Nina Dashti-Gibson

October 3, 2025

CCTR 691

FAIR Data Assessment Homework

1. A data(set) should be assigned a globally unique persistent and resolvable identifier when deposited with a data repository

The dataset has a unique PID (GSE276609).

1. When you deposit a data(set) in a data repository, you will need to provide discovery metadata in order to make the data(set) findable, understandable and reusable to others

There is information about the dataset provided including the sequencing platforms used, links to the GEO pages of reanalyzed samples, and descriptions of each new sample.

1. The data repository providing access to your data(set) should make the metadata describing your data(set) available in a format readable by machines as well as humans

There is a structured data table which is machine-readable.

1. Access to your data(set) may need to be controlled and that metadata should include license information under which the data(set) can be reused

There is not an explicit license, but the data is freely available to download from GEO. The publication includes a “Data Availability Statement.”

1. Metadata should remain available over time, even if the data(set) is no longer accessible

The metadata is listed independent of the links to the download the data.

1. The metadata describing your data(set) should use controlled vocabularies

There are consistent labels used to describe the data, but it seems like every sample has the same label (Basal-like triple-negative breast cancer patient-derived xenograft)

1. Provenance information about the collection and/or generation of data should be included in the metadata

The contributors, submission date, and last update date are provided.

1. Metadata describing your data(set) should follow the specifications of a community-endorsed standard

The standard naming conventions of the PDX models are used.

1. Your data(set) should be deposited preferably in a file format that is open and supported by the data repository for long-term preservation

Standard file formats are used (e.g. CLOUPE files).

1. Keeping your data(set) FAIR over time requires professional data curation and digital preservation

GEO is a trusted digital repository maintained by the NIH.