# **Welcome to the FSDH Proof-of-Concept Phase 2**

The FSDH PoC is an enterprise platform for invited GC scientists and researchers to store data, perform collaborative analysis, and manipulate data using advanced analytics tools.

# **Getting Started in FSDH – WORKSPACE LEAD**

## **Your Role – Workspace Lead**

* Responsible for monitoring Workspace usage and associated consumption of cloud credits, including work of collaborators
* Provide Workspace details (metadata)
* Invite collaborators – and ensure there will be sufficient credits available for them to actively participate in the Proof-of-Concept
* Use Case 1 – store and share data
* Use Case 2 – store, share and analyze data

## **STEP 1: Register and Login**

1a. Go to FSDH Registration page (Use Chrome or Edge browser)

At Email prompt enter valid GC email address

At Department prompt select your Dept/Agency from list (or add other)

Click **REGISTER**

*[image] Register page*

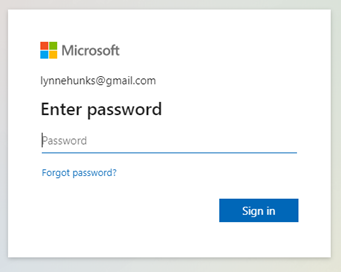
1b. You will be re-directed to the FSDH Login Page.

Login with your **GC email address and password**

*[image] FSDH Login*

1c. You will be presented with the Microsoft Authentication screen. Follow on-screen instructions to set up Multi-Factor Authentication (MFA). Note: we cannot provide more detailed instructions for this step because this is department specific.

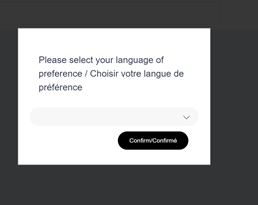
*[image]*



1d. At the Language Preference Screen, select your preferred working language from the dropdown. Click **Confirm**.

Note: this applies to the FSDH Portal only – tools such as Azure Storage Explorer and Azure Databricks are only available in English.

*[image] Language*



1e. At the Terms and Conditions screen, read and check the box “I agree to the Terms and Conditions” and then click Submit.

*[image] Terms and Conditions*

1f. You will then be directed to the FSDH Home page

*[image] Home page*

1g. You will also receive a FSDH New User Welcome email with the Terms and Conditions and a link to FSDH and the Getting Started Guide.

*IMPORTANT: if you have any issues with logging into FSDH please contact us*

*immediately using the* [FSDH Support Request Form](https://forms.office.com/r/zk82ehvUtv) *and choose* ***Login Issue.***

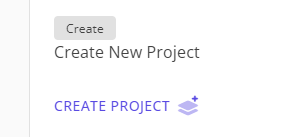
## **STEP 2: Navigate FSDH PoC Portal Landing Page**

* FSDH Home page
  + Complete when Home page available
  + Home
  + Workspaces
  + Resources
  + My Places
  + *[image]*

## **STEP 3: Create your Workspace**

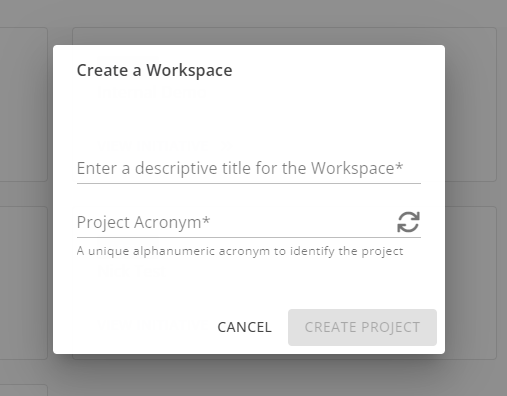
3a. From the FSDH Home page, click **CREATE WORKSPACE**

*[image]*



3b. Enter a descriptive name for your Workspace and add/create an Acronym. Click **CREATE WORKSPCE**

*[image]*



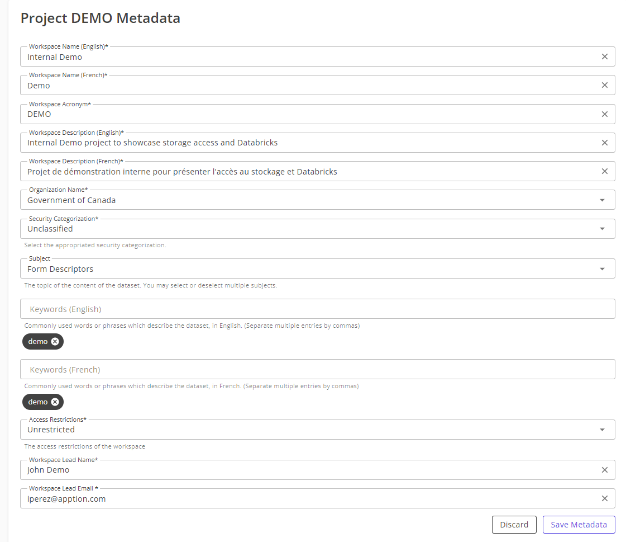
3c. A notice is displayed on the new Workspace page that the Workspace is currently pending approval. This could take up to XX minutes to complete.

*[image] Pending approval*

3d. When the new Workspace is available you must add the additional required metadata, in both official languages, in the Workspace Metadata form. **This mandatory step is required to request Storage and Databricks resources.** Refer to [*Metadata*](https://github.com/ssc-sp/datahub-docs/blob/main/UserGuide/Metadata/Workspace_Profile.md) for a description of each element.

Click **Save Metadata**

*[image]*



3e. You will be directed to your new Workspace page.

*[image] New Workspace*

## **STEP 4: Request Storage and Databricks**

* To be confirmed and screen shots identified
* Use Case 1 – request Storage Only
* Use Case 2 – request Storage and Databricks

## **STEP 5: Invite Collaborators**

5a. Before starting make sure collaborator has a valid GC email address

5b. Go to Workspace page and select USERS button

*[image] Workspace page*

5c. Select INVITE NEW USER Button, you will be redirected to the XX page

*[image]*

5d. Enter collaborator email address and select department and role and SAVE

5e. Collaborator will receive an email with this information:

FSDH New **Invited** User Welcome email with log in instructions, Getting Started Guide, T&Cs and name of Workspace Lead/Inviter.

*[image] email text*

## **STEP 6: Review Credits and Consumption Report**

(Describe report and what they are expected to do with it)

## **STEP 7: Upload a dataset to FSDH Storage and share with collaborators**

7a. Upload Data

Note: An email will be sent to all Workspace members when files are uploaded to the Workspace.

There aretwo options available for uploading files in the Workspace

* + **Storage Explorer**: provides a drag and drop interface and is accessible from the Workspace. This tool enables users to upload and download files.
    - *[image]*
  + **AzCopy** (command line interface): recommended option for working with large files and can also be integrated into scripts.
  + For details see [**Use AzCopy in FSDH**](https://github.com/ssc-sp/datahub-docs/blob/main/UserGuide/Storage/Use-AzCopy.md).
  + **FSDH Video Tutorial - AzCopy** focus has been created to walk users through the steps to upload files using AzCopy.

### **7b – Share Data**

You can share data in FSDH by inviting collaborators to your Workspace. See [Step 5: Invite Collaborators](#_STEP_5:_Invite) above.

## **STEP 8: (USE CASE 2 ONLY) Databricks**

**8a. Attend Databricks Introductory Session**

Introduce participants to the FSDH Azure Databricks – Why use Databricks; How to access Databricks in FSDH; How to setup your storage in Databricks; Frequently Asked Questions; How to get help using Databricks.

Walkthrough FSDH Azure Databricks to highlight features.

Date: TBD

**8b. Request additional Databricks Detailed Session**

A detailed individual or group session can be requested, if required. The content of the session will be determined based on your identified needs. Please send your request using the [FSDH Support Request Form](https://forms.office.com/r/zk82ehvUtv).

# **Support and User Documentation**

## **Support**

(client version of Amanda’s support table)

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |

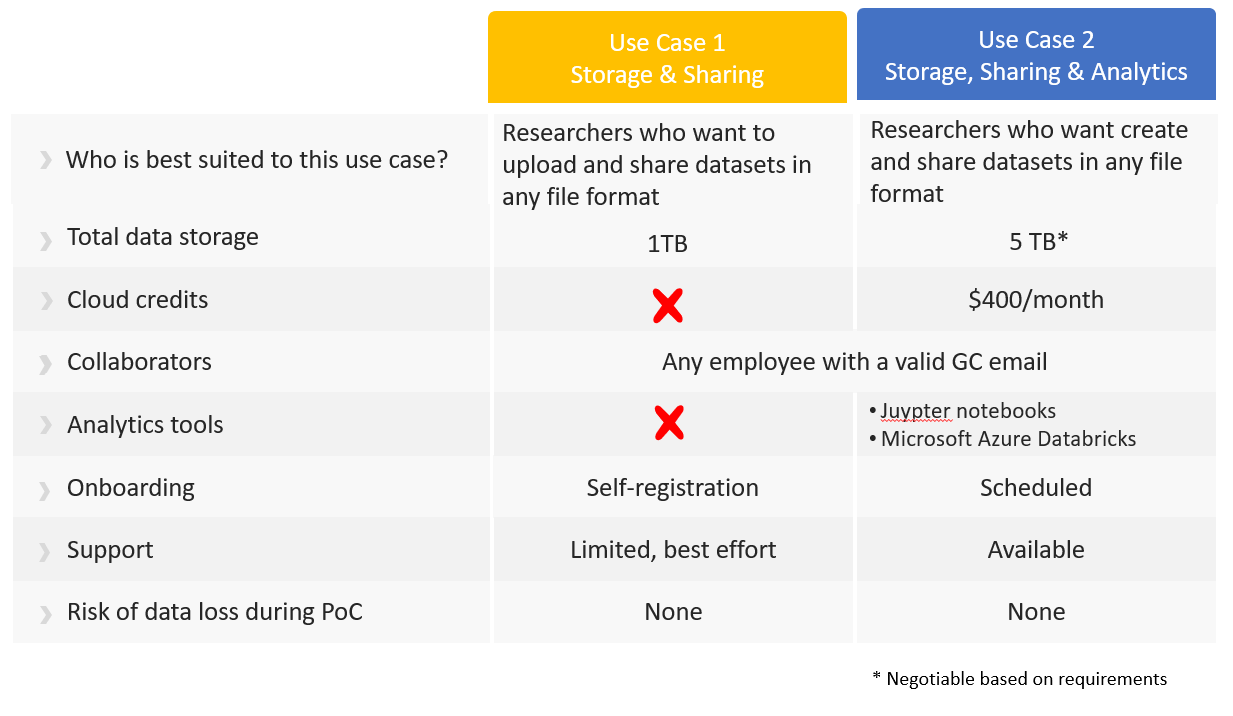
## **FSDH User Documentation**

*(in FSDH Resources section – should we change “Resources” to “User Guide”?)*

* **Workspaces**
  + [*What is a Workspace*](https://github.com/ssc-sp/datahub-docs/blob/main/UserGuide/Workspace/Workspace.md)
  + [*Workspace Profile (Metadata)*](https://github.com/ssc-sp/datahub-docs/blob/main/UserGuide/Workspace/Workspace-Profile-Metadata.md)
* ***Storage***
  + [*What is Azure Storage Explorer?*](https://github.com/ssc-sp/datahub-docs/blob/main/UserGuide/Storage/Datahub-AzureStorage.md)
  + [*Use AzCopy to Interact with Azure Storage Account*](https://github.com/ssc-sp/datahub-docs/blob/main/UserGuide/Storage/Use-AzCopy.md)
* ***Databricks***
  + [*What is Databricks?*](https://github.com/ssc-sp/datahub-docs/blob/main/UserGuide/Databricks/Databricks.md)
  + [*Databricks FAQ*](https://github.com/ssc-sp/datahub-docs/blob/main/UserGuide/Databricks/Databricks-FAQ.md)
  + [*Access your storage account in Databricks*](https://github.com/ssc-sp/datahub-docs/blob/main/UserGuide/Databricks/Access-your-storage-account-in-Databricks.md)
  + [*Data Processing using Databricks*](https://github.com/ssc-sp/datahub-docs/blob/main/UserGuide/Databricks/Data-Processing-using-Databricks.md)
  + [*Databricks Secure Access using ADSF Passthrough*](https://github.com/ssc-sp/datahub-docs/blob/main/UserGuide/Databricks/Databricks---Secure-Access-using-ADSF-Passthrough.md)
* ***Monitor Cloud Credits and Consumption***
  + *Report*
* ***Reference***
  + ***FSDH Proof-of-Concept Kick-off Deck (April 2023)***

# **Frequently Asked Questions**

1. **What are the 2 uses cases in this Proof-of-Concept?**



1. **What are the capabilities to be used and evaluated during the PoC?**
   * Cloud storage
   * Data Sharing
   * Notebook collaboration

1. **Where does the data ‘live’?**

SSC’s Azure Cloud – hosted in Canada.

1. **What will happen to the data at the end of the PoC timeline?**

Options are to delete/destroy or migrate to another environment specified by user.

1. **What is an Azure storage account?**

There are two key types of storage accounts – these are enabled by SSC and access provided by the FSDH team:

1. Azure Data Lake Storage (Gen2) - designed for big data analytics and provides Access Control Lists at the file and folder level
2. Azure Blob Storage - basic storage account - doesn't provide Access Control at the file and folder level but is required for compatibility in certain scenarios.

1. **Will I have issues using the FSDH with VPN?**

You may need to disconnect from your VPN to upload or download files.

1. **What if I need to more than the 5TB allocation or additional cloud credits?**

For the PoC we are trying to limit file sizes and stay within allocated budgets. However, if you feel you need more storage or credits, please open an issue with the [FSDH Support Request Form](https://forms.office.com/r/zk82ehvUtv) and your request will be reviewed.

* COST RECOVERY???

1. **What are the type of costs that my analytics project may incur?**

At high level, the costs for your project will depend on the following 3 factors:

* Compute – the type of virtual machines (Databricks Clusters) that you require for the project.
* Storage – the amount of data stored using the Blob Storage service
* Cost of any additional services that you may require for your project such as license cost for Power BI, Tableau, SQL Sever Database etc.

1. **Where can I see the costing model for the services? How can I track the cost that is being incurred for my project?**

Reports and alerts will be configured to enable you to track the ongoing usage and cost of the services consumed in your Workspace.

1. **What is the SSC Science Program?**

SSC created the Science Program to support GC priorities on modernising and securing research in Canada and to better enable scientists with digital infrastructure and services to support their respective mandates​. A key priority is to address and remove barriers that inhibit collaboration between SBDAs.

1. **What is the Data Solutions for Science (DSS) Team?**

The DSS team in the Science Program is mandated to deliver federal data solutions for Science Based Departments and Agencies (SBDAs) to address common data challenges and strengthen the posture of federal science in Canada​. In addition, we seek to eliminate ongoing investment in siloed departmental solutions that don’t meet the needs of the GC science community.

1. **What is the Federal Science DataHub (FSDH)?**

The FSDH PoC provides an enterprise platform for GC scientists and researchers to store data, perform collaborative analysis, and manipulate data using advanced analytics tools.

1. **Why a Proof-of-Concept?**

Allows for innovation and iteration as we discover and test data solution capabilities that will give scientists the tools they need to support their data management, storage, advanced analytics and collaboration requirements.