

DATA SET POLICY
Data Ark Data Commons

Departments of Genetics and Genomic Sciences and Scientific Computing and Data

Title:	Onboarding/Offboarding Data Ark data sets
Related to:	Data Ark: Data Set Policy
Effective date:	07/19/2022

Purpose:

- To ensure that the Data Ark data sets are actively used to the research community

Policy Summary:

- To host data sets with the data set size commensurate with the amount of active usage per data set

Onboarding Data Sets:

- PIs will complete a REDCap form and name expected research groups
- Approval process according to data set size:
 - ≤ 1 TB: Data Ark operations team will approve
 - > 1 TB: must be approved by the Data Ark Advisory Board

Retention Period:

- Expected usage after 1 year is commensurate
- If usage is low, then the data sets will be removed from Data Ark
- The original data owner will receive usage reports every quarter and be alerted when others have not used their sets

Costs Waived for Hosting Data Sets:

Data Size (in Terabytes)	# Of user groups/data sets	Cost waived/year
1	≥ 1	0
3	≥ 3	\$300
10	≥ 4	\$1,000 (\$500 for 6 months)
20	≥ 5	\$2,000 (\$1,000 for 6 months)
30	≥ 6	\$3,000 (\$1,500 for 6 months)
100	≥ 10	\$10,000 (\$5,000 for 6 months)

Contacts:

Wen Huang
Data Scientist/Bioinformatician, High Performance Computing
Scientific Computing
Icahn School of Medicine at Mount Sinai
One Gustave L. Levy Place, Box 1498
New York, NY 10029-6574
wen.huang2@mssm.edu

Lili Gai, PhD
Director, High Performance Computing Scientific Computing
Icahn School of Medicine at Mount Sinai
One Gustave L. Levy Place, Box 1498
New York, NY 10029-6574
lili.gai@mssm.edu

Maria Julia Castro, MS
Project Manager
[Department of Genetics and Genomic Sciences](#)
New York, NY 10029-6574
E mariajulia.castro@mssm.edu

Patricia Kovatch
Dean, Scientific Computing and Data
Professor, Departments of Genetics and Genomic Sciences and Systems Pharmacology
Icahn School of Medicine at Mount Sinai
New York, NY 10029-6574
patricia.kovatch@mssm.edu