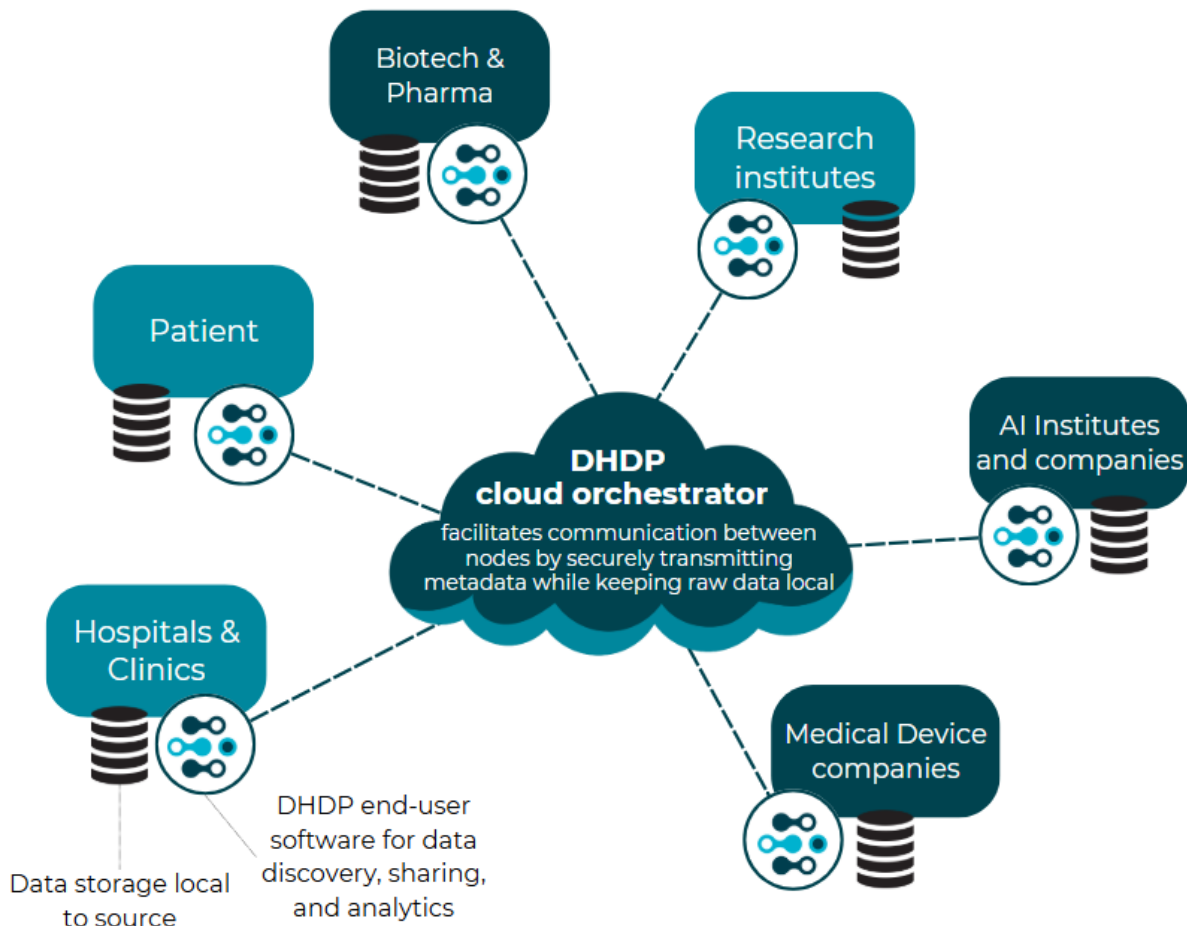


# Platform Functionalities

## Federated, secure and standardized data sharing

DHDP uses a **federated model to protect sensitive patient data**. Under this model, patient data remains local to the participating site where it is generated, and never crosses any institutional, local or provincial borders. Each site handles their own data and users with complete control over who can access each dataset. The DHDP will employ the OMOP common data model for data consistency and standardization. All data are encrypted in motion and at rest, and the Platform will comply with national and institutional regulatory requirements.





## **Diverse user base**

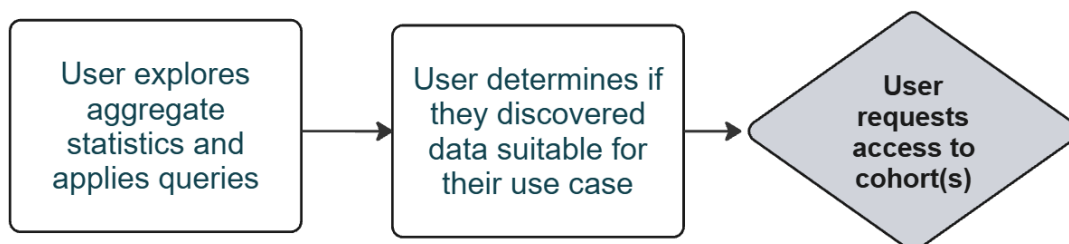
The Platform offers functionality to support a diverse set of end-users, including **researchers, innovators,** and **entrepreneurs**. End-users are typically classified as data providers if they share data, data users if they analyze data, or both, as data providers can also be data users.

- Examples of data providers: Principal Investigators, Physicians, R&D laboratory teams (academia, industry, not-for-profits)
- Examples of data users: Data scientists, Clinician Scientists, Machine Learning Engineers, Computational Scientist, Health Researchers and Bioinformaticians



## **Data discovery**

**Gain access to large-scale, de-identified datasets to answer critical research questions and accelerate progress in Canadian research programs.**



Through federated data discovery, authenticated users can explore aggregate statistics, visualizations, and data distributions across de-identified datasets. If, or when, they discover a patient cohort relevant to their research, users can request access to use the data for analytical purposes under the parameters data providers set through the Platform.

Currently, users can build data queries that combine clinical information, including demographic, diagnosis, and treatment parameters, with genomic variants to discover datasets of interest.

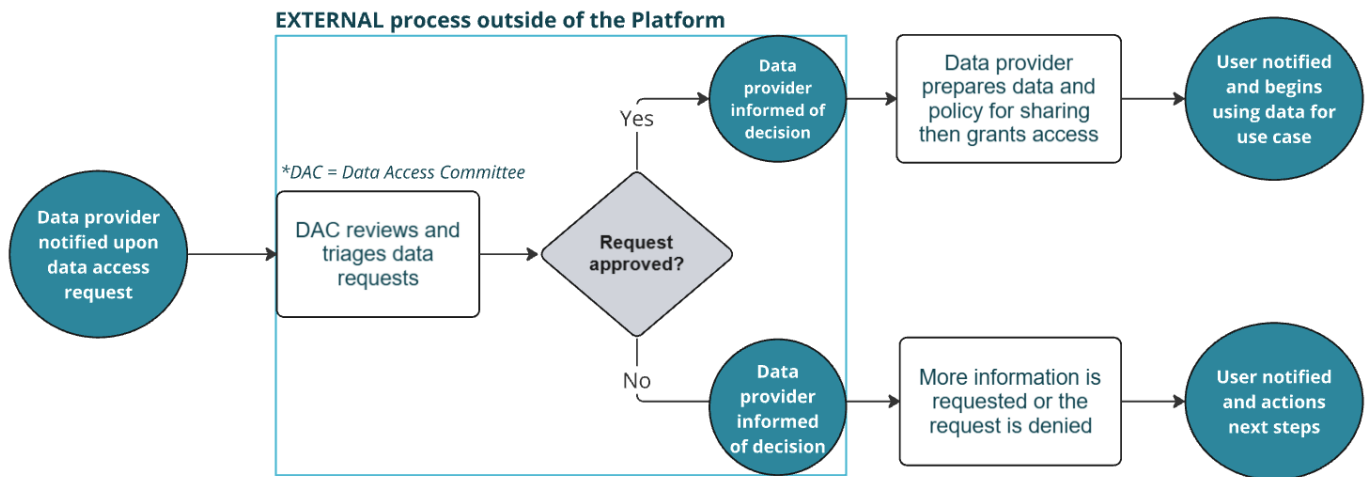


## Data preparation before sharing

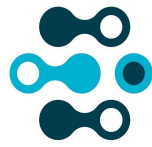
**Ensure the right individuals have access to the right data, at the right time.**

The Platform enables data providers to approve data access requests then share data without releasing raw datasets or creating copies. Data providers can customize data sharing by filtering or restricting access to sensitive information to then allow data users to analyze data without accessing the full, raw dataset. A complete log of jobs done on the shared data also becomes available for monitoring.

Important Note: DHDP is not responsible for triaging and reviewing data access requests. Data providers must undertake data access approval procedures outside of the Platform. Upon arrival at a decision, data providers will use the Platform to update the status of a data access request.



**Data preparation** — Through a user-friendly drag-n-drop workflow canvas, data providers can combine datasets for sharing, remove columns and/or rows, and apply other data transformations prior to sharing.



**Digital Health &  
Discovery Platform**

**Digital • Hôpital  
Découverte • Plateforme**

**Data policy management** — Data providers then need to create a policy which outlines the rules of engagement according to which the data will be accessed and used. This includes specifications for:

- User-based access control to set who, from which organization(s), data will be shared with.
- Time-bound data access to set how long access will be provided to data.
- Level of access permitted for the data user, including federated access (e.g. data can be used to train models but is not visible to the data user) and access to data within a controlled environment in which data previews may be seen.

## **Advanced analytics**

**Perform in-depth analyses, create visualizations, and train and validate machine learning models.**

The DHDP facilitates the coalescing of datasets from multiple sources for large-scale exploratory analysis, algorithm development, and data science application.

### **Details/features**

- Exploratory data analysis, visualization, or data transformation using SQL queries, Microsoft PowerBI, Jupyter Hub, and Platform built-in functionality
- Data science and machine learning model pipelines
- Leverage privacy-enabling technologies such as federated learning (FL) to scale access to data for the development of ML models across multiple decentralized nodes (e.g. hospitals, clinics) while preserving sensitive data by keeping it localized at each site
- Data augmentation or re-use of data for different analytical purposes and new insights

### **Use case examples**

- Predictive analytics (e.g. causal inference) for treatment trajectories and patient outcomes with regard to efficacy and safety
- Training and validation of diagnostic imaging tools
- Identify new predictive biomarkers and increase the number of available features for model development



## Digital Health & Discovery Platform

Digital • Hôpital  
Découverte • Plateforme

- Fuel new discoveries and detect meaningful patterns among patient cohorts
- Patient identification and stratification for clinical trial development
- External model validation across different patient cohorts, geographic regions, and institutions
- Scale access to data especially where data is scarce or siloed, such as for rare diseases

### **Important Notes:**

\*While the Platform is still in development, it will support a federated data governance model and offer functionalities currently supported by DHDP's technical partners' existing solutions. CanDIG is DHDP's open-source data querying partner, while bitnobi, a start-up solution acquired by Myant Inc, and integrate.ai have been engaged for data policy management and advanced analytical functionalities including data science and machine learning workflows.

\*The Platform will begin initial deployments in summer 2025 and build data acquisition processes for and with data providers. The Platform will Go-Live winter 2026 to support projects selected to receive the Digital Health Innovation Fund.

\*The Platform will not support data download and instead encourages the uptake of privacy-enabling technologies such as federated learning to scale access to data.