

## Plan Overview

---

*A Data Management Plan created using DMPTool-Stage*

**DMP ID:** <https://doi.org/10.48321/D1402Z>

**Title:** Data Management with the Broad Institute's Data Use Oversight System (DUOS)

**Creator:** Pamela Bretscher - **ORCID:** [0000-0003-4938-7990](https://orcid.org/0000-0003-4938-7990)

**Affiliation:** Broad Institute ([broadinstitute.org](https://broadinstitute.org))

**Funder:** National Institutes of Health ([nih.gov](https://nih.gov))

**Template:** NIH-GDS: Genomic Data Sharing

### **Project abstract:**

The Data Use Ontology System (DUOS), developed by the Broad Institute in conjunction with the NIH, is a mechanism for making data publicly available and managing access requests from secondary researchers. DUOS is a semi-automated system aimed at streamlining dataset registration and access requests guided by the Global Alliance for Genomics and Health (GA4GH) Data Use Ontology (DUO). DUO utilizes standardized terminology to codify the allowed uses of a dataset and align these with a secondary researcher's intended use case. This is achieved through both human and machine-readable data use terms that are used to tag a dataset, as well as a researcher's access request application. Primary researchers can use the Broad Institute's data access committee (DAC) to manage requests for access to their dataset from secondary researchers. This DMP is a mock research project for the purpose of creating a template DMP for researchers looking to use DUOS for their data management.

**Last modified:** 08-07-2023

### **Copyright information:**

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customize it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

---

## Data Management with the Broad Institute's Data Use Oversight System (DUOS)

The proposed research involves single-cell RNA sequencing data of melanoma tumor samples from 100 mouse-model individuals. In this study, we also plan to generate novel QC methods and pipelines for analysis which we also plan to share. The full study protocol and methods for collection will be included in our publication along with information for requesting access to the controlled access human data.

We aim to generate single-cell RNA sequencing data for 100 individuals with melanoma. For genomic data generated from this study, our Data Management Specialist will organize and share our data under controlled-access procedures by registering our data in the Broad Institute's DUOS (Data Use Oversight System), which is being piloted by the NIH, and leverage our institution's DAC. The DAC will receive and review data access requests and ensure that it is being used in accordance with the restrictions outlined in the consent forms for our study.

We aim to generate novel QC methods for preparing single-cell RNA sequencing data for analysis under this study, and plan to share these tools publicly. For the code and other materials associated with this study, we will share via open-access procedures by registering our data and tools in the Broad Institute's DUOS (Data Use Oversight System), so researchers are able to find and navigate them directly for personal download. All software and tools generated and shared will be available along with documentation for deployment and use.

Once generated and analyzed, our data will be deposited in the Terra system developed by the Broad Institute. Once this is complete, we will submit a dataset submission form to DUOS containing the data use limitations for our dataset to be approved by our institution's data access committee. Once this request is approved, secondary researchers will be able to request access to the dataset via DUOS, and these requests will be processed by the committee. We will serve as data custodians for the dataset, in that we will be responsible for providing data access to researchers who are approved by our data access committee.

We are using non-human data.

One limitation on the secondary use of this data is that it is only to be used for further research on cancer. Additionally, it is required that secondary researchers agree to collaborate with the original study team in any future research.

The study can be designated as "not sensitive" as it does not contain identifiable human data.

---