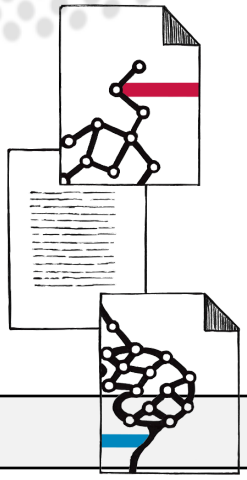
Metadata Curation



Version	Author(s)
1.0	Kahraman, K. Das, D., Sripad, P., Brown, T., Oh, H., Melloni, L., Bonacchi, N.
Date	**Editor(s)**
14.05.2024	Kahraman, K. Das, D., Sripad, P., Brown, T., Oh, H., Melloni, L., Bonacchi, N.

TABLE OF CONTENTS

Introduction

This document provides comprehensive guidance on the procedures for extracting and handling the metadata associated with the COGITATE project. For M/EEG and fMRI participants, metadata encompassed Case Report Forms (CRFs) and Exit Questionnaires (ExQus). For intracranial EEG (iEEG) studies, metadata encompass CRFs, ExQus, Multi Disciplinary Conference Files (MDCs), and Pathology files. While CRFs and ExQus directly relate to the experimental procedures, MDCs and Pathology files provide information into the epilepsy conditions of participants.

Description of Experiment Related Metadata Files

Case Report Form (CRF)

Case Report Forms (CRFs) are documents filled out by researchers at various stages: before, during, and after the experiment. These forms are used to confirm the readiness of the experimental setup, including ensuring that unnecessary devices are disconnected, trigger channels for photodiodes are set, and devices collecting ECoG data are properly configured. The version of the experimental code used is also recorded in the iEEG CRFs.

In addition to setup details, CRFs capture information about the participants. This includes, but is not limited to, demographic data, and the participant's condition such as feelings of tiredness or pain. Observations made during each experimental run are noted, which may include any comments from the patient or anomalies observed while running the study.

The structure of CRFs typically covers several areas. These areas document the setup of basic equipment, trigger settings, and the configuration of amplifiers and montages. Initial settings for eye trackers and screen brightness are adjusted and recorded before the experiment starts. Researchers document how instructions are given to the participants and confirm that the participant has understood them. The calibration of the eye tracker is also noted. Details concerning the timing of the experiment, events occurring during the experiment, and any social interactions involved are recorded. Observations about the participant's status during the test and additional patient information relevant to the experimental conditions are included.

For M/EEG experiments, notes on potentially faulty triggers and electrodes are recorded. For fMRI, notes on the runs are included. Both types of CRFs document whether the participant moved excessively or exhibited any other behavior that could affect the data quality.

The outlined structure of the CRFs was not fixed and deviations in the details regarding, for example, run or demographic information took place.

Exit Questionnaire (ExQu)

The Exit Questionnaire (ExQu) is a document that participants complete after the experiment to gather feedback on their experience during the study. It consists of several questions that cover various aspects of the experimental process, with many of the responses recorded using a Likert scale.

The questionnaire includes questions to assess the difficulty of staying focused throughout the experiment and whether participants noticed variations in the duration of stimuli. Specific questions inquire if participants were able to maintain focus on stimuli presented for short, medium, or long durations, although some responses to these questions were not provided.

Participants were asked to rate the difficulty of keeping targets in memory for the duration of a block and to evaluate the ease of recognizing different categories of stimuli, such as faces, objects, letters, and symbols. Additionally, participants responded to whether certain orientations of stimuli were harder to recognize than others and provided details on which ones, if any, were more challenging.

The questionnaire also sought opinions on the length of the blocks and whether a less variable stimulus duration would have made the task easier. Participants had the opportunity to add any additional comments regarding their experience, though this was not mandatory and some chose not to provide further comments.

Description of Epilepsy Related Metadata Files

Multi Disciplinary Conference (MDC)

The Multi Disciplinary Conference files compile detailed records from discussions among neurologists, surgeons, and other healthcare professionals who review and plan the treatment approaches for epilepsy patients involved in the study. These files provide information about the epilepsy condition of participants. Specific details include patient histories, diagnostic considerations, treatment options, and consensus on therapeutic strategies.

Pathology File

Pathology Files contain records of any pathological examinations conducted, including descriptions of resected brain tissue from participants undergoing surgery for epilepsy. These files offer information about the location and details of the brain resections performed.

iVEEG File

Some subjects had iVEEG (intracranial video EEG) files, while others did not. These files contain notes written by doctors during the recording for clinically relevant information. Furthermore they entailed electrode information such as number and implant hemisphere but also recording type (sEEG or ECoG). For subjects without an iVEEG file, information about

number of electrodes, implant hemisphere as well as recording type was provided from the teams.

Curation Process of Metadata Files

The main purpose of the curation process was to ensure that the metadata provided along with the Cogitate datasets are anonymous yet comprehensive. The metadata was checked for PHI information and either manually removed, or if possible, ignored by scripts to not get extracted. They are provided in non-proprietary easily readable formats such as CSV, TSV or JSON.

Curation of CRFs

The CRFs were originally in .docx or .pdf formats. A standardized template was created to include the most important information from the various types of CRFs. This template was then manually filled out for each subject to enable further processing.

The CRFs for every individual subject / session were stored under *CURATED_RAW_ROOT/SUBJECT/SESSION/METADATA/SUBJECT_CRF.json* for RAW data and under *BIDS_ROOT/derivatives/additional_metadata/sub-SUBJECT/METADATA/SUBJECT_CRF.json* for BIDS formatted data.

Curation of ExQus

The ExQus were mostly in .csv format and occasionally in .pdf format. If the ExQus were in .pdf format, the .csv template was filled out, and a .csv file was created for the subject. A script was used to extract the information from these files and convert it into a JSON format, which is a dictionary with the questions as keys and the answers as values. These JSON files were then combined with the CRFs into a consolidated file for all subjects to allow further processing.

The EXQU for every individual subject and sessions were stored under *CURATED_RAW_ROOT/SUBJECT/SESSION/METADATA/SUBJECT_EXQU.json* for RAW data and under *BIDS_ROOT/derivatives/additional_metadata/sub-SUBJECT/METADATA/SUBJECT_EXQU.json* for BIDS formatted data.

Curation of MDCs and Pathology Files

The MDC and Pathology files were in .docx format. Important demographic information and details about epilepsy were collected from these files and compiled into an Excel spreadsheet. These details were then converted into a JSON file allowing further processing.

The demographics and epilepsy information were included along with the data according to the release format.

The raw data format includes all of the demographics information as a CSV file (ECOG_demographics.csv) under the METADATA directory. Subject specific demographics information is provided as a JSON file under individual subject directories.

CURATED_RAW_ROOT/
METADATA/
MODALITY_demographics.csv
SUBJECT/
METADATA/
SUBJ_demographics.json

The BIDS data format included the demographics data in TSV format as specified by BIDS along with a sidecar JSON file. The data is split into two TSV files - participants.tsv and participants_epilepsy.tsv.

- The participants.tsv file consists of the demographics related columns including *'participant_id', 'sex', 'age', 'handedness', 'weight', 'height', 'secondary_language', 'race', 'primary_language', 'education', 'colorblind', 'visualcorrection', 'eyedominance', 'eyechartresults' and 'dioptry'.*
- The participants_epilepsy.tsv file consists of the epilepsy related columns including *'electrode scheme', 'number of implanted electrodes', 'implant hemisphere', 'iq [value, test]', 'wada', 'seizure type', 'age of onset', "compensation", "fluent_in_english (yes/no)", "visual_corrected_to_normal (yes/no)", "visual_lens_power_left", "visual_lens_power_right", "auditory_normal_hearing (yes/no)", "epilepsy_seizure_classification", "epilepsy_seizure_aura", "epilepsy_seizure_semiology", "epilepsy_seizure_frequency", "epilepsy_post_ictal_semiology", "epilepsy_trigger", "epilepsy_duration_uncontrolled", "epilepsy_seizure_onset_zone", "epilepsy_resection", "epilepsy_language_lateralization", "epilepsy_past_surgical_history", "epilepsy_past_medical_history", "epilepsy_family_history", "other_neurological_disorders", "epilepsy_mri_findings", and "epilepsy_pathology_findings".*

BIDS_ROOT/
participants.tsv
participants_epilepsy.tsv
participants.json
participants_epilepsy.json
derivatives/
additional_metadata/
sub-SUBJ/
METADATA/
SUBJ_demographics.json

Additional Metadata Files

Additional metadata includes information describing various aspects of the experiment such as analysis, protocols, tasks, devices and wiring. This information is represented in JSON format in both RAW and BIDS data for analysis, devices, protocols and tasks. The wiring diagram of the particular setup is made available in PDF format.

Analysis consists of the analysis steps, analysis order, links to similar analysis and location of the analysis code.

Devices consist of a list of devices from every lab involved in the experiment. The various measurement devices and connectors used as listed. This closely ties to the wiring diagram of individual labs that are provided as PDF.

Protocols include links to the data measurement protocols used to obtain the data. Mostly these are the Standard Operating Procedures followed by the labs during measurement.

Tasks describes the experiment in a condensed manner including links to a detailed description of the task and various stimuli and responses involved.

CURATED_RAW_ROOT/
METADATA/
analysis.json
devices_ECOG.json
wiring_ECOG.pdf
protocols.json
tasks_EXP1_ECOG.json
tasks_FingerLoc_ECOG.json

BIDS_ROOT/
derivatives/
additional_metadata/
METADATA/
analysis.json
devices_ECOG.json
wiring_ECOG.pdf
protocols.json
tasks_EXP1_ECOG.json
tasks_FingerLoc_ECOG.json

Code

The code used to extract, transform and load (ETL) the metadata is included in the Curation repository at https://github.com/Cogitate-consortium/cogcurate. The metadata ETL code is closely integrated with the curation process and should be run as such.