



Improved Access to Secure Human Genetic Data

Data Storage

Data is individually encrypted in a format that allows direct random access and indexing. Even very large files can be accessed efficiently; only the required section of the file needs to be decrypted.

Individual encryption allows files to be stored in potentially insecure storage locations, lowering cost and increasing flexibility.

- Secure Data Storage in the Cloud
- Secure Data Storage on Disk

Data Access API

- Portable
- Cloud-Ready
- Secure

{ File Download } { Metadata } { Account Linker } { Random Access } { GA4GH htsgat } { GA4GH Beacon }

Data Access Models

- Flexible
- API Developed to support multiple diverse secure access models

RESTful API

- Programmatic Access
- Command Line Client
- Web Interfaces

GA4GH htsgat

- Advanced Genomic Queries and Filtering
- Genome Browsers
- GA4GH Beacon
- Discovery

FUSE Layer

- Access Archived Data as a POSIX-like File System
- Use Data without first downloading it to local disk

- REST API calls
- Streaming File Transfer
- Whole file / Random Access Downloads
- GA4GH htsgat Genomic API Access
- Direct Connection with Genome Browsers
- FUSE Layer Front Ends
- Data Replication Interfaces

Data Access Use Cases

RD Connect

- Submit relevant project data to EGA for Archiving
- Visualize in GenomeMaps Genome Browser
- Access Genomic Data via GA4GH htsgat API
- Real-time visualization

...

CSC Finland:

- Cache EGA Archive Data locally
- Shared Access to Common Data Cache to CSC users
- FUSE Layer Mounted in Secure VMs for Analysis

Your User Case?

Data Security

- Data Access Permissions are centrally maintained by the EGA in secure databases.
- Data access is made available by OpenID Connect/OAuth2 server, which is backed by EGA DB.
- Identity and Authorization can be Federated

- Download Re-Encrypting Whole Files
- List Permissions
- List File Details
- Get restricted Metadata
- Containerized: Flexible Deployment

