Actionable event annotation in neuroimaging: HED-3G principles and practice

***K. ROBBINS[1],** D. TRUONG[2,3], A. DELORME[2,3], S. MAKEIG[2,3];[1] Univ. of Texas San Antonio; [2]Univ. of California San Diego, La Jolla, CA; [3]Inst. of Neural Computation, UCSD/INC/SCCN, La Jolla, CA;

Abstract

Introduction: HED-3G is an evolution of the Hierarchical Event Descriptor system to move experiment annotation beyond cryptic keywords and scattered documentation to a standardized annotation of experimental structure that can be directly used by analysis tools to perform sophisticated analyses based on events and other metadata.

Rationale: Event markers associate a time-delimited phase of or change in an identifiable process with a defined time point or period. While event-markers play a central role in organizing neuroimaging and other experimental data, current annotation practice mostly relies on local laboratory strategies for labelling events and other information such as task and experimental design. As a result, data analysts must often resort to assembling information scattered across publications, stored metadata, and internal laboratory documents and must then write custom computer code to perform analyses.

Methods: HED-3G provides a vocabulary, software infrastructure, and tools for detailed event annotation. HED-3G standardizes strategies for deciding which events to report, how to annotate them, and how to incorporate the information into analyses and information extraction tools. Researchers using HED-3G first declare the relevant concepts needed to describe their experiment including presented experimental stimuli, participant responses, tasks, and experimental conditions, and describe the temporal structure of their experiment. Researchers then express this information using a standardized, structured vocabulary. The HED-3G infrastructure provides a mapping from defined concepts to event annotations that enables powerful cross-study analysis. Validation tools available online and as Python and JavaScript modules ensure consistency and compliance with the HED-3G specification. The standardized vocabulary and structure allow researchers to use a set of standard tools, eliminating the need for creating customized code. HED-3G is integrated into the BIDS (Brain Imaging Data Structure) system for organizing and sharing datasets.

Results: We have developed several case studies that demonstrate the effectiveness of HED-3G in annotating a variety of features and experimental concepts in machine-actionable form. Other aspects of the HED-3G ecosystem are addressed in related posters (Truong: tools), (Makeig: extensions), (Delorme: EEGLAB integration) (Pal Attia: clinical use), (Denissen: fMRI).

Conclusions: HED-3G can dramatically improve the utility of neuroimaging and behavioral data. Our challenge is now to encourage researchers to incorporate HED annotation into their new, existing, and shared data.

Keyword (Complete): HUMAN; NEUROIMAGING; EEG; MEG; BOLD; NEUROINFORMATICS
Grant/Other Support: : Swartz Foundation gift; NIH Grant 5R01NS047293-16

Extending HED-3G event annotation using Library schemas

***S. MAKEIG**[1,2], D. TRUONG[1,2], A. DELORME[1,2], K. ROBBINS[3]; [1]University of California San Diego, La Jolla, CA; [2]Inst. of Neural Computation, UCSD, La Jolla, CA; [3]University of Texas San Antonio

Abstract

A major current gap in standards for storing and sharing neuroimaging data is a system for detailed description of experiment events. The Hierarchical Event Descriptor (HED) system, first proposed a decade ago by Nima Bigdely-Shamlo (2012), has recently undergone a major developmental revision to HED-3G (K. Robbins, this meeting), whose architecture and syntax is now based on a root schema (set of supported terms) and syntax, plus a provision for extending the annotation vocabulary established by the root schema using specialized HED Library Schemas. HED Library schemas are designed to eliminate the need for the root HED schema to grow unwieldy, and to promote incorporation of maximally efficient and accepted sets of terms for use by the relevant research communities who build and maintain them. Thereby, HED-3G can be made suitable for annotating all types of time-series data and experiments. For example, annotations of neuroimaging datasets recording brain activity during performance of language tasks will require terms such as noun, verb, dependent-clause, etc., defined in a Language library schema (M. Denissen, this meeting). One HED library schema under development incorporates terms from the international standard SCORE vocabulary for clinical reporting of human scalp EEG data features (T. Pal Attia, this meeting), including terms such as 'sleep spindle', 'interictal spike', 'alpha burst', etc. Four more HED library development projects are now planned or underway: one for experiments involving viewing Moving-Pictures, a second for annotating events in Story-Narratives (presented in either text, audio, or audiovisual formats), a third for annotating Spatial-Relationships, and a fourth for describing Task-Design. Other research community areas needing development include: Human-Anatomy, Health-Disease, Human-Kinematics, Rodent-Kinetics, Sport, Virtual-reality, Music, Driving, etc.. HED-3G is designed to work efficiently with data stored per the BIDS (Brain Imaging Data Structure) or other archiving frameworks. As the amount of shared, well-annotated neuroimaging data accumulates, the foreseen rewards of identifying new information by applying machine learning and related methods to large, inhomogeneous neuroimaging data collections can be realized. Leveraging this technology to apply statistical machine learning methods to large collections of well-annotated neuroimaging data, cognitive neuroscience researchers should be able to identify yet unknown patterns of behavior-related brain dynamics and apply the results in healthcare and other application areas.

Keyword (Complete): HUMAN; NEUROIMAGING; EEG; MEG; BOLD; NEUROINFORMATICS

Grant/Other Support: : Swartz Foundation gift
Grant/Other Support: : NIH Grant 5R01NS047293-16

HED-3G tool ecosystem for machine-actionable event annotation

***D. TRUONG**[1,2], S. MAKEIG[1,2], A. DELORME[1,2], K. ROBBINS[3]; [1]Univ. of California San Diego, La Jolla, CA; [2]Inst. of Neural Computation, UCSD/INC/SCCN, La Jolla, CA; [3]University of Texas San Antonio

Abstract

Current standards for storing and sharing neuroimaging data such as BIDS (Brain Imaging Data Structure) focus on dataset-level organization and only mandate the inclusion of minimal metadata. However, BIDS allows a much richer specification of machine-actionable experimental event description using Hierarchical Event Descriptors (HED). The HED system has recently undergone a major developmental revision, to HED-3G (K. Robbins, this meeting). HED-3G introduces a more user-friendly way of annotation, greatly increasing the usability of the framework. It also enables the documentation of experiment structure, including tasks, condition variables, and temporal structure as related to experimental events. HED-3G introduces the concept of Library Schemas that allows the expansion of HED to cover all types of time-series data experiments without obfuscating (or bloating) the base schema (S. Makeig, this meeting).

Our working group has been developing a collection of tools to support HED-3G, a key component to making the standard usable and *widely adopted*. Validation tools, which ensure consistency and compliance of dataset annotation, have been updated to support new features of HED-3G and have been integrated into analysis software such as EEGLAB in MATLAB. The graphical user interface tool CTagger guides users through the annotation process, providing hints and shortcuts to expedite event annotation. The EEGLAB *HEDTools* plug-in incorporates CTagger to annotate and perform HED-based analysis on EEGLAB datasets and studies.

HED-3G not only provides a more holistic and detailed way of understanding shared data, but also supports the prospect of automatically aggregating and jointly analyzing data from collections of HED-annotated datasets across laboratories, studies, and data repositories. HED-enabled data discovery requires ongoing development of HED tools for data, tools and compute resources such as the NEMAR portal (*nemar.org*) to OpenNeuro (*openneuro.org*). An important planned component is integration of HED tools with the metadata handling mechanisms of Datalad, the data management software layer underlying current major data portals including OpenNeuro. Future developments also include HED-enabled analysis tools for HED-enabled multiple regression, event-related context analysis, and advanced statistical and machine learning methods suitable for application to large-scale neuroimaging data.

Keyword (Complete): HUMAN; NEUROIMAGING; EEG; MEG; BOLD; NEUROINFORMATICS
Grant/Other Support: : Swartz Foundation gift
Grant/Other Support: : NIH Grant 5R01NS047293-16

Meta- and mega-analysis of electroencephalographic data via the Neuroscience Gateway using the BIDS/HED framework in EEGLAB

*****A. DELORME[1,2], D. TRUONG[1,2], S. SIVAGNANAM[1,4], K. YOSHIMOTO[1,4], K. ROBBINS[3], S. MAKEIG[1,2], A. MAJUMDAR[1,4]; [1]Univ. of California San Diego, La Jolla, CA; [2]Inst. of Neural Computation, UCSD/INC/SCCN, La Jolla, CA; [3]University of Texas San Antonio; [4]San Diego Supercomputer Center

Abstract

EEGLAB is an interactive Matlab toolbox for processing continuous and event-related EEG, MEG, and other electrophysiological data incorporating independent component analysis (ICA), time/frequency analysis, artifact rejection, event-related statistics, and several useful modes of visualization of the averaged and single-trial data. EEGLAB integrates the Brain Imaging Data Structure (BIDS) standards and HED (version 3G) system for event annotation, allowing meta-analysis across a large corpus of datasets. The advantages of using the BIDS standards for analysis use and reuse are several. First, they do not rely on a complex database schema; they simply store data and relevant metadata in text and binary files within an ordinary file folder structure. Second, BIDS text files are both human and machine-readable, making it easier for new users to examine. BIDS data may be imported directly into EEGLAB for processing as an EEGLAB study (a group of related EEG datasets).

BIDS allows defining time-marked data events using dedicated text files that can contain detailed event information using the Hierarchical Event Descriptor (HED) system that is currently being upgraded to its third generation (HED-3G). Here we demonstrate an EEGLAB pipeline processing multiple BIDS-formatted, HED-3G-annotated datasets made publicly available on *OpenNeuro.org* and *NEMAR.org*. The pipeline imports BIDS datasets, preprocesses the data and extracts data epochs of interests based on HED tags in common across (here) BIDS-formatted datasets from different experiments and laboratories. We then process the data using the Neuroscience Gateway (*nsgportal.org*) platform using the EEGLAB *nsgportal* plug-in. The Neuroscience Gateway (NSG) project facilitates neuroscientist access to and use of National Science Foundation (NSF) funded high performance computing (HPC) resources. As a simple demonstration, we compute a 'target' event-related potential containing a P300 positivity, by averaging event-related data epochs collected from more than one experiment. We discuss the planned incorporation of BIDS and HED tools in the *NEMAR.org* data, tools, and compute resource.

Keyword (Complete): HUMAN; NEUROIMAGING; EEG; MEG; BOLD; NEUROINFORMATICS
Grant/Other Support: : Swartz Foundation gift
Grant/Other Support: : NIH Grant 5R01NS047293-16

Brain Imaging Data Structure (BIDS) compatible Hierarchical Event Descriptor (HED) library schema for clinical (i)EEG data annotation

T. PAL ATTIA[1], KAY ROBBINS[2], SCOTT MAKEIG[3], DUNG TRUONG[3], BEN H. BRINKMANN[4], GREG A. WORRELL[4], DORA HERMES[1.]; [1]Multimodal Neuroimaging Laboratory, Department of Physiology and Biomedical Engineering, Mayo Clinic, Rochester; [2]Department of Computer Science, University of Texas at San Antonio; [3]Swartz Center for Computational Neuroscience, Institute for Neural Computation, University of California San Diego; [4]Bioelectronics Neurophysiology and Engineering Laboratory, Department of Neurology, Mayo Clinic, Rochester

Abstract

Sharing data in standardized, reproducible formats enables powerful mega-analysis that lead to advances in neuroscience and informatics. In particular, in the clinical setting of epilepsy research which must deal with many different data formats acquired and analyzed using various software packages, the standardization of clinical terminology holds great potential. The Brain Imaging Data Structure (BIDS) has moved these standardization efforts forward for neuroimaging data. Within BIDS, there are clearly defined extensions for combining many different data types acquired in neuroimaging studies. The Intracranial Electroencephalography (iEEG) extension standardizes the data and metadata storage and the description of events during an iEEG recording. Additionally, BIDS has accepted the Hierarchical Event Descriptors (HED) system for detailed data and experiment event annotations. The new HED-3G schema library concept gives researchers the option to extend the base HED schema vocabulary, so as to adequately describe events of specific interest to their field, by adding library schemas to the base HED schema to support use of specialized vocabularies.

In this study, we developed a HED schema library for the Standardized Computer-based Organized Reporting of EEG (SCORE). SCORE is a textual description now implemented in commercial software for annotating EEG and ictal clinical events using standardized terms to create a standardized report. Our SCORE implementation in HED tackles these textual reports' lack of machine readability and makes SCORE available for and machine-readable by open-source software. To further make epilepsy terms available within BIDS, we use the HED convention to enrich BIDS datasets with epilepsy classification and seizure onset zone terms.

These efforts, including a HED schema library for the SCORE standard to annotate EEG record events and features including epilepsy classification and seizure onset zone annotation with participant and electrode information, support the machine-readability of clinical annotation. These tools can be used by many researchers worldwide alongside other standard machine-readable format data for data sharing using standard definitions, to reduce errors in clinical evaluations, and to perform advanced mega-analyses, ultimately advancing the understanding of the human brain.

The use of Hierarchical Event Descriptors with fMRI

M.J.M DENISSEN[1], M. PAWLIK[1], A.N. RAVENSCHLAG[1], N. HIMMELSTOß[1], S. HAWELKA[1], S. MAKEIG[2,3], K. ROBBINS[4], F. HUTZLER[1]; [1]Paris Lodron Universität Salzburg; [2]Univ. of California San Diego, La Jolla, CA; [3]Inst. of Neural Computation, UCSD/INC/SCCN, La Jolla, CA; [4]University of Texas San Antonio

Abstract

Data-sharing is becoming a more common practice in neuroscience. Currently, the majority of shared neuroimaging data consist of fMRI recordings. Many of these datasets include task-based fMRI, where participants engage with controlled sensory stimulation following task instructions. A major flaw is that these datasets are often minimally annotated. While these experiments can include complex events and event combinations, descriptions are often generalized on the level of conditions, at best allowing users to replicate the original analysis but leaving little opportunity to use more of the information about brain dynamics supporting cognition contained in the data to its full extent. The Hierarchical Event Descriptor (HED) system provides detailed and machine-actionable descriptions originally developed to describe neuroimaging experiments recording M/EEG data. HED has been applied successfully to EEG data to facilitate analysis across multiple datasets. Its current release (3G) provides a user-friendly environment for annotation of events using a sparse, human-readable tagging system (K. Robbins, this meeting), and several available tools facilitate annotation (D. Truong, this meeting). Applying HED to fMRI data is straightforward as fMRI and EEG experiments follow a similar structure. Despite the differences in the measured signals, both have similar requirements for analysis. By applying HED annotation to fMRI datasets researchers cannot only improve its reusability, but also benefit from the continuous development of HED-enabled tools that aid in developing and/or automatize parts of the data-analysis. Methods for fMRI analysis are developing rapidly; the use of multivariate rather than standard univariate analysis methods is increasingly common. Such experiments are particularly well suited for reanalysis, since they engage many aspects of cognition that can be linked to brain activity. Detailed information about these experiments, experiment events, and time structure is needed to use these valuable datasets to their full potential, and is also vital for helping researchers find relevant datasets in continuously growing databases. Applying the same HED standard to describe events and structure of fMRI and M/EEG experiments will also allow researchers to better connect results from these datasets on the basis of the underlying brain processes they record. Standardized, machine-actionable descriptors for data are vital to accelerating knowledge accumulation. HED-3G provides a framework for detailed description that is not only useful and necessary for M/EEG datasets, but equally so for fMRI data.

Keyword (Complete): HUMAN; NEUROIMAGING; EEG; MEG; BOLD; NEUROINFORMATICS

Grant/Other Support: : FWF W1233-B