One of the goals of this project is to build a user-friendly, scalable, and sustainable platform to support data sharing, discovery, and access. This centralized system will manage research datasets and metadata, facilitate collaboration among stakeholders, and improve data transparency. As part of the evaluation phase, the CS-DCC team conducted a comparative review of several open-source platforms, in-house solutions, and commercial tools suitable for building this infrastructure.

Open-source platforms

a. CKAN

- **Status**: Actively tested and configured in a development environment.
- **Pros**:
  - Well-established open-source platform with strong community support.
  - Rich plugin ecosystem.
  - Designed for open data portals with rich metadata handling and full API access.
  - Highly customizable and extensible.
  - More general-purpose data portal platform that can integrate and harvest metadata from a wide range of external sources.
- **Cons**:
  - Relatively Basic Access Control.
  - Doesn't support real-time federated search.

a. Gen3

- **Status**: Basic deployment tested using Helm and Kind.
- **Pros**:
  - Designed for large-scale biomedical research data.
  - Aligns well with the principles of a data mesh architecture
  - Granular access control and data commons architecture.
  - Support for real-time federated search.
  - Integrated with a data analysis platform.
- **Cons**:
  - Complex to deploy and manage.
  - Primarily designed to support the Gen3 data commons framework, with a strong focus on biomedical and genomic research data.

c. Bento

- **Status**: Preliminary evaluation completed.
- **Pros**:
  - Canadian-developed platform with a strong focus on biomedical data.
  - Good for collaborative metadata curation.
- **Cons**:
  - The data catalogue function is still under development.
  - Smaller community and fewer integration options.

In-house solutions

Data Explorer (Laval)

- **Status**: Actively developed and presented to the CS-DCC team.
- **Pros**:
  - Schema-driven exploration using Semantic Engine.
  - Vue.js frontend, Node.js backend, PostgreSQL database.
  - Supports CSV export, dataset merging, and basic filtering.
  - Future plans include Boolean search logic and ontology-based variable display.
  - Designed to support integration with CSAFS catalogue and Borealis.
  - Uses Amazon S3 for large file storage; metadata stored in PostgreSQL.
  - Access control features in place; OAuth2/ORCID integration planned.
- **Cons**:
  - Still under development, some features (e.g., advanced filtering, semantic harmonization) are planned but not yet implemented.
  - Requires schema creation before dataset upload.

Commercial solution

a. DNAstack

- **Status**: Contact established. Engagement expected in September because they are stuck with other projects.
- **Pros**:
  - Commercial support and ongoing development.
- **Cons**:
  - Licensing cost.
  - May be less flexible than open-source options for customization.

Preliminary recommendation

Based on the current evaluation, **CKAN and Gen3** stand out as the two most promising solutions for our data portal:

- **CKAN** provides a mature, flexible framework ideal for open or general-purpose data repositories, with strong support for customization and APIs.
- **Gen3** is a powerful choice for research-focused data portals requiring fine-grained access control, secure sharing, and structured metadata.

- **Laval Data Explorer** offers a tailored, schema-driven interface aligned with CSAFS needs, with strong potential for semantic integration and harmonization.

Each has distinct strengths, and the final decision may depend on the specific priorities of our project — such as openness, compliance, or data sensitivity.