# OBI Data Quality Framework

*Stage 2 – Recommended Vocabularies*

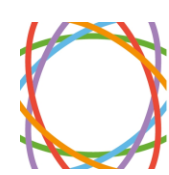Version: 1.2 - Date: August 8, 2024

## 1. Recommended Vocabularies

The following is a list of health research-related controlled vocabularies and ontologies that *should* be used for OBI related data descriptions, study descriptions, and data management descriptions. Adopting vocabularies and ontologies will support the findability, accessibility, interoperability, reusability (FAIR) of the data that are collected and stored on OBI's Brain-CODE platform.

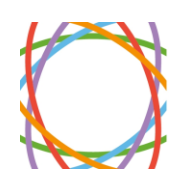| Name | Description | URL | Use cases | Use on Brain-CODE |
|---|---|---|---|---|
| **Disease Ontologies** | | | | |
| **MedDRA (Medical Dictionary for Regulatory Activities)** | MedDRA is an international medical terminology with an emphasis on use for data entry, retrieval, analysis, and display. | https://bioportal.bioontology.org/ontologies/MEDDRA | • Diagnosis labelling in medical history forms | • Medical History |
| **MONDO (Mondo Disease Ontology)** | Mondo Disease Ontology by the Monarch Initiative. The Mondo Disease Ontology (Mondo) aims to harmonize disease definitions across the world. It is a semi-automatically constructed ontology that merges in multiple disease resources to yield a coherent merged ontology. | https://bioportal.bioontology.org/ontologies/MONDO | • Adopted by GA4GH Phenopackets[1] | |

---

[1] GA4GH Phenopackets Recommended Ontologies: https://phenopacket-schema.readthedocs.io/en/latest/recommended-ontologies.html

| | | | | |
|---|---|---|---|---|
| **LOINC (Logical Observations Identifier Names and Codes)** | Laboratory and Clinical Observations. LOINC is used to code questions in electronic medical records (e.g., "what organism is in this blood sample?") | https://bioportal.bioontology.org/ontologies/LOINC | • Used for lab information | |
| **SNOMED Clinical Terms** | SNOMED Clinical Terms. SNOMED is Typically used to code for clinical terminology values (e.g., "Proteus mirabilis") | https://bioportal.bioontology.org/ontologies/SNOMEDCT | • Used to code clinical diagnoses and other clinical terminology where MedDRA cannot be used (LOINC to be used for lab results) | |
| **CDISC Controlled Terminology** | Clinical Data Interchange Standards Consortium (CDISC) collection of controlled terminology (CT) primarily used for registered clinical trials. | https://www.cdisc.org/standards/terminology/controlled-terminology | • CDISC CT for Protocol Representation Model (PRM) for registered trial planning.<br>• CDISC CT for CDASH to store data for registered trial submissions.<br>• CDISC CT for Standard Data Tabulation Model (SDTM) for registered trials reporting.<br>• CDISC CT for ADaM for registered trial analysis reporting | |
| **Phenotype Ontologies** | | | | |
| **HPO (Human Phenotype Ontology)** | The HPO provides a comprehensive logical standard to describe and computationally analyze phenotypic abnormalities found in human disease. | https://hpo.jax.org/app/ | • Phenotypic abnormalities in human disease<br>• Adopted by GA4GH Phenopackets | |
| **Genetics Ontologies** | | | | |
| **HGNC** | The HGNC provides standard names, | https://www.genenames.org/ | • A comprehensive collection of gene | |

| | | | | |
|---|---|---|---|---|
| **(HUGO Gene Nomenclature Committee)** | symbols, and IDs for human genes. | | • names, symbols, and ids for human genes<br>• Adopted by GA4GH Phenopackets | |
| **NCBI Gene** | NCBI gene vocabulary. | https://www.ncbi.nlm.nih.gov/gene/ | • An NIH catalogue of genes | |
| **NIH GDC Schemas** | Standard field descriptors from the NIG Genomic Data Commons (GDC). | https://docs.gdc.cancer.gov/Data_Dictionary/viewer/ | • Lab protocol standard field descriptors for genomic data | |
| **Units** | | | | |
| **UO (Units of Measurement Ontology)** | The Units of Measurement Ontology (UO) provides terms for units commonly encountered in medical data. The following table shows some typical examples. https://pubmed.ncbi.nlm.nih.gov/23060432/ | https://bioportal.bioontology.org/ontologies/UO | • A comprehensive collection of unit terms<br>• Adopted by GA4GH Phenopackets | |
| **Medications** | | | | |
| **DrugCentral** | DrugCentral integrates a broad spectrum of drug resources related to chemical structures, biological activities, regulatory data, pharmacology and drug formulations. https://pubmed.ncbi.nlm.nih.gov/33151287/ | https://drugcentral.org/ | • A comprehensive collection of medication information<br>• Adopted by GA4GH Phenopackets | |
| **Tasks** | | | | |
| **Cognitive Atlas - Tasks** | The Cognitive Atlas is a collaborative knowledge building project that aims to develop a knowledge base (or ontology) that characterizes the | https://www.cognitiveatlas.org/tasks | • A comprehensive collection of cognitive tasks commonly used in neuroscience, psychology, and cognitive science | |

| | state of current thought in cognitive science. The project is led by Russell Poldrack, Professor of Psychology at Stanford University. | | | |
|---|---|---|---|---|

## 2. Recommended Tools to Support Semantic Coding

Please find below a list of tools that can help with semantic concepts and code lookups to support coding.

- o **Athena by OHDSI**
  - o https://athena.ohdsi.org/
  - o Athena is a powerful tool to reverse look up terms and concepts and find controlled vocabularies or ontologies that have definitions for them. It can also expand the term to show related concepts in its contextual hierarchy.
- o **SciCrunch & Interlex**
  - o SciCrunch General: https://scicrunch.org/scicrunch/data/search#all
  - o Interlex: https://scicrunch.org/scicrunch/interlex/dashboard
  - o Neuroimaging: https://scicrunch.org/nidm-terms
  - o Interlex on SciCrunch (by the FAIR data informatics lab) is another powerful tool to look up terms and concepts to find appropriate vocabularies. Various research communities have generated vocabularies now listed on Interlex. Interlex has its own "ilx" coded terms but also shows mapping terms to other vocabularies. Neurolex was integrated into Interlex and contains a large number of brain-related concepts.

## 3. Ontology Databases

Please find below a list of useful ontology concept databases. Additional databases will be added in the future.

- Ontobee: https://ontobee.org/
- Bioportal: https://bioportal.bioontology.org/
- EBI OLS: https://www.ebi.ac.uk/ols/