

2023 Data Ark: Data Commons User Survey Comments and Responses

Scientific Computing and Data

April 1, 2024

The 2023 Data Ark: Data Commons user survey—distributed in Feb 2023—solicited feedback from **183** Data Ark users (all users to date and their PIs). Of these, **19** responded (a **10%** response rate)

Survey Questions:

1. How would you rate your satisfaction with the Data Ark: Data Commons data quality and availability at Mount Sinai?

- Very Satisfied: **42%**
- Satisfied: **37%**
- Neither Satisfied Nor Dissatisfied: **16%**
- Dissatisfied: **0%**
- Very Dissatisfied: **5%**

2. How would you rate your satisfaction with Data Ark support?

- Very Satisfied: **42%**
- Satisfied: **42%**
- Neither Satisfied Nor Dissatisfied: **5%**
- Dissatisfied: **5%**
- Very Dissatisfied: **5%**

3. What barriers exist preventing your usage of Data Ark?

- A lack of knowledge of how to query data, and a lot of red tape to get access

Data Ark provides training to Mount Sinai researchers on hosted high-demand datasets. Additionally, our dataset-specific webpage provides details of the data source and other data-related information. We are exploring how best to collect exemplary data analysis code from our userbase to facilitate data use and analysis for less experienced researchers.

Our service is free-of-charge and access restrictions are imposed by data owners on the Data Ark website.

- It has become more difficult to easily upload and store files, even if they are publicly available. I have ended up storing these files in my personal space - which it is exactly the opposite that the reason Data Ark was created for

For onboarding data to data ark, please fill in [this online form](#). We will review the request and discuss with you according to our [onboarding policy](#).

- It is a pain to navigate and figure out how to remotely access the SAS / MarketScan system and do not have much information in terms of how to upload our own

datasets and offload result of analysis. Information on that will help new comers tremendously.

We migrated the MarketScan data to Minerva HPC, which provides the ability to access users' data files while analyzing MarketScan data. We are sending out the instructions on this.

In addition, we are exploring how best to collect exemplary data analysis code from our userbase to facilitate data use and analysis for less experienced researchers.

- Remembering to check if a dataset is there, which is my fault

4. Please enter any other feedback for improving our services on Mount Sinai Data Ark.

- - Provide a platform or a upload form for people to upload publicly available files and datasets - Provide guidelines in a website for data sharing and upload by users, so everyone can contribute - Facilitate sharing of files, or at least a directory with information about current resources. I am not sure this is available at the moment?

For onboarding data to data ark, please fill in [this online form](#). We will review the request and discuss with you according to our [onboarding policy](#).

- Great support team! MSDW dataset is very useful.

Thank you for your positive comments!

- Have people who've used the system (or designed it) for complex work in the Data Commons to explain how to use it better. Going for help and getting the response "I don't know" or "Just pay us and we'll get the data for you" is very frustrating.

We appreciate your comments. Data Ark provides training to Mount Sinai researchers on hosted high-demand datasets. Additionally, our dataset-specific webpage provides details of the data source and other data-related information. We are exploring how best to collect exemplary data analysis code from our userbase to facilitate data use and analysis for less experienced researchers.

- I've appreciated the improvements to the Data Ark website over the past 6 months or so. But further improvements in clarity and specificity regarding data use agreement requirements, IRB requirements for specific data sets, procedures for use, and guidelines around acknowledging the CTSA would be valuable.

Data use agreements (DUAs) are detailed on the <https://dataarkforms.hpc.mssm.edu/> webpage and are updated according to the regulations and terms set out by the data owners.

All datasets hosted on Data Ark are de-identified or anonymized, and therefore IRB approval is not required for accessing datasets.

The procedure of data access is detailed on the dataset-specific webpage. The guidelines for acknowledging CTAS is outlined on our website <https://labs.ica hn.mssm.edu/minervalab/resources/data-ark/>.

- More support on SAS, MarketScan environment because this a tremendously valuable resource and would love to see it more widely utilized.

Data Ark has facilitated the MarketScan training by the data owner Merative to Sinai-wide researchers on Jan 31, 2024. Additionally, Data Ark provides training on accessing this dataset via Minerva HPC on April 24, 2024. Also, we will consider collecting exemplary data analysis code from our userbase with the MarketScan User Group to facilitate data utilization and analysis for less experienced researchers.