OBI Data Quality Framework

Stage 2 – Glossary and List of Acronyms

Version 1.2 - Date: August 8, 2024
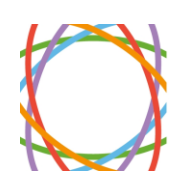
Introduction

The terminology used in the Data Quality Framework (DQF) leverages existing definitions using standards from BIDS and DAQCORD and has been adapted and integrated with how these terms have been used previously in OBI-funded research to meet the needs of the Data Quality Framework in describing data across a range of modalities. This document provides descriptions for terms used throughout the DQF to classify data into "levels" and to describe data capture and curation.

References:

- BIDS
- Ercole, A., Brinck, V., George, P., Hicks, R., Huijben, J., Jarrett, M., Vassar, M., Wilson, L., & DAQCORD collaborators (2020). Guidelines for Data Acquisition, Quality and Curation for Observational Research Designs (DAQCORD). Journal of clinical and translational science, 4(4), 354–359. https://doi.org/10.1017/cts.2020.24

Glossary

Term	Definition
Data Classification Levels	
Study dataset	Contains all data from a single study, organized into one or more data packages.
Data package	An organized collection of folders/files containing data and related metadata for one modality.
Modality	Describes a broad category of data based on what is being measured and/or how it is measured (e.g., clinical, imaging, wearables, and genomics).

	Note: in neuroimaging, "modality" often refers to imaging techniques or pulse sequences (e.g., fMRI, T1-weighted MRI, EEG). In the Data Quality Framework, these are referred to as instruments.
Instrument	A clinical questionnaire/form that is often collected using REDCap (e.g., Brain-CODE Demographics, WHOQOL-BREF), or a scanner or EEG amplifier, etc.
Record	Data from one participant from one instrument at one study timepoint. A timepoint may be defined in different ways, such as the baseline visit in the study schedule, or further divided into individual days within a multi-day observational data collection (e.g., individual days during week-long continuous monitoring with actigraphy). For tabular data, a record is equivalent to a row in a table/database.
Item	An individual value in a tabular data record (e.g., question #5 on the WHOQOL-BREF assessment for one participant at one study timepoint, number of hours slept measured by actigraphy on one night for one participant), or a neuroimaging scan (e.g., fMRI scan during one task for one participant at one timepoint).
Data Capture and Creation	
Integrated Discovery Program	A collaborative research program funded by the Ontario Brain Institute. Integrated Discovery Program members fall into the category of Data Producers.
Data Producer	An investigator and their delegate(s) affiliated with a participating institution. Data Producers are responsible for creating or capturing data that are uploaded to Brain-CODE.
Data curation	The management of data throughout its lifecycle (acquisition to archiving) to enable reliable use and retrieval for research purposes (modified from DAQCORD).
Minimally curated	A dataset that is processed to minimize errors and missing information.
Data Quality Rules	REDCap module that conducts item-level missingness and validation analysis on a project's data.
Data Resolution Workflow	A REDCap module that functions with the Data Quality Rules tool and enables users to open a workflow for tracking and documenting the process of resolving issues with data in the project.

List of Acronyms

Key Acronyms

BIDS – Brain Imaging Data Structure

CSV – comma-separated value

DP – Data Producer

DQF – data quality framework

DRW – Data Resolution Workflow (REDCap tool)

(e)CRF – (electronic) case report form

EDC – electronic data capture [tool]

FAIR – Findability, Accessibility, Interoperability, Reusability

IDP – Integrated Discovery Program

OBI – Ontario Brain Institute

OHIP – Ontario Health Insurance Plan

PHI – personal health information

QA – quality assurance

QC – quality control

REB – research ethics boards

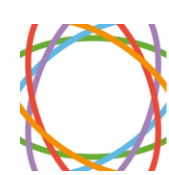SOP – standard operating procedure

Other Acronyms

ADAM – Analysis Data Model

BAM – binary alignment map

CDISC – Clinical Data Interchange Standards Consortium

CDASH – Clinical Data Acquisition Standards Harmonization Implementation Guide

CRAM – compressed reference-oriented alignment map

CT – computed tomography OR controlled terminology (CDISC ontology)

DICOM – digital imaging and communications in medicine

EDF – European data format

FDA – Food and Drug Administration

FLAIR – fluid-attenuated inversion recovery

fMRI – functional magnetic resonance imaging

GA4GH – Global Alliance for Genomics and Health

HGNC – HUGO Gene Nomenclature Committee

HPO – Human Phenotype Ontology

LOINC – Logical Observations Identifier Names and Codes

MedDRA – Medical Dictionary for Regulatory Activities

MEG – magnetoencephalography

MONDO – Mondo Disease Ontology

MRI – magnetic resonance imaging

NCBI – National Center for Bioinformation Technology

NCI GDC – National Cancer Institute Genomics Data Commons

NGS – next generation sequencing

NIFTI – Neuroimaging Informatics Technology Initiative

NIH – National Institutes of Health

PET – positron emission tomography

SDTM – Standard Data Tabulation Model

SNOWMED (CT) – Systemized Nomenclature of Medicine (Clinical Terms)

UO – Units of Measurement Ontology

UTF – Unicode transformation format

VCF – variant call format

WHOQOL-BREF – World Health Organization Quality of Life (Brief)