Introducing Data Passports

Jonathan Lawson
Thomas Conner

August 2025

GA4GH Data Passports: Proposing a New Technical Specification for Data Policy Representation

Proposing GA4GH Data Passports

A machine-readable, visa-based framework—modeled after GA4GH Passports—for representing data governance policies, making sure data policies reside with the data.

- New technical specification for data policy representation
- Addresses fragmentation in dataset governance requirements
- Provides transparent and portable policy representation
- Improves automation of compliance checks

Why Data Passports?

- Fragmented landscape of data governance requirements
- Inconsistent enforcement across various jurisdictions
- Lack of a standardized data policy representation
- Automation of compliance checks is challenging

Benefits and Opportunities with Data Passports

`Benefit	Description
Policy Portability	Encodes rules with the data, reducing the time/cost delays of system-specific policy implementation and expediting access
Interoperability	Facilitates cross-platform, cross-institution, cross-border data exchange
Auditability	Improves traceability and alignment with ethical/legal obligations[4]
Implementation Efficiency	Reduces burden of custom data onboarding and governance logic

Proposed Visa Types

ConsentedDataUseTerms

- Represents participant consent.
- Leverages the **Data Use Ontology (DUO)**

ApplicableLawRegulationsPolicies

- e.g. GDPR, HIPAA, tribal governance policy, relevant data sharing policies

ApprovedUsers

- Defines a specific list of users permitted to access the data.
- Applicable to controlled-access datasets, registered access, or consortia use.

OversightBodies

- Describes entities responsible for data governance, such as Data Access Committees (DACs) or Institutions with ownership, regulatory oversight, stewardship, or liability interest in the data

RequiredAgreements

- Lists contractual or policy agreements users must adhere to, e.g. institutional agreements aligned with GA4GH Model Data Access Agreement clauses[12] by way of direct links or persistent IDs to required documents

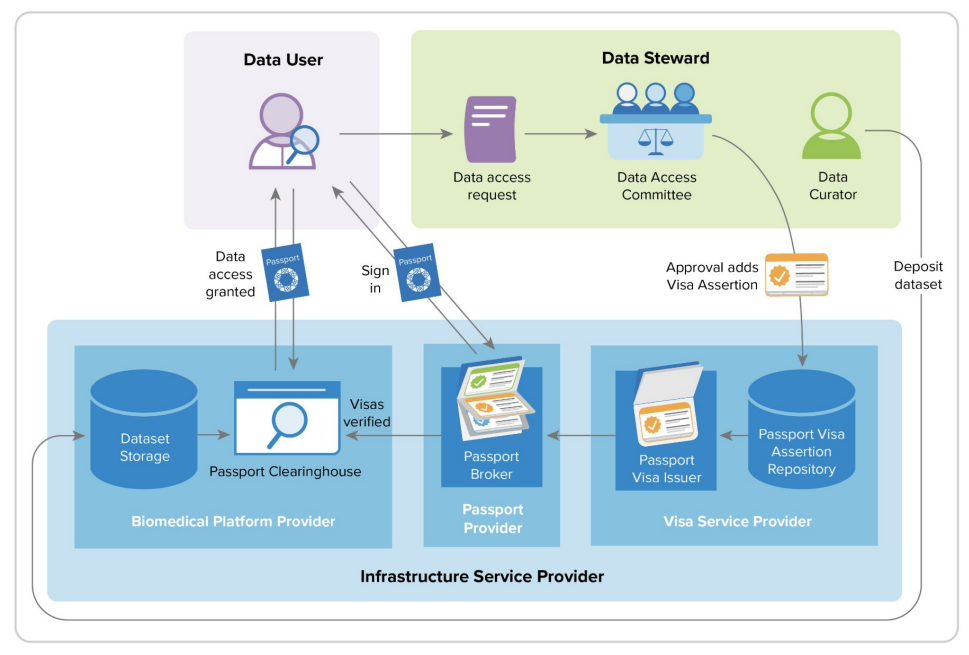
Interplay with Existing GA4GH Passports and AAI

Data Passports are designed to work **in tandem** with the existing GA4GH Passport standard.

Passports could be renamed **Researcher Passports**

This complementary relationship enables:

- **Dynamic access control**: Only granting access when both the Researcher Passport and Data Passport visas align.
- **Policy-aware analysis orchestration**
- **Auditable and reproducible access decisions**



Implementation Considerations

Data Passport Structure

- Each Data Passport consists of one or more visas.
- Each visa includes metadata: issuer, type, value(s), conditions, expiration (if applicable).
- Format can mirror GA4GH Passport JSON schemas for interoperability.

Binding to Data

- Passports may be:
  - Embedded in metadata schemas and linked to GA4GH DRS IDs[13]
  - Referenced via stable URLs or object identifiers
  - Managed in registries or access brokers

Policy Enforcement and Access Workflows

- Access systems (e.g., DUOS[14], TREs) can evaluate Data Passports alongside researcher Passports to make access decisions.
- Data Passports enable **validation prior to ingest and analysis** and **post-access auditability**.

Institutional Benefits and Passport Visibility

While Data Passports are designed to support **federated interoperability**, they can also provide substantial benefits to institutions for **internal governance and data management**.

Even when not externally shared, issuing Data Passports for datasets allows institutions to:

- Centrally catalog governance requirements.
- Automate internal access review and audit workflows.
- Prepare datasets for future sharing or platform onboarding.
- Monitor the policy status of datasets over time.

To support use cases that span **private to fully open sharing models**, Data Passports may optionally include a visibility field that specifies **access scope** for the passport metadata itself. Suggested values include:

- **public** – Metadata is discoverable and viewable by all.
- **institution-only** – Only members of the issuing institution can view the metadata.
- **group-specific** – Only designated user groups (e.g., consortia or cohorts) can view the metadata.
- **private** – Metadata is restricted to the issuer and explicitly authorized users.

This visibility control allows institutions to **balance transparency and confidentiality**, adapting to organizational policy, legal requirements, and stakeholder concerns.