



# **PRED+CT: A Patient Repository of EEG Data & Computational Tools**

## **Statement of Procedures**

James F. Cavanagh, PhD  
Assistant Professor  
Department of Psychology

Abdullah Mueen, PhD  
Assistant Professor  
Department of Computer Science

University of New Mexico

The *Patient Repository for EEG Data + Computational Tools* (PRED+CT) project is led by Dr. James F. Cavanagh and Dr. Abdullah Mueen at the University of New Mexico. At the present time, the PRED+CT website will be hosted by the UNM computer science department (<http://predict.cs.unm.edu/>). This site will not only be the first open-source EEG database for patient data, but it will facilitate the standardization of assessment and analytic tools towards the overarching goal of collaborative data collection and open data mining.

PRED+CT will have three major components, each of which will be fully open-source and downloadable to any user:

- 1) A suite of configurable and extendable software executables for task administration and EEG triggering.
- 2) Data storage of patient and control group EEG files.
- 3) Matlab-based pre-processing and classification tools for discriminating patients from controls





### **What about human subjects issues?**

Users who wish to share their data using PRED+CT should first receive approval from their IRB for sharing of de-identified data. Most IRB's do not consider data sharing itself to fall under the definition of "human subjects research", but interested users should request such a determination from their IRB before sharing data. Including such language in the informed consent is the best way to ensure that ethical issues are well-managed; however this step may not be necessary depending on the user's IRB. In addition, users of the data contained in the repository will make no effort to re-identify individual level data.

### **How are the data licensed?**

Unless otherwise noted, this database and its contents are made available under the Public Domain Dedication and License v1.0 whose full text can be found at: <http://www.opendatacommons.org/licenses/pddl/1.0/>. We hope that all users will follow the [ODC Attribution/Share-Alike Community Norms](#); in particular, while not legally required, we hope that all users of the data will acknowledge the PRED+CT project in any publications.

### **How will data be uploaded?**

To upload data, users will need to create a login account (username, password, email, institution and position). PRED+CT administrators will receive all login requests and initiate email contact with a dropbox link. Users will then upload their data files (in either native EEG system formats or as .mat or .set files) to this dedicated dropbox folder.



PRED+CT administrators will download the data and verify the following:

- 1) No user-identifying information or protected health information is included and the data are fully de-identified. Specifically, each data file should have a unique alpha-numeric or numeric ID that does not contain any information like participant initials or date of data collection. If these requirements are not met, the data may be re-labeled by PRED+CT administrators. Examples for appropriate labeling include:
  - a. Patient\_1, Patient\_2, Control\_1 ...
  - b. Anxiety\_1, Anxiety\_2, Control\_1 ...
  - c. S1, S2, S3, S4 ...
  - d. S1\_ONmeds, S1\_OFFmeds, S1\_Ctl ...
- 2) General experiment information will be required, including at least:
  - a. Lead investigator / team / location and contact information
  - b. Patient type
  - c. Number of patients / controls
  - d. Task, including a description of the trigger types
  - e. EEG system / number of channels / sampling rate / reference
  - f. Any pre-processing already completed
  - g. Publication link (if possible)
  - h. Funding source (if relevant)



- 3) Information will need to be included for each data file. The following fields will be necessary:
  - a. Patient or control
  - b. Age
  - c. Sex
- 4) Additional information will be welcome provided it does not contain user-identifying information. Examples may include:
  - a. Symptom scores
  - b. Demographics like socio-economic status, years of education, race and ethnicity
  - c. Neuropsychological test outputs
  - d. Session number, if longitudinal (note that exact dates will not be allowed)
- 5) The information in items #3 and #4 above could be included as an .xls or a .mat file, or could be integrated into the eeglab format EEG array as a subfield (i.e. EEG.demographics.sex, EEG.demographics.age, etc.).
- 6) PRED+CT administrators will curate the donated data and upload each file as a .mat file with EEGLab structure (i.e. the EEG. Array). This step is not only beneficial to present all data with a *lingua franca* of Matlab, but it is necessary to remove all dates and other potentially identifying information which can reside in header files. All datasets will include the Matlab scripts used to translate the native EEG formats into .mat files.
- 7) In addition to the data files, relevant task files (Eprime, Matlab, PsychoPy, Presentation, etc.) or relevant pre-processing scripts will be welcome additions.

### **How do I cite PRED+CT?**

We anticipate a *Frontiers in Neuroinformatics* submission in early summer 2017.

### **How is PRED+CT funded?**

This work was conducted as part of a Brain Initiative Seed Award supported by the University of New Mexico Office of the Vice President for Research.