

Metadata Quality in OSDR

SMD Workshop 2024

San Huei Lai Polo
she/her
Open Science Data Repository/GeneLab
KBR
san-huei.lai@nasa.gov

BPS

Biology & Physical Sciences



National Aeronautics
and Space
Administration



What is Metadata Quality?

Metadata quality is defined by its ability to accurately describe data and meet specific criteria like completeness, accuracy, and compliance with standards such as the FAIR principles. High-quality metadata ensures that data is understandable, discoverable, and reusable by both humans and machines. It is closely tied to the fitness-for-use of the data, making it essential for effective data sharing and reuse.

(Kindling & Strecker, 2022)

“The level of data curation performed at repositories and, as a result, the quality of metadata varies (Koshoffer et al. 2018).”



No curation = data quality is the responsibility of the submitter

High level of curation = high metadata quality

QUALITY CRITERIA	NASA OSDR
Accessibility	OSD identifier and DOI
Open data license	CC0 license
Open data format	Open and non-proprietary format
Appropriateness of metadata/data documentation	Clear source of data, data handling and processing (tool and software used), version history
Consistency	Metadata are compliant with metadata schema and is consistent within the repository; use of ontology
Accuracy	Built-in validation; primary and secondary curators in addition to submitter
Reuse potential	Our AWGs continually help us evaluate the fitness-for-use of our datasets. Feedback from people interested in a wide variety of data usage makes our data/metadata quality especially well rounded

Challenges of Metadata Quality in OSDR

- Human error inevitable— programmatic/AI tools in development
- Data submitter/PI time is limited and thorough metadata is not always prioritized— provide metadata templates early on
- Commercial spaceflight partners sometimes do not share all relevant metadata
- Hardware environmental data can be very hard for curators or data submitters to obtain
- Ontology maintenance is variable