ONTARIO BRAIN INSTITUTE | INSTITUT ONTARIEN DU CERVEAU

OBI Data Quality Framework

Stage 1 – Study Design, Training, and Testing

Version: 1.2 - Date: July 17, 2024

1. Introduction

OBI seeks to ensure the highest possible quality from data produced by its sponsored research studies and collected on Brain-CODE. OBI aims to maximize opportunities by sponsored research teams and external groups to conduct multi-modal, multi-site, cross-domain studies to generate new discoveries previously impossible to achieve. By serving the broader research communities with outstanding data that meet FAIR[1] principles, OBI aims to establish trust and attract further collaborations in its mission for brain discovery and patient impact. By leveraging experience gained by OBI's Integrated Discovery Programs (IDPs) since 2010, OBI has developed principles and guidelines to maximize the quality, trustworthiness, and usability of data that will be collected in its next phase of discovery. This Data Quality Framework (DQF) aims to capture required and recommended data practices to achieve this goal.

This document is Stage 1 of two stages. Stage 1 addresses study design, training, and testing, while Stage 2 addresses study run-time and post-collection activities. This document should be used by any Data Producer (including IDPs) that will be contributing data to OBI's Brain-CODE platform. Items stated with "***must***" indicate a process or output that is mandatory to complete

[1] Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

by Data Producers such as OBI's Integrated Discovery Programs (IDPs). Items stated with "***should***" or "***recommends***" indicate highly recommended processes or outputs, but variations can be implemented in consultation with OBI.

For any questions regarding this Data Quality Framework please contact: help@braincode.ca

2. Principles of Data Quality

A variety of expert groups in health informatics data management have developed principles to achieve a high level of data quality. One of the most comprehensive frameworks was established by Brown et al. who propose the following dimensions as guiding principles[2][3]; we denote these as the Data Quality Guiding Principles:

1. **Accuracy**: Data are correct and reflect the truth.
2. **Reliability**: Data are consistently collected and entered in a standard way across data collectors.
3. **Timeliness**: Data are current to routine data entry and available for near real-time reporting.
4. **Completeness**: There are no missing essential data elements.
5. **Precision**: Data have necessary detail to address study management and research questions.
6. **Integrity**: Data are secure and protected from bias or manipulation.

OBI's Brain-CODE platform was designed to collect, store, and analyze multi-modal data from a broad range of brain disorders. Multiple data modalities are collected on Brain-CODE: clinical data, imaging data, molecular/genomic data, neuropsychological, eye-tracking, etc. Other modalities and data types can also be collected on Brain-CODE such as wearables data and

[2] Brown W. 2007. Data quality assurance tool for program level indicators. MEASURE Evaluation. (Reviewed in Gass et al. 2017).

[3] Often, other terms are used in the literature to denote similar concepts, for instance "Concordance" and "Reliability" are often used interchangeably, as are "Accuracy" and "Correctness", and "Timeliness" and "Currency" for instance (for a systematic review see Weiskopf & Weng, 2013). Generally, 5 to 6 key dimensions as the ones listed above are consistently used in the literature.

population survey data, among others. Each modality may have numerous data types and file formats under use. Gass et al. identified steps to achieve quality principles in research studies[4].

The following sections outline key steps that will help OBI-supported research initiatives achieve best in-class standards to achieve the six (6) Data Quality Guiding Principles. By maximizing good practices in data collection and curation with these principles, Data Producers via Brain-CODE can achieve excellence in data collection, sharing and use in accordance with the internationally recognized principles of findability, accessibility, interoperability, and re-usability (FAIR[1]). By contributing data to Brain-CODE research studies should, as a result, gain recognition and reputation as highly impactful in increasing neuroscientific knowledge and supporting the improvement of patient health.

3. Protocol Design and Required Data Elements

The following requirements ***must*** be considered as part of the initial study design phase:

1. Data Producers ***must*** include all elements defined in the OBI Protocol Guidelines to the extent possible. The protocol and consent forms ***must*** be shared with OBI for review prior to ethics submission to help ensure that appropriate language is used.
    - See Stage 1 document #2: "OBI DQF Standard Protocol Requirements"

2. Data Producers ***must*** adopt the OBI Common Data Elements (CDEs) in accordance with the Common Data Elements Guidelines (Stage 1 document #3). These guidelines will be developed in consultation with Data Producers (e.g., IDPs). Data Producers are also expected to add their own measures and modalities to meet data collection and research objectives. Participant burden should be considered. Data Producers must adopt measures that allow collected data and scores to be shared (note: norms and documents used for scoring will not be shared by OBI).
    - See Stage 1 document #3: "OBI DQF Common Data Elements Guidelines"

3. Data Producers ***must*** notify OBI if their study will be submitted for clinical trial approval (CTA) to Health Canada or other international regulatory bodies.

[4] Gass, J.D., Misra, A., Yadav, M.N.*S. et al. 2017.* Implementation and results of an integrated data quality assurance protocol in a randomized controlled trial in Uttar Pradesh, India. *Trials* **18,** 418 (2017). https://doi.org/10.1186/s13063-017-2159-1

- When Data Producers are planning to conduct a Health Canada- or FDA-registered clinical trial, studies ***should*** be designed in accordance with existing guidelines[5] and ***should*** consider adopting CDISC Standards. Data Producers in non-registered trials ***should*** also adopt Good Clinical Practices and related standard operating procedures (SOPs).
- Materials are available via the N2 Canada resources site and training can be achieved by providers such as CITI Program.
- Please contact OBI if you require further assistance with data collection for clinical trials.

4. Implementing a Monitoring and Evaluation Team

The following requirements must be considered as part of the initial study development phase. Data Producers must implement a data collection, monitoring and evaluation team that will provide robust training and facilitate clear and efficient operations and communications.

4.1 Recommended Team Structure for Data Producers

1. Data Producers ***must*** appoint a Neuroinformatics Lead who will act as a "central coordinator" for all data collected across sites and modalities for a particular Data Producer and its studies. They will report to the Data Producer's Program Manager and liaise with OBI & Indoc Neuroinformatics teams.

2. Data Producers ***must*** ensure all staff collecting data are trained on data collection protocols and standard operating procedures (SOPs).
    - Training and best practice ***should*** cover how to administer surveys, scoring of surveys (if applicable), the process for data entry, and how to report cases and issues to supervising staff.

3. Neuroinformatics Leads with support from OBI ***must*** administer EDC training sessions covering topics including data collection, quality assurance, and quality control in accordance with the following training materials:

[5] https://www.canada.ca/en/health-canada/services/drugs-health-products/compliance-enforcement/good-clinical-practices/guidance-documents/guidance-drugs-clinical-trials-human-subjects-gui-0100/document.html

- See Stage 1 document #6: "OBI DQF Data Collection, Assurance and Quality Training"
- For Data Producers collecting data on their own systems, please ensure training is provided according to your own institutional policies.

4. Each data collection site ***should*** have site Coordinators/Research Assistants/Technicians be in direct contact with the Neuroinformatics Lead.
    - Site coordinators ***should*** report on site specific study activities and data collection progress on a weekly basis to the Neuroinformatics Lead.

5. Program Managers and Neuroinformatics Leads ***must*** be in regular contact with Site Leads and OBI in order to coordinate study activities, report on progress, and resolve issues.

6. Neuroinformatics Leads ***must***:
    - Monitor progress towards set targets as defined in Data Quality Framework and study protocol. In the cases of direct data capture on Brain-CODE, OBI will facilitate study progress monitoring via web dashboards.
    - Provide supportive supervision to data collection staff.
    - Perform monitoring of data on a regular basis (e.g., weekly)
    - Coordinate quarterly preliminary analyses to ensure proper data coding.
    - Be involved in any cross-program projects or initiatives.
    - Support site teams with data monitoring, locking and transfer to Brain-CODE when necessary.

4.2 Roles and Responsibilities

The following table describes the roles and responsibilities across Data Producer (DP) roles and OBI roles (including Indoc Research roles).

Legend
DP
DP/OBI/Indoc
OBI/Indoc

Role & Responsibility	Person/Team Responsible
Planning and Communications	
1. Develop and maintain DQF	OBI/Indoc Neuroinformatics teams
2. Inform and train DPs on DQF	OBI Neuroinformatics team
3. Develop and maintain study protocol	DP Research team
4. Develop and maintain a Data Management Plan (DMP) [see Stage 1 document #4]	DP Neuroinformatics Lead
5. Develop and maintain study SOPs	DP Neuroinformatics Lead
6. Develop, test, and deploy eCRFs, imaging pipelines, genomic pipelines, etc.	DP Neuroinformatics team / OBI / Indoc
7. Share study protocol, DMP, and initially completed Study Quality Checklist (SQC) to OBI for review and approval.	DP Neuroinformatics Lead
8. OBI review of protocol, DMP, and DQC and adjust with DP	OBI Neuroinformatics / DP Neuroinformatics Lead
Training and Testing	
1. Train DPs on Electronic Data Capture (EDC) systems	OBI/Indoc Neuroinformatics teams
2. Train study coordinators and technicians across all sites on study Protocols, DMP and SOPs	DP Neuroinformatics Lead, DP Site Clinical Managers, DP Site Technical Leads
3. Initial and ongoing testing of EDC systems and quality assurance procedures	DP Neuroinformatics Lead, DP Site Clinical Managers, DP Site Technical Leads
Study Operations and Monitoring	
1. Complete the Study Quality Checklist as study progresses and share updated version with OBI	DP Neuroinformatics Lead
2. Share REB approvals/amendments with OBI prior to recruitment	DP Neuroinformatics Lead
3. Oversee weekly data collection and monitoring by coordinators and technicians (imaging, molecular, etc.) including resolution of any issues	DP Neuroinformatics Lead
Data Locking and Curation	
1. Locking monitored data on a quarterly basis or as required per modality	DP Neuroinformatics Lead
2. Locked data extraction on Brain-CODE (or transfer to Brain-CODE)	DP Neuroinformatics team, OBI / Indoc Neuroinformatics teams
3. Curate the data to achieve “minimal curation” level as agreed upon with OBI	DP Neuroinformatics team, OBI / Indoc Neuroinformatics teams
4. Review Data Quality Report after curation completed [see Stage 2 document #3]	DP Neuroinformatics Lead, OBI / Indoc Neuroinformatics teams
5. Resolve issues from Data Quality Report, if any	DP Neuroinformatics team
6. Manage Brain-CODE curation software, tools, and data access	OBI / Indoc Neuroinformatics teams

Table 1. Roles and Responsibilities across Data Producer and OBI roles.

5. Establishing a Data Management Plan & Quality Checklist

The following requirements ***must*** be implemented as part of the final study design phase:

1. For each study, Data Producers ***must*** create a standard Data Management Plan (DMP) that will be implemented and adopted. The DMP ***must*** be shared with OBI for review and to inform of the study data management plan. Any updates to the study protocol ***must*** be reflected in the DMP and the updated version ***must*** be shared with OBI. The DMP will clearly outline the data to be collected, documentation and metadata created, the storage and backup of the data during studies, preservation of data in the long term, data sharing and reuse plans, responsibilities and resources for data management, as well as the ethics and legal compliance in place. Establishing a DMP aligns with the CIHR guidelines for research management and a DMP template was designed in alignment with the federal Digital Research Alliance recommendations. For Data Producer studies, many aspects related to storage, preservation, data sharing, compliance and shared responsibilities are provided by OBI's Brain-CODE platform and services. OBI has developed a DMP template with guidance from CIHR and the Portage Network[6]. Please refer to the following Data Management Plan Template to be completed for each study:
    - See Stage 1 document #4: "OBI DQF Data Management Plan - Template"

2. For each study, Data Producers ***must*** complete the OBI Study Quality Checklist (SQC) based on the DAQCORD criteria[7] at each stage of their study and **submit this checklist to OBI** as it becomes updated at each stage of the study (Stage 1 and Stage 2). This checklist will help Data Producers and OBI verify that key criteria for data management have been established by all teams.
    - See Stage 1 document #5: "OBI DQF Study Quality Checklist - Template"

[6] Portage Network. (2020, August 25). Primer - Data Management Plan. Zenodo. https://doi.org/10.5281/zenodo.4495631

[7] Ercole, A., Brinck, V., George, P., Hicks, R., Huijben, J., Jarrett, M., . . . Wilson, L. (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

6. Support

If you have specific informatics questions regarding your research project, please contact the Neuroinformatics Lead and/or Program Manager for your program. For any general questions or support requests regarding this Data Quality Framework please contact help@braincode.ca.