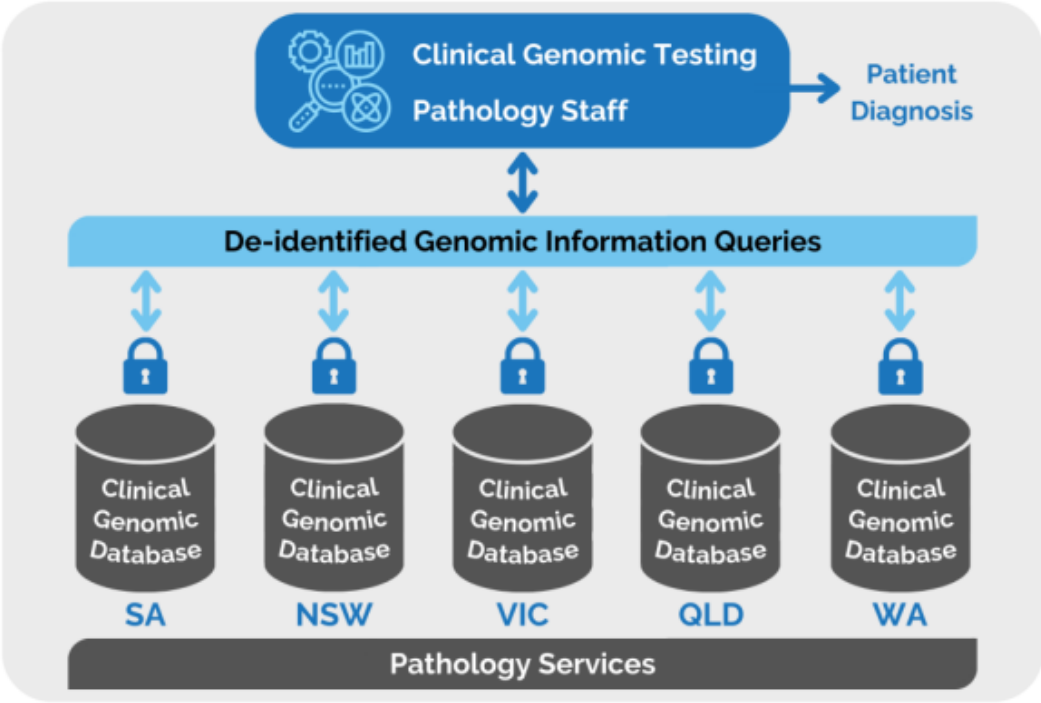


Figure 1. Exemplar model for federated access to healthcare genomic information in Australia

Initial Approvals

- **Executive summaries** of the Clinical NAGIM project were provided to leadership and executive teams at each participating pathology service.
- All four pathology services gave **in-principle approval** to participate in the cross-jurisdictional working group and progress a clinical data sharing pilot.

Data type

- The pathology services working group prioritised a **short-term aggregated genomic data sharing pilot** as the most feasible approach, to allow initial technical elements around diagnostic data sharing to be progressed, within a highly constrained timeframe (technical delivery in <12 months) and limited resourcing (in-kind/leveraged only).
- Laboratories agreed to share genetic variants generated from their diagnostic investigations, in an aggregated file format.

Cloud infrastructure

- All four laboratories already used a **cloud-based service** within their existing health service infrastructure. Three laboratories had access to Amazon Web Services (AWS) cloud-based systems, with the fourth in procurement with AWS when the project began.
- The pilot leveraged this common AWS infrastructure across laboratories to develop a cloud-based data-sharing pilot system.

Data sharing tool

- Various commercial, open-source and bespoke options for data sharing exist *(Appendix S1)*, however most were either not feasible for this pilot or not fit-for-purpose for Australian pathology services. These systems should be assessed further in the context of patient-level data access in a future phase.
- The project working group determined that a bespoke tool supporting federated data ingestion - with a central data processing layer - was the most feasible option, due to the diversity in databases across laboratories and presented a good opportunity to pilot a basic federated system. (*see Technical Specifications document*)[2]

Data standards

- Some laboratories were using standardised medical ontologies for clinical information (e.g. Human Phenotype Ontology (HPO)). However, services generally had limited machine-readable and extractable data in a standardised format. For example, free text for clinical information, PDF formats for consent, and alignment to different reference genomes.
- For flexibility and feasibility, this pilot did not enforce data standardisation, with manual data transfers and conversions used when required.

Outcomes

Data tool and system

- **This project implemented a federated approach (with central processing layer) for aggregated data to build initial foundations for genomic data sharing and access across diagnostic services.**
- Currently, genetic variants identified during diagnostic testing are stored in each lab's AWS cloud setup and not accessible by other diagnostic laboratories.
- A custom tool built by Australian Genomics/UMCCR ingested and processed laboratory-level data and generated aggregate variant counts in a central environment *(Appendix S2 and S3)*
- Aggregate variant count data was returned to each laboratories's infrastructure
- Variant count data is helpful for diagnostic laboratories in determining the pathogenicity of rare variants, and excluding non-relevant variants, thereby improving the accuracy, safety and efficacy of genomic testing.

Data sharing

- Of the four participating pathology services, three successfully shared aggregated genomic data into the federated system (SA, VIC, QLD).
- One pathology service required a legal agreement, which could not be executed during the project's remaining time-frame.
- Of the three contributing services, two succeeded in establishing their AWS cloud system (VIC, QLD), with data shared through this mechanism.

Considerations and Challenges

Governance and Legal Framework for Genomic Data Access

Clinical genomic data sharing requires formal governance approval through Data Custodians, Governance Officers, Executive Directors, and Legal representatives, aligned with applicable health data laws and privacy legislation. Organisational risk appetite significantly influences data sharing policies.

Aggregated genomic data

Three of the four participating laboratories were able to share variant count data into a federated platform without requiring formal legal agreements, demonstrating acceptability and feasibility of sharing low-risk, anonymized data across services without complex governance requirements.

Additional data (aggregated ethnicity, or phenotypes) would have enhanced clinical utility of the variant counts. However, this would have required more comprehensive governance approval, which was not pursued in this project due to the brief timeline.

Patient-level genomic data and clinical phenotypes

Executive support and legal/governance approval processes for this level of data sharing across pathology services remains untested. It was not feasible for participating laboratories to seek governance approvals for this level of data access in the project's timeframe.

No consistent executive approval pathways across services were identified to determine theoretical governance approaches for progressing patient-level data access, in the absence of an immediately accessible system for data sharing.

Considerations

- **Flexibility in data sharing arrangements.** Organizational risk appetites for data sharing varies across services. This includes acceptability of the different data types that can be shared. As organisational positions can be challenging to shift, flexibility in data and access processes will be important for different health services to contribute to a national data access system.
- **Consultations around legal agreements.** Legal agreements are designed to protect the interests of an administering entity and align to local regulatory requirements. Engaging executive and governance leaders in health services, to develop a legal data sharing framework, will be needed to ensure broad acceptance of national data sharing terms associated with patient-level data sharing.
- **Legal assessment for healthcare data sharing is complex and contextual**. It requires case-specific evaluation that integrates legal considerations with technical factors, data types, system environment, and privacy risks, specific to a data use scenario. This is best achieved once specifics of data types, access processes and technical platforms are known.
- **Choice of platform architecture** (federated vs. centralised) for national data sharing will likely impact types of data organisations are willing to provide, overall risk appetite, and associated legal requirements.

Conclusions

- Governance processes vary significantly across health services and jurisdictions and will be impacted by the specific choices of data, accessibility and platforms involved.
- Aggregate data sharing can progress across health services with minimal governance barriers, though risk-averse services may still require legal agreements.
- Clarity and guidance is needed around how relevant legal frameworks apply to sharing patient-level genomic and clinical data across jurisdictions and health services for diagnostic use.

Patient Consent

Variations exist across health services in data use consent statements. However, most clinical genomic testing consent forms include compatible statements about data use.

Examples include:

> **Lab 1:** *"Data and sample sharing: I understand and agree that the sample, genomic data and related health information may be shared and stored to help advance scientific knowledge (in a de-identified format – that is, without personal identifiers such as name and address)."*
>
> **Lab 2:** *"Results and related health information may be shared with genomic and medical databases that are used for patient care. All identifying information will be removed."*

Aggregated genomic data (variant frequencies, counts)

Most clinical genomic consent forms include non-optional statements about sharing anonymised data to scientific/medical databases as standard diagnostic practice. This typically covers sharing variant classifications and curation evidence to clinical platforms like Shariant and ClinVar.

However, consent is not necessarily required for aggregate data sharing, as demonstrated by broad submissions to these platforms and labs in this project successfully providing aggregated data within their jurisdiction's consent or general governance parameters.

Patient-level genomic data and clinical phenotypes

Since patient-level data sharing was not implemented, general requirements for supporting more granular data sharing remain untested with pathology laboratory governance and executives. Whether existing consent statements are sufficient for diagnostic access to patient-level data - from ethical, legal and governance, or social perspectives - is unclear.

Public Acceptability

Public acceptability is high for sharing clinical genomic data to diagnose other patients (~80%, AG Genomics Acceptability Project). However, community concerns may exist around lack of ability to opt-out; and the potential impact of this on willingness to have a test among priority populations, such as Aboriginal and Torres Strait Islander communities (AG Community Advisory Group).

Preferences and Technical Feasibility

Societal preferences for improving consent and provision of granular options and opt-out should be investigated further, and considered alongside the technical feasibility of health services to implement these processes. Machine-readable processes will be necessary for managing consent data at scale. Digital systems for managing consent are being considered or progressed in some jurisdictions and health services. However, manual, PDF-based approaches for storing consent information remain standard practice across laboratories.

Conclusion

- **For aggregated data:** Current consent clauses are not a barrier to broader sharing and diagnostic access.
- **For patient-level data:** Requirements for granular data access remain untested, and whether existing consent processes are sufficient is unknown.
- **Improvements to consent** should be considered to ensure broad acceptability of embedding patient-level data access for diagnostic use into routine testing processes, however both societal preferences and technical feasibility should be considered in parallel.

Indigenous Data in Clinical Genomics Testing

Pathology services ideally consider ethnicity of patients, including Indigenous Australian identity, as part of clinical genomic testing. This information ensures accuracy of genomic data analysis and test interpretation, particularly for ethnic minorities.

Aboriginal and Torres Strait Islander identity**, and ethnicity more broadly, can be self-reported by patients or genetic ancestry can be bioinformatically imputed by pathology services** using genomic data. However, in practice, while laboratories sometimes receive self-reported patient ethnicity or can calculate genetic ancestry from the genomic data, these processes are not yet routine practice across pathology services. In addition, there are some community concerns around computation and use of ethnicity data, in particular for Aboriginal and Torres Strait Islander patients.

Of the diagnostic laboratories in this pilot, Aboriginal and Torres Strait Islander identity was not consistently recorded across pathology services. There are uncertainties around best practice and community acceptability around generating and accessing genomic data from Indigenous patients.

Data on Indigenous identity were not requested or shared by laboratories in the technical pilot due to sensitivity and uncertainties around these data. While labs identified that ethnicity data, including Aboriginal and Torres Strait Islander identity, were highly important for diagnostic accuracy and clinically useful, stronger governance review and approvals are required before broader sharing of ethnicity and Aboriginal and Torres Strait Islander ancestry can be considered.

Conclusion

- No national guidelines currently exist for healthcare genomics services around best practice for managing Indigenous data in routine clinical genomic testing.
- There is a need to consider Indigenous Data Sovereignty principles and community perspectives around management, access and use of data from Aboriginal and Torres Strait Islander patients.

Public Acceptibility of Genomics Data Use in Clinical Settings

There is strong public support for healthcare genomic data sharing to improve healthcare and advance research. Specifically, our research reveals **high public acceptability for using healthcare genomic data for diagnosing others (82%)**, as well as for improving medical services for genetic testing (79%) and for medical research (76%). (AG Genomics Acceptability survey, N=1400 members of the public)

Improving understanding of the demographics of those who found clinical genomic data sharing "unacceptable" or were "unsure", and the reasons for these responses, can inform future implementation strategies for progressing data sharing for diagnostic use, including consent, patient information and public engagement and clinical workforce education.

Further implementation considerations include ensuring appropriate choice of organisation responsible for storing and managing the data, which significantly impacts public acceptance. There are clear preferences for public institutions as data custodians. Government agencies (69%) and local health services (66%) have significantly higher public acceptability than commercial providers (37%).

Commercial Providers

The higher acceptability of public institutions over commercial providers as data custodians may create complexities, as private pathology services conduct a substantial portion of genetic testing in Australia.

Additionally, there are hesitations among some pathology services in sharing genomic data that can convey commercially sensitive information about service delivery activities to commercial laboratories, e.g. testing volumes and associated test types.

Future Infrastructure Considerations

This pilot demonstrated that multiple pathology testing services can successfully share aggregated genomic variant data using existing agreements, processes and common commercial cloud providers, together with a centralised service to merge and redistribute the relevant information.

Future implementations could range from fully centralised knowledgebases to fully federated models, each with trade-offs in complexity, governance, cost, functionality, and vendor lock-in. **A fully centralised model** is technically and operationally ideal. It has fewer points of failure, a smaller data breach attack surface, and lower overhead for participating laboratories, as minimal local ITC support is required. Centralised data management also enables consistent quality control, streamlined access restrictions, and ensures all parties view the same data simultaneously. However, some laboratories may lack the appropriate governance frameworks or face legal and security restrictions that prevent even de-identified data from being transferred externally.

A federated model can address these concerns by retaining data at the source and transferring only necessary information upon request. While this enhances data sovereignty, it requires each site to maintain complex technical infrastructure, harmonised authentication and authorisation processes,

and synchronised system updates. Implementing effective data discovery and quality control across all sites also increases complexity.

The hybrid model piloted in this project (federated laboratory data nodes with a central data aggregator) aimed to balance data ownership concerns with system manageability. While it supports aggregated analysis of genomic variation in the Australian population, it lacks patient-level detail needed to fully support diagnostic accuracy. Nonetheless, it represents a practical interim approach while governance issues surrounding centralisation are resolved.

In conclusion, where possible, a centralised model is recommended as the preferred long-term solution as it maximises functionality and clinical utility, while minimising cost and vendor lock-in risks. National screening programs provide a suitable context to implement such a model, while simultaneously addressing governance and legal barriers**. If data cannot be centralised due to insurmountable restrictions, a federated approach may be pursued**—provided the additional technical and funding demands can be met.

In the interim, the hybrid (semi-federated) model can be maintained or expanded to support the existing clinical network, build distributed technical expertise, and further explore the benefits and limits of data aggregation.

Expansion of the Shariant central knowledgebase could also be considered, such as incorporating additional clinically significant variant subsets (e.g. rare or strong candidate variants). However, Shariant's infrastructure does not support granular patient-level genomic and clinical data and is designed to support sharing of *laboratory-classified evidence*.

Development of appropriate infrastructure for accessing patient-level genomic data across diagnostic services remains a priority.

General Considerations

There are a range of guiding principles for designing or choosing future solutions. These include balancing:

- Complexity
- Central Cost
- Node Cost
- Governance
- Functionality
- Vendor lock-in

In the current pilot, the decision to implement federated nodes was based on low complexity, central cost, node cost and governance, balanced by lower functionality and high vendor-lock in (reliant on AWS). **Continuation or progression of the system can increase functionality, or decrease vendor lock in, but with associated trade-offs of increased costs, difficulty of governance and complexity of the system.**

Costs and Sustainability of Data Sharing Systems

i) Aggregate Data Systems

The current pilot service for sharing aggregate genomic data has low operational costs and requires minimal resources at both participating sites and the central aggregation platform. Major future costs are expected to arise from further system development and the addition of features driven by new clinical use cases.

The use of native cloud services (e.g., AWS) contributes to low ongoing costs but may limit platform neutrality. All sites must have access to AWS and ensure their governance arrangements permit storage of data with a commercial cloud provider.

Estimated resource requirements include:

- **0.25 FTE per site**, shared between ITC staff (initial setup, security monitoring) and data managers (automating exports, metadata harmonisation).
- **2 FTE central coordination**, including research software engineers and a project manager to maintain infrastructure, support security audits, and manage new clinician query requests.
- **Approx. AUD 10,000/year** for commercial cloud operating costs.

While current operating costs are modest, the primary investment will be in future feature development as additional use cases emerge.

ii) Patient level Data Systems

The costs for patient-level data infrastructure are highly variable and depend on the chosen model (e.g. centralised, federated, hybrid) and the scope of requirements, including metadata harmonisation, quality controls, security audits, and legal review.

A meaningful cost estimate cannot be provided without detailed planning and coordination with participating laboratories. The next step is to undertake a structured evaluation process once there is consensus from laboratory, executive, and jurisdictional stakeholders that patient-level data access is a strategic priority.

Proposed next steps:

- **6-month project scoping phase**, led by a dedicated project manager, to consult with pathology services, assess needs, define technical options, and develop detailed requirements.
- **Resource estimates** for this phase include:
    - **1 additional central FTE** to support project planning, legal engagement, and technical design.
    - **0.5 FTE per participating site** to develop user stories, support requirements gathering, and contribute to the project planning process.

This phase would result in a comprehensive project plan, initial budget estimates, legal framework considerations, and potential procurement pathways (e.g. request for tender) for implementation

Diagnostic Research Processes

Research Activities for Diagnosis

Clinical genomic testing operates at the clinical-research interface, with ongoing development of improved analysis methods and continuously advancing knowledge of genetic variant-disease links.

Diagnostic genomic testing often requires research approaches to establish a patient's diagnosis. Supporting processes for diagnostic research to be undertaken for diagnostic purposes as part of routine clinical genomic testing is crucial for improving patient outcomes.

Current healthcare policies inadequately distinguish between "diagnostic research activities" (making diagnostic discoveries to establish a patient diagnosis) and "broad medical research" (generating new medical knowledge across populations).

This distinction is critical for rare disease diagnostics, as one laboratory representative noted: "*All rare disease patients are research patients.*"

Consent for Research

Consent and governance processes must clearly distinguish between diagnostic research and broad medical research to avoid confusion.

Patients undergoing clinical genomic testing typically receive an optional "research tick box" on consent forms for data use in medical research. However, the AG audit found limited utilisation of this data, confusion among ethics committees and clinical services regarding its scope, and some jurisdictions have removed this option.

Health services may adopt new consent mechanisms for patient healthcare genomic data research use, presenting an opportunity to improve consent and governance processes that better support diagnostic research versus broader medical knowledge generation

Recommendations

Clinical Genomic Data Sharing

- Fund or incentivise establishment of a national clinical system for pathology services to access clinical genomic data for diagnostic use.
- This can build on the **aggregated variant count data** from diagnostic laboratories as a first phase of a coordinated national genomic data access strategy for Australian health services. But, critically, should advance the next phase of **patient-level genomic and clinical data access** for diagnostic use to support the most clinically-impactful outcomes.

Data and Infrastructure

- Support adoption of genomic **data standards** and associated clinical test data, including machine-readable consent, ethnicity, and phenotypes; to support national harmonisation of data and scalability of data access processes across services.
- Resource or incentivise **IT infrastructure** upgrades within health services to ensure system interoperability required for national genomic data sharing platforms.

Legal and Governance

- **Seek formal Executive endorsement** across health services or jurisdictions, for in-principle support of clinical genomic data access for diagnostic use.
- Seek guidance on applying existing legal frameworks to sharing **patient-level genomic and clinical data** across jurisdictions / health services for diagnostic use
- Develop a template **legal agreement for diagnostic data access** across services for patient-level data, including consultations with health service legal representatives for compatibility with individual services' legal and governance requirements.

Consent Processes for Diagnostic Data Access

- Confirm **acceptability of existing consent clauses and processes** for including patient-level data access for diagnostic use as part of routine clinical practice, from public and governance perspectives.
- Determine **societal preferences and technical feasibility** for improving consent processes, including the consideration of opt-out processes, and integrate these preferences within governance mechanisms.

Indigenous Data Sovereignty

- **Convene a national conversation** to determine best practice for generating and accessing Indigenous data as part of routine clinical genomic testing. This should be undertaken with pathology and health services, accreditation standards representatives (eg NPAAC), and Indigenous data sovereignty experts.
- **Investigate acceptability in Indigenous patients and community** of allowing data access for diagnostic use in routine clinical genomic testing practice, and the impact on willingness to have a genomic test.

Public Acceptability

- Ensure the current **strong public support for genomic data use is maintained** by establishing national clinical genomic data infrastructure within trusted organizations (such as government agencies or public health services)
- **Build on existing research to better understand the demographics and reasons** for low acceptance and uncertainty in individuals, to inform strategies for consent, and public engagement.
- Determine the **nature of participation for commercial / for-profit pathology services** in a future national data access system, considering preferences and lower trust of patients and the public, and commercial sensitivities of pathology services for sharing test-based data.

Diagnostic research activities

- Reform healthcare policy, governance frameworks and consent language to recognise and **support research activities as integral components of standard diagnostic processes**, with clear policy distinction between patient-centred diagnostic research and broad medical research.

Conclusions

Building on the Clinical NAGIM Project Foundations

The 2024-25 Clinical NAGIM project successfully established a core working group across public/not-for-profit pathology services to advance clinical genomic data sharing.

The project identified two key use cases of clinical data access that would improve diagnostics:

1. **Aggregate genomic data sharing**
   - Sharing genomic variant counts as an aggregate
   - Moderate diagnostic value, allowing:
     - Exclusion of common variants unlikely to be pathogenic
     - Better identification of rare variants likely to be pathogenic
   - Feasible, low-risk, and economical initial approach.

2. **Patient-level genomic data and clinical phenotype access**
   - Sharing clinical and patient-level (non-aggregate) genomic information
   - High diagnostic value, allowing:
     - Accurate interpretation of genomic variants in context of clinical presentation
     - Diagnostic reanalysis
   - Greater governance complexity requiring executive commitment, dedicated resourcing, and robust coordination.

The project demonstrated basic feasibility of a simple federated infrastructure for aggregate data sharing and strong cross-jurisdictional willingness to advance genomic data sharing for diagnostic benefit.

From here, a range of infrastructures can be progressed including:

- A centralised or hybrid/semi-federated platform limited to aggregated data
- Extension of the existing Shariant knowledgebase
- A centralised trusted data environment for patient-level data
- A federated platform for patient-level data

Critically, the key next step is to **move beyond aggregate genomic count data and support access to insights from patient-level data to improve diagnostic yield, efficacy, safety and equity.**

Next Steps

1. **Commit funding or incentives for genomic data sharing systems, strengthen policies and mandates for diagnostic data access, and clarify consent processes to delineate research data sharing vs access for clinical diagnostic practice;**

   This can include further mandates for MBS-funded test data, and strengthening NPAAC pathology accreditation requirements around data provision for diagnostic use and around associated consent statements.

2. **Establish "Executive or inter-jurisdictional agreements"** as formal Executive endorsement for genomic data access across health services for diagnostic use.

   This approach was identified by Genome Canada as a key step they should have implemented from the outset, to progress effective data sharing across provincial healthcare boundaries. Implementation could be coordinated through the HTGC.

3. **Include this report's recommended activities as priority projects to** progress under the refreshed National Health Genomics Policy Framework, and implemented with prompt action through Genomics Australia.

Urgency to Act

Prompt action is needed to **maintain stakeholder trust and engagement, and compatability of systems being established across Australia.** A national approach to genomic data management has been on the national agenda for over a decade, with multiple consultations, recommendations, and minor or unfunded pilot initiatives (National Health Genomics Policy Framework (2017), NAGIM Blueprint (2020), NAGIM Implementation Recommendations (2023)). [3-7]

Without tangible action during this lengthy time period, there is risk of disengagement from genomics and healthcare stakeholders and divergent systems being built across jurisdictions and healthcare organisations.

Action is urgently needed to avoid further fragmentation and preserve stakeholder support for a national genomic data strategy.

References

1. Stark, Z., Glazer, D., Hofmann, O. *et al.* A call to action to scale up research and clinical genomic data sharing. *Nat Rev Genet* **26**, 141–147 (2025). https://doi.org/10.1038/s41576-024-00776-0

2. Australian Genomics. *Clinical NAGIM Variant Counts Project* (version 0.3). Technical Specifications Document. (2024). Available by request.

3. Australian Government Department of Health and Aged Care. *National Health Genomics Policy Framework 2018-202. https://www.health.gov.au/resources/publications/national-health-genomics-policy-framework-2018-2021?language=en*

4. *Queensland Genomics. Blueprint for a National Approach to Genomics Information Management (NAGIM). 2020. https://queenslandgenomics.org/national-approach-to-genomics-information-management/*

5. Australian Genomics*. National Approach to Genomic Information Management (NAGIM) – Implementation Recommendations. Overview*. 2023. https://www.australiangenomics.org.au/wp-content/uploads/2021/06/NAGIM-Implementation-Recommendations-Overview_April-2023.pdf

6. Australian Genomics*. National Approach to Genomic Information Management (NAGIM) – Implementation Recommendations.* 2023 https://www.australiangenomics.org.au/wp-content/uploads/2021/06/NAGIM-Implementation-Recommendations_January-2023.pdf

7. Australian Genomics*. National Approach to Genomic Information Management (NAGIM) – Implementation Recommendations. Supplementary.* 2023 https://www.australiangenomics.org.au/wp-content/uploads/2021/06/Supplementary-Information-for-NAGIM-Implementation-Recommendations_January-2023.pdf

Appendices

S1. Genomic Data Sharing Systems

Exemplar commercial, open-source and bespoke genomic data sharing systems in the national and international landscape, considered in the Clinical NAGIM project and data sharing pilot for diagnostic laboratories.

- Illumina - Emedgene
- DNAnexus federated platforms and tools
- AWS HealthOmics
- Global Alliance for Genomics and Health - Beacon
- University of Washington Genotype Phenotype Browser – Geno2MP
- NSW Health Health Pathology Gen-Phen platform
- UNSW rare disease genotype-phenotype platform
- CSIRO Transformational Bioinformatics group federated genomic data platform
- UMCCR Bespoke system (Piloted in this project; See Supplementary Information)

S2. Scripts and Software

Australian Genomics Clinical NAGIM data sharing pilot (2024-25) scripts and software:

https://github.com/umccr/nagim-clinical-data-sharing

Appendix S3 Example Aggregated Data Extracts from the Clinical NAGIM Laboratories

contig --- str	position --- u64	ref --- str	alt --- str	hom_count --- u32	het_count --- u32	labs_contributing_count --- u32	most_severe_consequence --- str	gene_id_set --- list[str]	transcript_id_set --- list[str]
NC_000002.12	178542755	G	A	9	18	2	missense_variant	["ENSG00000237298", "ENSG00000…	["ENST00000657023", "ENST00000…
NC_000002.12	178542910	G	A	22	17	2	missense_variant	["ENSG00000237298", "ENSG00000…	["ENST00000657023", "ENST00000…
NC_000002.12	178543986	A	G	921	1500	2	missense_variant	["ENSG00000237298", "ENSG00000…	["ENST00000657023", "ENST00000…
…	…	…	…	…	…	…	…	…	…
NC_000002.12	178802255	C	A	184	248	2	missense_variant	["ENSG00000155657"]	["ENST00000460472", "ENST…

contig --- str	position --- u64	ref --- str	alt --- str	hom_count --- u32	het_count --- u32	labs_contributing_count --- u32	most_severe_consequence --- str	gene_id_set --- list[str]	transcript_id_set --- list[str]	allele_string --- str	id --- str	seq_region_name --- str	start --- u64	end --- u64	…	canonical --- i64	cdna_start --- i64	cdna_end --- i64	cds_start --- i64	cds_end --- i64	protein_start --- i64	protein_end --- i64	variant_allele --- str	hgvsp --- str	hgvsc --- str	codons --- str	amino_acids --- str	consequence_terms --- list[str]	distance --- i64	protein_id --- str
NC_000002.12	178542755	G	A	9	18	2	missense_variant	["ENSG00000237298", "ENSG00000…	["ENST00000657023", "ENST00000…	G/A	rs2020664385	2	178542755	178542755	…	1	97324	97324	97099	97099	32367	32367	A	ENSP00000467141.1:p.Arg32367Cys…	ENST00000589042.5:c.97099C>T	Cgt/Tgt	R/C	["missense_variant"]	null	ENSP00000467141
NC_000002.12	178542910	G	A	22	17	2	missense_variant	["ENSG00000237298", "ENSG00000…	["ENST00000657023", "ENST00000…	G/A/T	rs560274402	2	178542910	178542910	…	1	97169	97169	96944	96944	32315	32315	A	ENSP00000467141.1:p.Thr32315Il…	ENST00000589042.5:c.96944C>T	aCc/aTc	T/I	["missense_variant"]	null	ENSP00000467141
NC_000002.12	178543986	A	G	921	1500	2	missense_variant	["ENSG00000237298", "ENSG00000…	["ENST00000657023", "ENST00000…	A/G	rs62621236	2	178543986	178543986	…	1	96383	96383	96158	96158	32053	32053	G	ENSP00000467141.1:p.Ile32053Th…	ENST00000589042.5:c.96158T>C	aTt/aCt	I/T	["missense_variant"]	null	ENSP00000467141
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
NC_000002.12	178802255	C	A	184	248	2	missense_variant	["ENSG00000155657"]	["ENST00000460472", "ENST00…	C/A/T	rs3568376	2	178802255	178802255	…	1	403	403	178	178	60	60	A	ENSP00000467141.1:p.Asp60Tyr	ENST00000589042.5:c.178G>T	Gat/Tat	D/Y	["missense_variant"]	null	ENSP00000467141