BRAINBASE: A RESEARCH AND DATA MANAGEMENT PLATFORM FOR HUMAN EEG

Tara Thiagarajan

Sapien Labs, Arlington, VA
tara@sapienlabs.org

***Abstract*— Human beings have a wide range of personalities, proclivities and cognitive capabilities, some of which may be genetically determined, but much of which is shaped by experience. These diverse differences reflect in various ways in the fine scale structure and dynamics of the human brain. It is therefore important to study the human brain in larger sample sizes across diverse human populations. This paper describes v1.0 of BrainBase, a cloud based research platform that enables large scale data management and collaboration for EEG studies with the view of building an open access aggregated database of human EEG at scale. It includes metadata standards for strong experimental controls and cross-dataset comparisons, back-end intelligence to ensure data quality, and the ability to search and compile unique dataset configurations based on keywords. The ease of use of EEG equipment and the framework provided by BrainBase opens opportunities for a wider range of researchers to contribute to human brain science.**

***Keywords*—EEG, BrainBase, database, data standards**

I. INTRODUCTION

A. Human Brain Diversity

Humans differ among each other far more than any other species. We are different in our personalities, inclinations, preferences and capabilities and vary more in the gene expression in our brain than any other species [1]. Furthermore, although the gross architecture of the human brain is genetically encoded, its final shape and function is largely determined by experience [2-4]. The result is that each human brain is fine-tuned in terms of connectivity by its unique history of stimulus. This leads to substantial experience dependent differences in brain structure, dynamics and function. A musician's brain, for example, is organized differently from a dancer's brain, which is different from a London Taxi driver [5, 6].

When considered in a global context, with diverse languages, cultures, social structures, educations, occupations, physical environments, incomes and access and use of technology, our experiences vary widely. For example, a recent study by Sapien Labs found up to 1000X differences in dynamical features across a broad sample from remote villages to urban hubs that were linked to elements of modernization [7, 8]. Such wide difference indicates that there is no average brain and cautions against extrapolating findings from small homogeneous groups to global populations. It also points to a need to build a normative database that encompasses the diversity across humanity. This is essential in order to identify approaches and solutions to brain health tailored to different populations and even to the individual.

B. EEG as a Critical Tool in Human Neuroscience and Neurotech

There are many ways to measure human brain function from fMRI and PET to MEG and EEG. While each has its advantages and disadvantages from the perspective of understanding the human brain, EEG stands out in its potential.

EEG provides a noninvasive view of electrical activity on millisecond timescales at a cost many times lower than any other technology for brain measurement. Its significant cost effectiveness and relative ease of use (device costs ranging from $1,000 to $100,000 compared to several $100,000s to millions for fMRI, PET and MEG) make it more accessible to a larger community of researchers.

Many EEG systems are now easily portable and therefore, unlike any other technology, allow measurement of human brain activity in natural environments rather than within a lab setting. Some examples of cost effective and portable hardware are Emotiv EPOC and OpenBCI, which offer 14 channel systems with good signal quality for less than $1,500.

Furthermore, new emerging approaches to analysis of the EEG signal can deliver deeper insights.

Taken together, this opens up previously unimaginable possibilities beyond any other brain imaging technology today by enabling participation of a broader range of researchers and expanding the scope of experimentation to difficult to reach populations in natural conditions.

C. Moving beyond small sample sizes and across borders

With most studies coming from individual labs, sample sizes in much of the scientific literature are small, on the order of 20 to 50 subjects. This has inherent limitations given the diversity of human populations and the large number of confounds. Many things influence brain activity and how the brain responds to an experimental intervention is conditional on various aspects of the participants state of mind. In small samples, it is impossible to control for the numerous confounds. Consequently, the reliability and extensibility of research results are always in question.

Furthermore, a large majority of studies are focused on college educated subjects in the United States and Western Europe. This sample is largely unrepresentative of the global population (for example, less than 10% of adults globally are college educated) and cautions against extrapolating published results to a global population. There is much to be gained by building a collaborative research community that works together with standardized tools to build large scale datasets representative of a global humanity.

D. The opportunity of scale and big data

Shifting the research paradigm from one of individual labs producing small studies to a network of labs working with data standards that enable cross study comparisons and collaborating to deliver large scale datasets can accelerate and amplify the impact of human neuroscience, and advance the pace of neurotech applications. Large scale datasets allow better controls for inter and intra-person diversity as well as the opportunity for the application of the tools of big data and machine learning for more individually tailored and precise results across a host of applications.

However, scale also requires more effective data management to ensure consistent data quality, data standards, data validation as well as different paradigms for analysis. BrainBase provides a data management system that allows researchers to better manage their individual lab data as well as to collaborate and share data more effectively across borders without extensive efforts around data formatting and cleaning. Importantly, the backend of BrainBase will employ intelligence to ensure that data submitted is of high quality and that associated metadata can be parsed into a searchable database, as well as mechanisms to standardize data.

II. The BrainBase Platform

A. Summary

BrainBase is an open cloud based platform accessible on the web at BrainBase.io that enables management of EEG and associated subject information in one place. The data is stored according to standardized data formats to allow the data to be searchable and easily combined with other datasets for large scale analysis. Beyond this it also provides libraries of protocols and the ability to define, publish and participate in large scale global experiments.

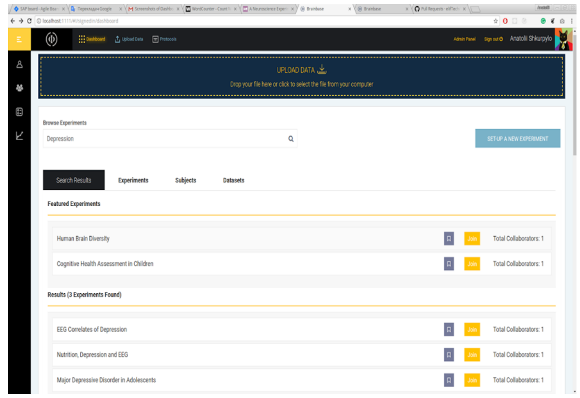

Figure 1. Screenshot of the BrainBase user dashboard

In the final form of this platform users will be able to:

- Manage subject and session information with associated EEG files.
- Search and utilize protocols for a range of cognitive and other tasks from a large library.
- Join experiments in progress around the world.
- Create experiments and new protocols that others on the platform can join and participate in.
- Search and compile datasets according to specific criteria.
- Submit code in Python or R that can be run on large datasets without the need to download the dataset.
- Access automated analysis of EEG files
- Submit results (graphs with descriptions) that are associated with compiled datasets and analysis code.
- Create annotated links between results posted by users to enable faster communication and insights into webs of results.

B. Data Management

Data in the system is of three types – (i) EEG recording files (ii) Metadata related to parameters of the EEG recording and (iii) Subject Metadata.

The standard format for EEG files on BrainBase is the increasingly popular European Data Format (.edf). However, other file formats will be accepted and converted to EDF. The number of file formats accepted will be enhanced over time.

Subject Metadata can be uploaded in multiple formats including excel templates (for multiple subjects) that meet our guidelines or fillable pdfs using our existing templates. These metadata files can be zipped along with the EEG files and dragged and dropped in the database. Alternatively, users can enter data for each subject using our online forms and subject management features.

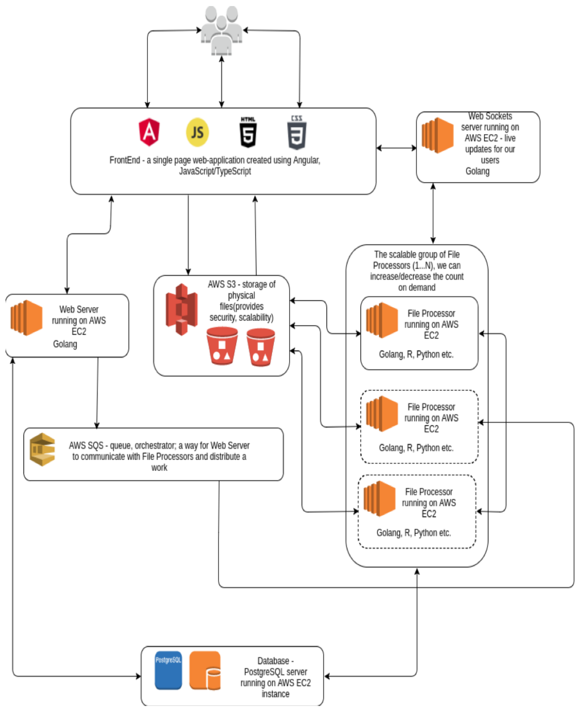

Figure 2. Overview of the architecture

1) *EEG File Handling*

Files uploaded are first checked for supported EEG extensions and then unpacked to extract header information. The header is validated to ensure complete information is available with respect to channels sampled, sampling rate, filters applied and any internal referencing, and read into the database. Users are notified of any incomplete or unreadable information to be corrected.

2) *Metadata*

Metadata typically is of three different types – (i) information on the protocol or task performed by the subject during the EEG, (ii) demographic or other information about the subject and (iii) scores or other task related outcomes. Metadata can be in two formats – Tabular (excel or csv files) or Text files.

a) *Tabular Data*

The database accepts meta data in excel spreadsheets with certain structure and naming protocols. Metadata is checked for content, associated with relevant subjects and/or EEG files and parsed into searchable tables. Mismatches and unreadable data are flagged to the user for clarification. Metadata information is then parsed and stored and made available for search.

b) *Text Files*

Currently BrainBase has the capability to read clinical records provided in text form. This includes reading medications, symptoms and diagnosis based on an initial and growing database of terms and parsing these into tables. Over time, we anticipate natural language processing applications that can extend the scope of text inputs.

3) *Data Parsing*

All uploaded files (or entered subject information) are unpacked and automatically checked for discrepancies. These include (i) EEG files that are not associated with a subject or vice versa (ii) missing information that is essential to interpretation (e.g. information about the protocol/task performed during the recording or the recording parameters such as sampling rate) (iii) unrecognized data formats or data headers in metadata files or (iv) conflicted data. Following this check, an email is automatically generated and sent to the user describing the discrepancy and requesting them to fix it. A work queue in the admin panel also allows BrainBase data support to manually check these data issues and look for ways to fix problems with reading of the data.

EEG files are then checked for association with metadata with respect to the subject and for associated information about the specific task performed during the recording. Absent or unreadable metadata are flagged to the user for resolution.

Data that is not problematic in any of these ways is automatically parsed into the database and becomes searchable.

4) *Data Encryption and Anonymity*

While most subject metadata is searchable, identifying information such as User Name, address and contact information are encrypted and available only to the user who uploaded the information in the account to maintain anonymity. Each subject in the system receives a unique ID that is system generated and serves as the reference for any record downloaded from BrainBase. The form will be available in multiple languages and slightly modified forms so that it can be used across cultures and demographics.

C. *Facilitating Research Design & Collaboration*

Besides simply managing data uploaded to the platform, BrainBase provides a workflow to define, publish and collaborate on experiments. The workflow and features are designed with the purpose of making the process of sound experimental design fast and easy as well as enabling easy collaboration across borders.

1) Creating or Joining an Experiment

Users can define experiments by outlining the goals and methods, defining profiles of multiple participant/study groups and uploading any associated protocols or questionnaires. Once an experiment is published, other users can search experiments by key words and join them. Joining an experiment allows users to contribute data to the experiment, access contributed data and participate in the experiments forum. The workflow ensures that important details necessary to replicate an experimental paradigm are clearly defined making it easy for people to join and participate.

2) Protocols and Survey Tools

BrainBase provides a searchable library of protocols and survey tools that are either freely available on the platform or available for purchase through third parties. These include task applications and tools for measuring various cognitive functions such as attention and working memory, intelligence tests and a host of survey tools covering different aspects of experience and behavior.

Users can also define and add their own protocols to the database for others to use.

3) Metadata Standards and Controls

There are various aspects of individual participants that can vary outside the elements of the experimental paradigm. These include aspects of the individual's demographics, context and capabilities as well as the individual's state of mind and body at the time of the experiment. BrainBase provides standard metadata questionnaires for Subject Information and their State of Mind at the time of the experimental session that cover the important factors in both aspects and can serve as a standard across all experiments. While controlling for so many factors is difficult within small samples, as datasets are aggregated across experiments or expanded by collaboration, the effects of these factors will be more apparent.

a) Demographic Controls

Various aspects of the individual have been found to influence the EEG including income, access/use of technology, nature of occupation and so on. While every aspect of an individual cannot be captured, BrainBase provides a demographic and mental health history form (Basic Subject Information) that is suggested as a standard for all experiments. Specialized forms on other aspects of an individual's life such as technology use, mobility and communication, music training and appreciation and so on have been contributed by users.

b) Session State of Mind

The dynamical features of the EEG change over time within an individual depending on a vast number of factors from mood (e.g. anxiety), blood sugar, psychoactive substance intake (such as caffeine), medications (e.g. ibuprofen) and sleep. However, most studies rarely ask participants about these factors at the time of recording. Differences arising on account of these factors are therefore potential confounds that have not been controlled for. BrainBase provides a standard Session form that covers these aspects and is suggested for all experiments.

c) Experimental Controls

While many experiments may involve EEG during specific cognitive paradigms, the simple paradigms of continuous EEG recordings with eyes closed and open provide a resting baseline or control in various experimental paradigms. BrainBase therefore recommends these as standard base protocols along with all EEG experiments. This also contributes to a standard normative database that can be aggregated across all datasets and can provide important insights on its own as well as deliver a set of normative standards for future use in the field.

D. Search and Download Datasets

The last major feature of BrainBase is the ability to search datasets according to different criteria and create and download uniquely configured datasets that are aggregated across contributors. The search allows selection of recording features (e.g. specifying sampling rate and channels), experimental tasks used and categories of metadata available. Records meeting the criteria, which may span contributions from multiple researchers, are aggregated into a single downloadable dataset. All metadata is aggregated into a single csv file and all EEG recordings are converted to .edf form and available in a .zip file.

As a second phase, search will be expanded to include individual elements of the metadata. For instance, one might search for individuals who have participated in a particular attention task who have consumed caffeine within 2 hours of the experiment and rated their mood as very anxious to create one dataset, and individuals who have consumed caffeine but rated their mood as not anxious to create another dataset for comparison.

III. Conclusions

In conclusion, we propose BrainBase as a global platform for collaborative EEG research to accelerate and amplify human neuroscience research. BrainBase will provide standardized, well annotated EEG data serving both as a source of normative data as well as enabling big data and machine learning approaches that can deliver deep insights into human mental and cognitive health as well as a host of neurotech applications.

Acknowledgment

We thank our partners at UNSAM (Argentina), PHFI (India), Temple University (USA), Macquarie University (Australia) and others who have provided important inputs and perspectives for the development of BrainBase.

References

[1] W. Enard, P. Khaitovich, J. Klose, S. Zollner, F. Heissig, P. Giavalisco*, et al.*, "Intra- and interspecific variation in primate gene expression patterns," *Science,* vol. 296, pp. 340-3, Apr 12 2002.

[2] P. Medini, "Experience-dependent plasticity of visual cortical microcircuits," *Neuroscience,* vol. 278, pp. 367-84, Oct 10 2014.

[3] K. A. Gordon, B. C. Papsin, and R. V. Harrison, "Activity-dependent developmental plasticity of the auditory brain stem in children who use cochlear implants," *Ear Hear,* vol. 24, pp. 485-500, Dec 2003.

[4] C. Xerri, "Plasticity of cortical maps: multiple triggers for adaptive reorganization following brain damage and spinal cord injury," *Neuroscientist,* vol. 18, pp. 133-48, Apr 2012.

[5] E. A. Maguire, D. G. Gadian, I. S. Johnsrude, C. D. Good, J. Ashburner, R. S. Frackowiak*, et al.*, "Navigation-related structural change in the hippocampi of taxi drivers," *Proc Natl Acad Sci U S A,* vol. 97, pp. 4398-403, Apr 11 2000.

[6] J. Sherwin and P. Sajda, "Musical experts recruit action-related neural structures in harmonic anomaly detection: evidence for embodied cognition in expertise," *Brain Cogn,* vol. 83, pp. 190-202, Nov 2013.

[7] D. Parameshwaran, T.C. Thiagarajan, "Alpha Oscillations and Modernization," 2016.

[8] D. Parameshwaran, T.C. Thiagarajan "Complexity of EEG reflects socioeconomic context and geofootprint," *In Review,* 2016.