Gen3 2025 Development Roadmap

Gen3 Community Forum
26 February 2025

THE UNIVERSITY OF CHICAGO
CENTER FOR TRANSLATIONAL DATA SCIENCE

Australian BioCommons

NeSI New Zealand eScience Infrastructure

The Agenda

- Announcements
- Review of 2024 Roadmap
- 2025 Roadmap
- Steering Committee Discussion
- Q&A

Announcements

- New documentation site released! https://docs.gen3.org
- Next community forum will be on May 7, but the topic has not been finalized

2024 Gen3 Product Roadmap

Summary of 2024 roadmap

1. Gen3 Open-source Support **Complete**
2. Frontend Framework **Mostly Complete**
3. Helm for Deployment **Mostly Complete**
4. Integration Test Suite **Mostly Complete**
5. Observability **Complete**
6. Gen3 Lite **Mostly Complete**
7. Nextflow Integration **Phase 2 Complete**
8. RAG Interface for dataset search **Complete**
9. Data Lakehouse **Partially Complete**
10. Annual product roadmap **Complete**

Other requests provided at last year's roadmap meeting

GEN3 DATA COMMONS

1. Better visibility into roadmap progress **Partially complete**
2. More fine-grained access control **Not started**
3. Trusted research environment like capabilities **Partially complete**
4. Better data access request management **Partially complete**
5. Implementation of a native graph database like neo4j or neptune **Not started**
6. Better documentation for each microservice to help with extending Gen3 (API definitions and service interactions) **Partially complete**
7. Self-assembled working groups of interest **Not started**

2025 Gen3 Product Roadmap (Draft)

2025 Gen3 Product Roadmap (Draft)

- Commons Services Operations Center (CSOC) Support
- Task Execution Service (TES) Support
- Integrated AI capabilities
- Improved Documentation
- Frontend Improvements
- Mesh and Node Cards
- TBD
- TBD

CSOC Improvements

- A commons services operations center (CSOC) is used by organizations that run more than one Gen3 system.
- A CSOC allows a team of engineering and security staff to set up, configure, secure, operate, and monitor two or more data commons or data meshes.
- The CSOC Working Group is developing a resource to automate both infrastructure set up and Gen3 deployment
- This is a key step so that organizations can easily run 2 or more Gen3 commons

CSOC Vision: A Unified Gen3 Management Portal

One Portal for End-to-End Gen3 Lifecycle Management

- Seamlessly deploy, configure, and manage **multiple** Gen3 instances.
- Enable **zero to production** Gen3 deployments effortlessly.

Multi-Cloud Infrastructure Provisioning

- Support **Kubernetes** deployments across multiple cloud providers.
- Simplify cloud infrastructure setup and maintenance.

Community-Driven

- Incorporate **IaC contributions** from AU Biocommons, Krumware, OCC and other community partners.
- Collaborate on design and other features

Comprehensive Monitoring & Management

- Unified visibility into **Gen3, Kubernetes, and cloud resources** for proactive issue resolution.

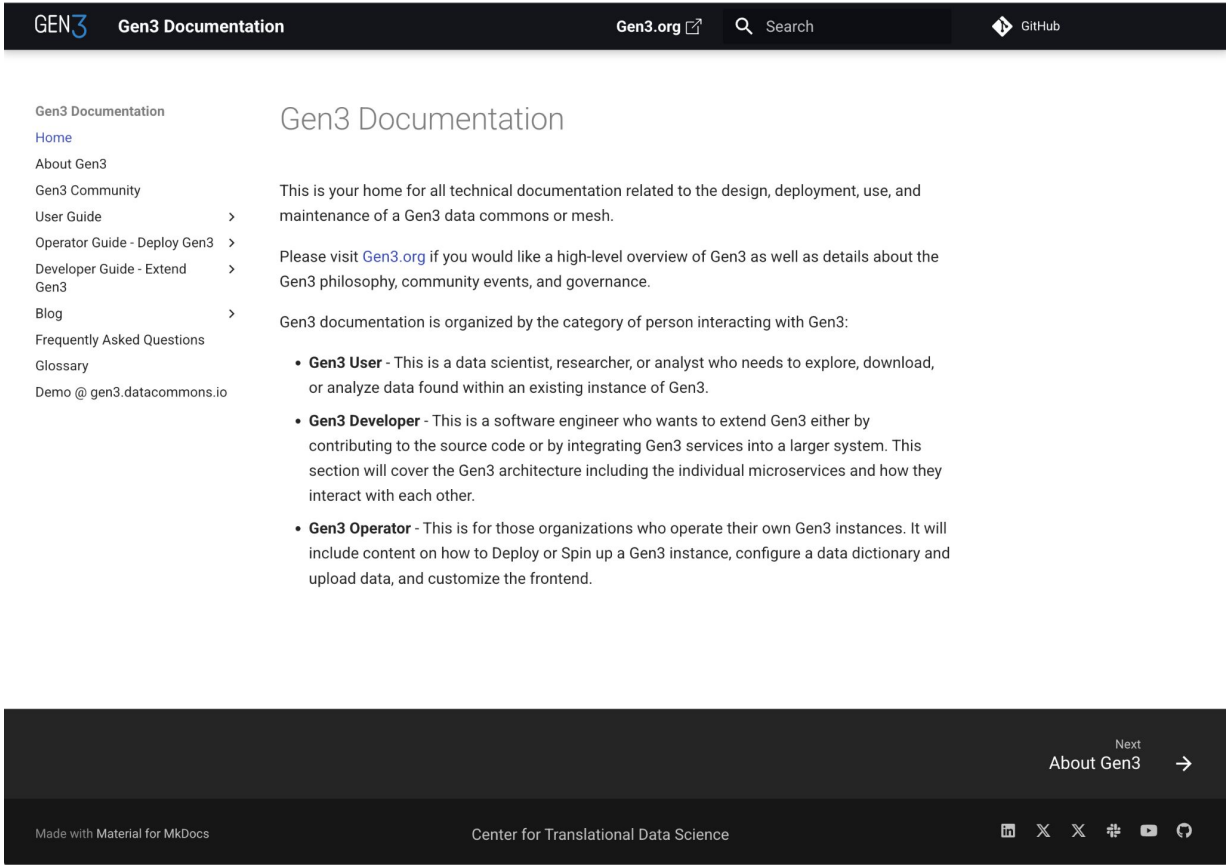
Task Execution Service (TES) Support

- Support container execution by implementing the GA4GH TES (Task Execution Service) standard, through an integration of the distributed task execution tool Funnel
- Support workflow execution by implementing support for Nextflow workflows through the Gen3 TES API
- The Gen3 TES API improves our ability to isolate tasks and workflows and allows for greater flexibility in running jobs outside of a workflow.
- This will allow us to uncouple Nextflow from workspaces and simplify the addition of other workflow specification languages in the future

Integrated AI Capabilities - Towards AI Commons

- Integrate open-source vector store into Gen3
- Add APIs for embeddings
- Add Gen3 services to create embeddings
- Add services to fine tune open source AI models over data in commons
- Add services for creating synthetic data from data into commons
- Explore architectures for building small to midscale LLM/GenAI over data in Gen3 commons
- Explore architectures for supporting federated / distributed AI in Gen3 meshes
- This will be a multi-year project different from the other updates included in the roadmap

Improved Documentation



Improved Documentation

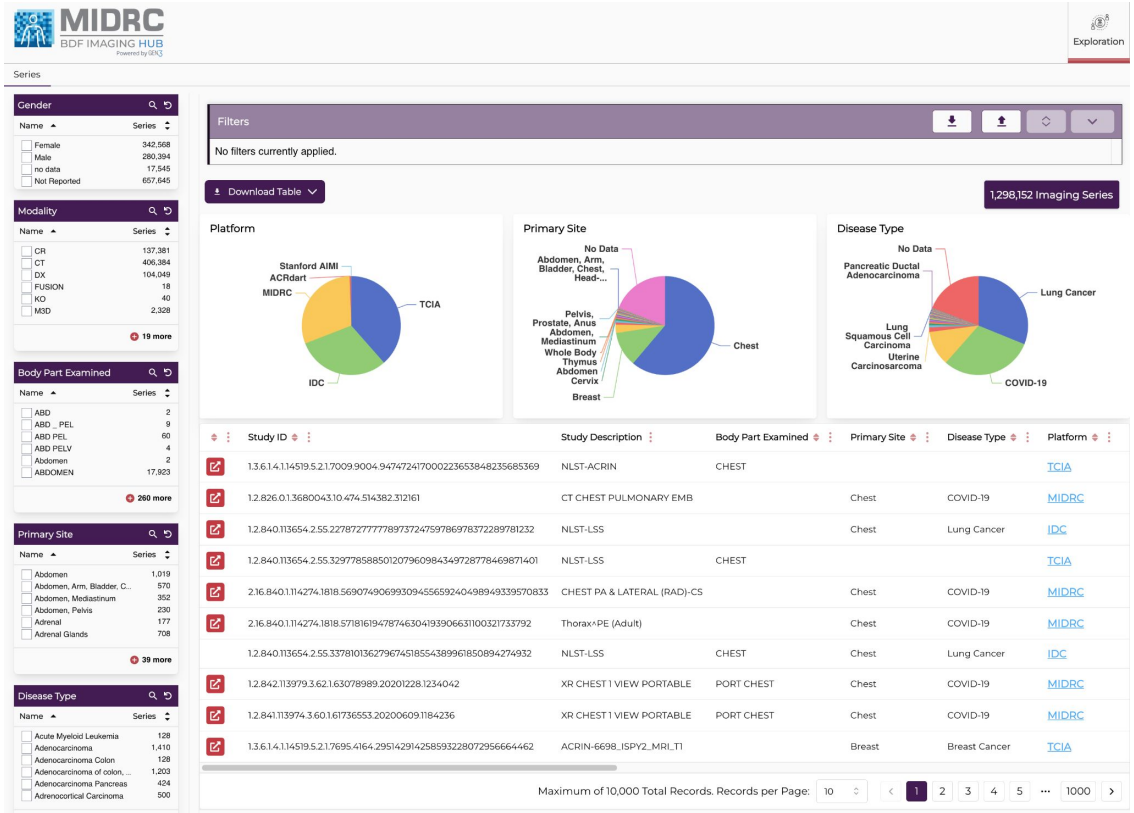
- Migration of all previous technical content on gen3.org to docs.gen3.org. Revision and addition of many new pages.
- Overall structure more clearly organized into separate sections based on use cases: Users, Developers, and Operators.
- Integration of helm documentation previously found at docs.gen3.org. Old versions now found at old.docs.gen3.org.
- New site is now generated using MkDocs and nearly all content is now exclusively in markdown format.
- Contributions by community to docs are better supported due to simpler content and contributor instructions.
- Content will now be versioned to align with Gen3 software.
- **More content to come in 2025!**

Frontend Improvements

GEN3 DATA COMMONS

Gen3.2

- Roll out remaining features (workspaces and dictionary view)
- Migrate existing data commons and meshes to Gen3.2
- Improve documentation
- Additional features in the form of apps

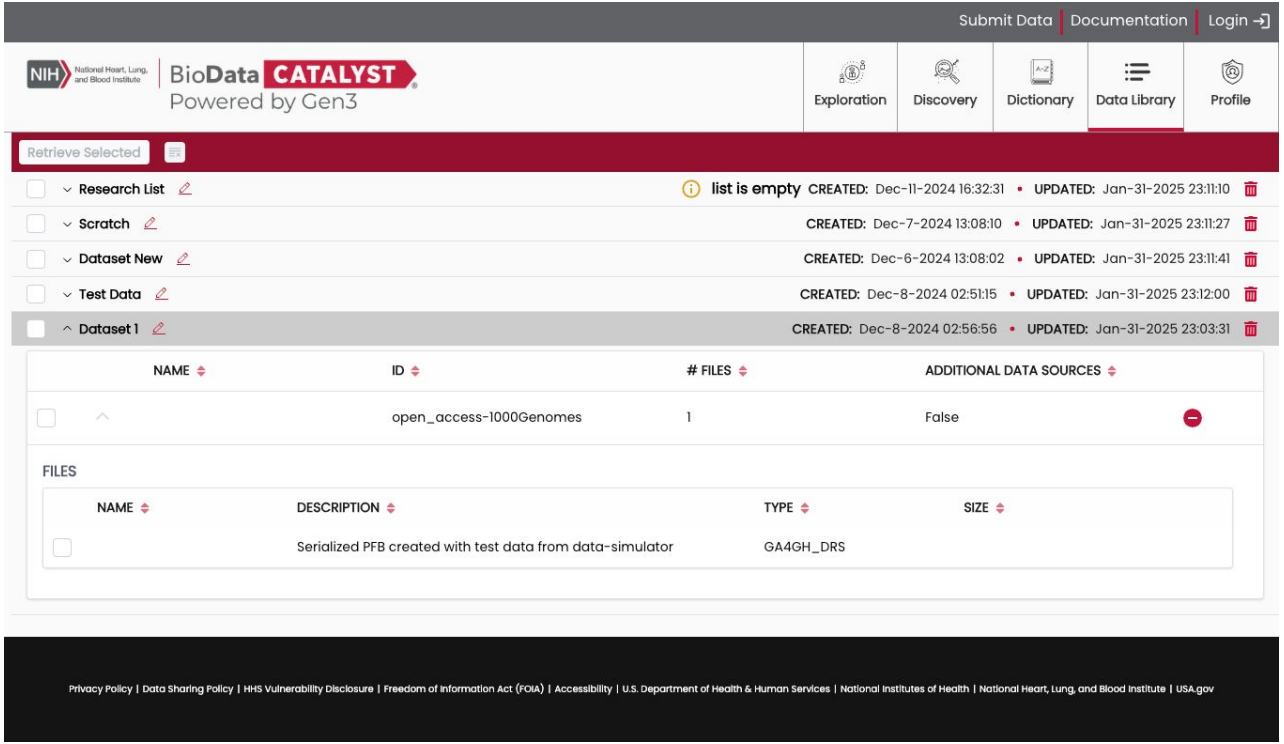


Frontend Improvements

Data Library

User managed list of data sets/cohorts for:

- Workspaces
- Exporting
- Sharing
- Applications

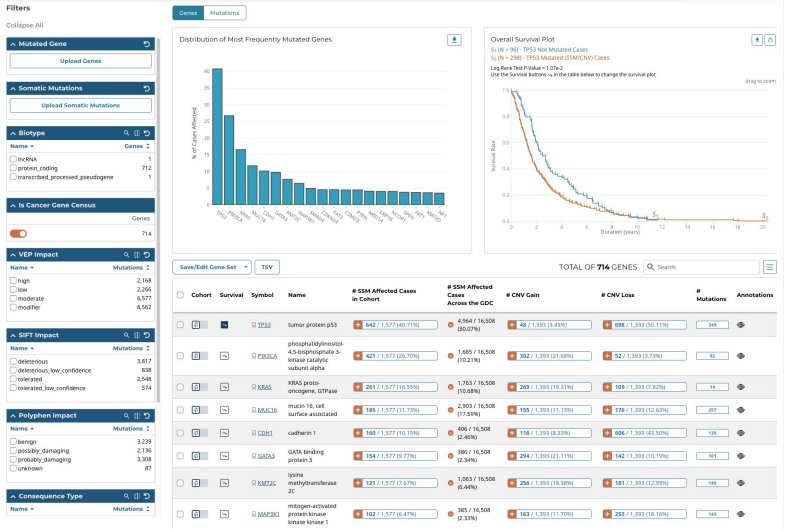


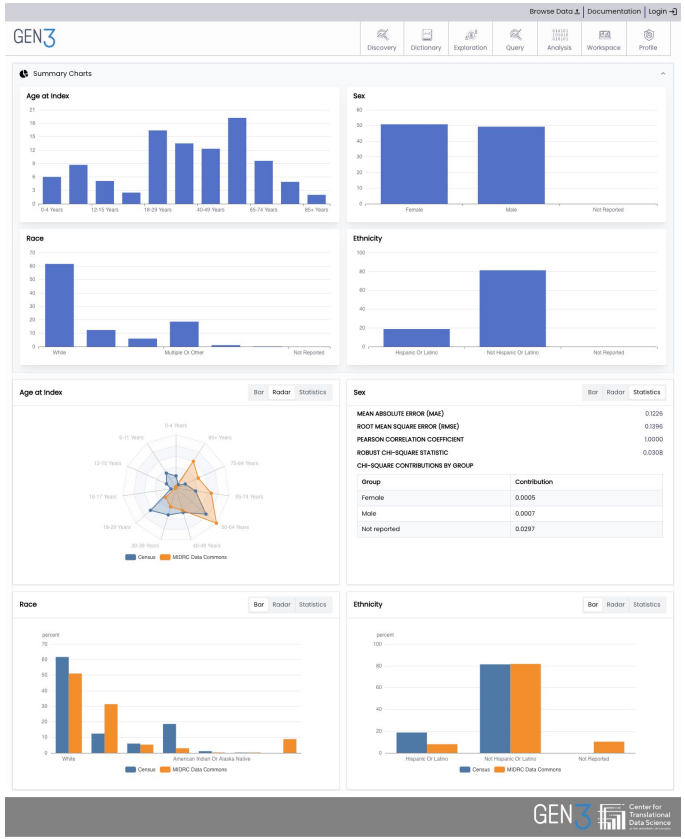
Frontend Improvements

GEN3 DATA COMMONS

Greater data visualization capabilities

- Will model many improvements based on the GDC to allow interactive gene and mutation visualizations





Mesh and Node Cards

- Machine readable document provided by a data commons to a data mesh (and vice versa) in order to participate in the mesh.
- Example mesh card:

```
{
      "card_type": "mesh",
      "meshcard_version": "1.0.0",
      "description": "description of data mesh",
      "data_type": "imaging",
      "usage_endpoint": "/usage",
      "metadata_endpoint": "/mds/metadata",
      "metadata_valid_apis": ["data_connect", "gen3_mds"],
      "auth_valid_apis": ["passports", "gen3_fence"],
      "data_valid_apis": ["DRS", "gen3_indexd"],
      "nodes": [
      {
            "card_type": "node",
            "id": "example-node",
            "meshcard_version": "1.0.0",
            "description": "description of a node in the mesh",
            "node_endpooint": "example.com",
            "metadata_api": {
                  "standard": "data_connect",
                  "endpoint": "example.com/mds/metadata/"
            },
            "auth_api": {
                  "standard": "passports",
                  "endpoint": "example.com/ga4gh/passports/"
            },
            "data_api": {
                  "standard": "DRS",
                  "endpoint": "example.com/ga4gh/drs/"
            }
      }
      ]
}
```

Mesh and Node Cards

GEN3 DATA COMMONS

- Example Node Card:

```
{
      "card_type": "node",
      "id": "example-node",
      "meshcard_version": "1.0.0",
      "description": "This would be the description of a node in the mesh",
      "node_endpooint": "example.com",
      "metadata_api": {
          "standard": "data_connect",
          "version": "1.0.0",
          "endpoint": "example.com/mds/metadata/"
      },
      "auth_api": {
          "standard": "passports",
          "version": "1.0",
          "endpoint": "example.com/ga4gh/passports/"
      },
      "data_api": {
          "standard": "DRS",
          "version": "1.5",
          "endpoint": "example.com/ga4gh/drs/"
      }
}
```

Other topics from the community

GEN3 DATA COMMONS

Gen3 Panel Discussion

Questions