Hierarchical Event Descriptor library schema for EEG data annotation

Tal Pal Attia[1], Kay Robbins[2], Sándor Beniczky[3,4], Jorge Bosch-Bayard[5], Arnaud Delorme[6], Brian Nils Lundstrom[7], Christine Rogers[8], Stefan Rampp[9,10], Pedro Valdes-Sosa[5], Dung Truong[6], Greg Worrell[7], Scott Makeig[6], Dora Hermes[1,*]

[1] Multimodal Neuroimaging Laboratory, Department of Physiology and Biomedical Engineering, Mayo Clinic, Rochester, Minnesota, USA
[2] Department of Computer Science, University of Texas San Antonio, San Antonio, Texas
[3] The Danish Epilepsy Centre, Filadelfia, Denmark
[4] Aarhus University and Aarhus University Hospital, Aarhus, Denmark
[5] Department of Biological and Health Psychology, Faculty of Psychology, Universidad Autónoma de Madrid, Spain
[6] Swartz Center for Computational Neuroscience, Institute for Neural Computation, University of California San Diego, La Jolla, California
[7] Bioelectronics, Neurophysiology, and Engineering Laboratory, Department of Neurology, Mayo Clinic, Rochester, Minnesota, USA
[8] McGill Centre for Integrative Neuroscience, Montreal Neurological Institute, Montreal, Canada
[9] Department of Neurosurgery, University Hospital Erlangen, Erlangen, Germany
[10] Department of Neurosurgery, University Hospital Halle (Saale), Halle (Saale), Germany
* Corresponding: Dora Hermes (hermes.dora@mayo.edu)

Abstract

Standardizing terminology to describe electrophysiological events can improve both clinical care and computational research. Sharing data enriched by such standardized terminology can support advances in neuroscientific data exploration, from single-subject to mega-analysis. Machine readability of electrophysiological event annotations is essential for performing such analyses efficiently across software tools and packages. Hierarchical Event Descriptors (HED) provide a framework for describing events in neuroscience experiments. HED library schemas extend the standard HED schema vocabulary to include specialized vocabularies, such as standardized clinical terms for electrophysiological events. The Standardized Computer-based Organized Reporting of EEG (SCORE) defines terms for annotating EEG events, including artifacts. This study makes SCORE machine-readable by incorporating it into a HED library schema. We demonstrate the use of the HED-SCORE library schema to annotate events in example EEG data stored in Brain Imaging Data Structure (BIDS) format. Clinicians and researchers worldwide can now use the HED-SCORE library schema to annotate and compute on electrophysiological data obtained from the human brain.

Introduction

Electroencephalography (EEG) recordings capture brain electrical activity in many different environments. In the clinical setting, a clinician reviews the recordings and, typically using free text, annotates events of interest to identify normal and abnormal findings. Currently, visual analysis and interpretation of the EEG data are critical for diagnosis, and agreement between experts interpreting the same data is higher when they use standardized terminology [1-3]. To facilitate consistency, the International League Against Epilepsy (ILAE) and the International Federation of Clinical Neurophysiology (IFCN) have provided definitions of terms used to describe data features in EEG recordings. Here, we integrate one set of defined terms with an open science standard such that these terms can be used in a system of structured fields in a machine-readable and actionable format.

Recent advances in the Open Science community have developed systems to share definitions of terms for event annotations in order to make these definitions available to a broad user base. The Hierarchical Event Descriptors (HED) system (https://www.hedtags.org/, https://doi.org/10.5281/zenodo.7930927) enables researchers to annotate their data using a formally specified framework for describing, annotating, and validating events occurring during the recording of time-series data [4]. In the HED framework, events are described using sets of terms drawn from several branching term hierarchies. Each event annotation consists of a comma-separated string of terms drawn from the HED vocabulary. This structure allows detailed and accurate annotation of events, machine validation of the annotations, and event description-based search across data collected in any number of studies.

The HED framework, vocabulary, and tools have evolved significantly over the past few years [5-7] in alignment with FAIR (Findable, Accessible, Interoperable, and Reusable) guiding principles [8]. HED enables researchers to share, search, and reuse annotated time series data across studies and projects, thereby promoting collaborations and facilitating meta-analyses. The Brain Imaging Data Structure (BIDS) (https://bids.neuroimaging.io) is a growing set of widely used specifications for structuring magnetic resonance imaging (MRI), EEG, magnetoencephalography

(MEG), and intracranial EEG (iEEG) datasets [9-11]. Incorporating HED annotation into BIDS data formatting facilitates large-scale data-sharing and analysis.

One of the key features of the HED framework is its extensibility. HED allows researchers to add new event types, attributes, and values as needed through schema libraries [4]. Schema libraries extend the standard HED schema with structured vocabulary, including terms unique to specific research fields.

To characterize normal as well as pathological (ictal) EEG data using standardized terms, Beniczky et al. [12] presented the Standardized Computer-based Organized Reporting of EEG (SCORE). SCORE aims to improve EEG reporting quality and facilitate data sharing and collaborative research by providing standardized terms and definitions that researchers and clinicians use. SCORE is based on broad international consensus in defining graphoelements, distinct waveforms identified in EEG recordings having unique characteristics. The first SCORE version was endorsed by ILAE-Europe, and the second SCORE version was endorsed as a reporting guideline by the IFCN [13]. SCORE aims to give experts standard terminology that can be used in clinical practice to annotate EEG data and maximize interobserver agreement. SCORE was built on multiple previously proposed guidelines and vocabularies [14-22] and on influential EEG textbooks [23,24]. The second SCORE version added terms drawn from numerous classifications, glossaries, and standard terminologies [17,25-30].

Since its publication, SCORE has been adopted and implemented in many clinical sites and EEG research projects. These studies have used SCORE to better document and study normal and abnormal EEG patterns in different neurological conditions [31-41]. Other studies have investigated the user experience and acceptability of SCORE, reporting that using SCORE has improved the consistency and accuracy of EEG reports [42-45].

To allow the broad scientific and clinical community to add EEG event annotations to BIDS-formatted EEG data in a standardized fashion, this study describes a HED library schema for SCORE (https://doi.org/10.5281/zenodo.7897597). Figure 1 shows a schematic view of how a few terms from the HED-SCORE library schema can be used to annotate graphoelements, including artifacts and seizure activity occurring in an

electrophysiology time series recording. Here, we demonstrate the use of HED-SCORE annotation applied to public BIDS data examples.
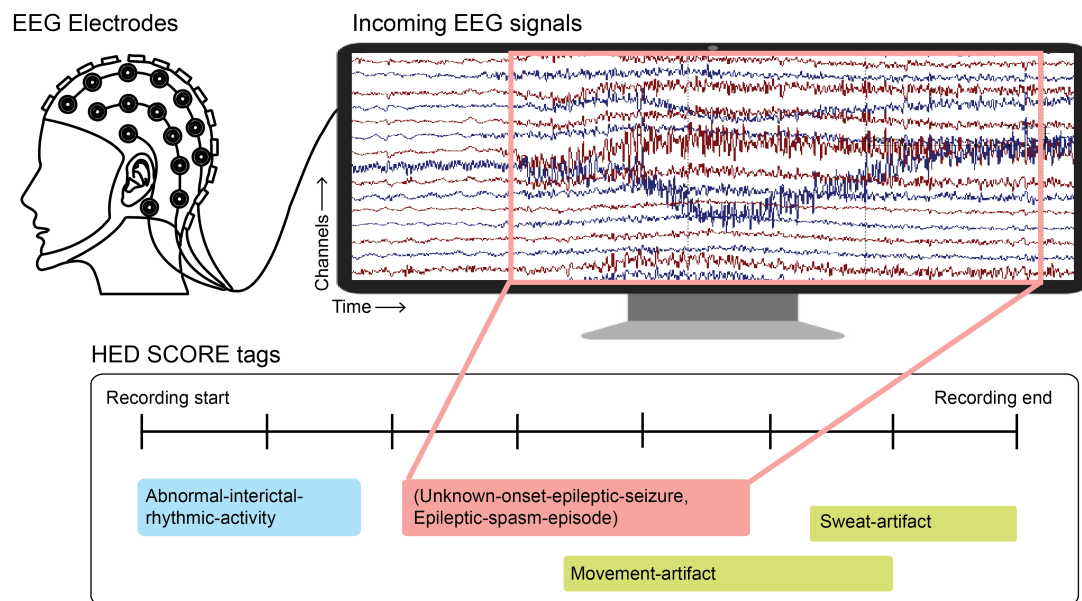

Figure 1 - Schematic of the SCORE implementation in a HED library schema. EEG signals are recorded from the scalp and visualized on a computer screen. A clinician interprets the signals, and HED-SCORE tags can be given for, e.g., interictal activity, seizures, and artifacts. Note that this is a schematic, and these features may not correspond to the data shown.

Results

We developed a HED library schema for SCORE. This library schema, which is the first extension of the standard HED schema, allows neurology researchers to annotate electrophysiology recordings per international standards. An interactive view of the HED-SCORE library schema is also available through an expandable HTML viewer (https://www.hedtags.org/display_hed.html). Figure 2 shows the hierarchical structure of the HED-SCORE library schema, as seen through the HTML viewer. The top level shows the main types of EEG graphoelements included in the SCORE report (Figure 2A). Each level can be expanded to show the lower-level nodes (Figure 2B). Hovering over each level displays the description of the SCORE term (Figure 2C).
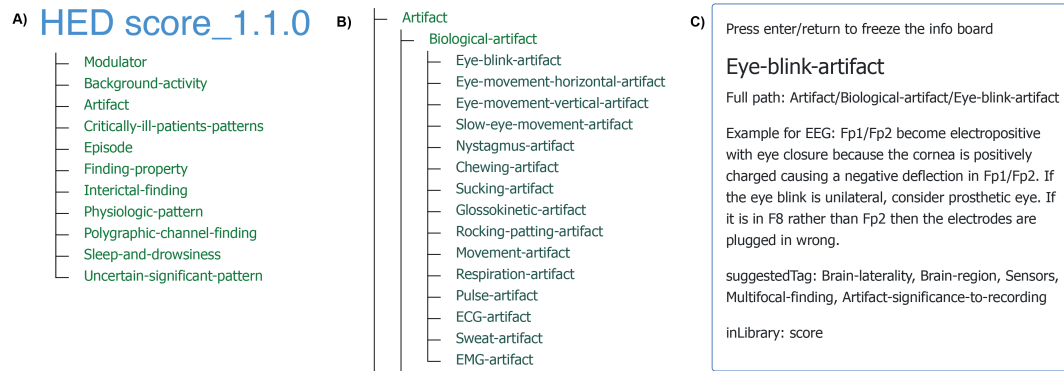
Figure 2 – HED library schema expandable HTML browser allows users to dive deeper into the schema terms and their descriptors. A) HED-SCORE library schema top-level describing the main types of EEG graphoelements included in SCORE. B) Top-level term 'Artifact' is expanded to show the lower-level nodes of 'Biological-artifact'. C) Extended description for 'Eye-blink-artifact' node shown when hovering over the term. The 'Attribute' section shows suggested tags to recommend additional terms that a user might want to include along with this tag. For data annotations, short forms are typically used (e.g., 'Eye-blink-artifact'), and each short form can easily be mapped to its long-form paths ('Artifact/Biological-artifact/Eye-blink-artifact').

HED-SCORE implementation in example BIDS datasets

We provide a set of BIDS examples to test the implementation and validation of HED-SCORE in BIDS (https://github.com/tpatpa/bids-examples/tree/master/xeeg_hed_score). These examples show the HED-SCORE implementation in three settings: annotating seizures, artifacts, and modulators. Examples are based on data from the Temple University Hospital EEG Corpus (TUEG) [46], an open-source collection of clinical EEG recordings performed at Temple University Hospital (TUH) and Mayo Clinic Rochester, MN. In BIDS-EEG formatting, each data file can be accompanied by a metadata file that describes events occurring at certain times during the corresponding EEG recording. When using HED annotations in these events files in BIDS, the HED library schema must first be defined at the project level in the dataset description (an example *dataset_description.json* file is shown in Figure 3). Within this dataset description, library schemas are referenced by their version number (e.g., score_1.1.0). HED-SCORE library schema version 1.1.0 is a partnered schema which is a library schema that is merged with a standard schema (8.2.0). Therefore, annotations can include tags from both the HED-SCORE library schema and its partner standard schema without using schema prefixes.

HED-SCORE library schema

```
"Name": "HED schema library for SCORE annotations example",
"BIDSVersion": "1.7.0",
"HEDVersion": "score_1.1.1"
```

Figure 3 - implementation example 'dataset_description.json' showing defined HED schemas used in the dataset.

After the schema is defined in the BIDS dataset description, there are two ways to add HED-SCORE tags to BIDS events files. Events files (*..._events.tsv*) are tab-separated files where each column has a name. One way to add HED-SCORE tags is to define a HED column in the events file that contains short-form HED-SCORE tags (e.g., *Aware-focal-onset-epileptic-seizure* rather than the full path in the schema). The short form is enabled by the HED rule defining that each term must appear in only one place within the schema. Thus, each term can be easily mapped to its long-form paths as needed. For example, the short-form HED-SCORE: *Aware-focal-onset-epileptic-seizure* can be mapped to its complete long-form term hierarchy: *Episode/Epileptic-seizure/Focal-onset-epileptic-seizure/Aware-focal-onset-epileptic-seizure*, allowing computer-based searches for events annotated with terms at any level of the hierarchy (e.g., searching for '*Epileptic-seizure'* returns all types of seizures appearing below this term in the hierarchy).

The second way to add HED-SCORE tags to a BIDS formatted recording is to accompany the BIDS events file (*..._events.tsv*) with a human and machine-readable field-value JSON sidecar file (*..._events.json*) that lists the HED tags associated with event types in a column of the file. The two ways to add HED-SCORE tags can be combined, as downstream analysis tools combine all annotations for each row in the BIDS events file. To show the ability to customize annotations according to needs while still being compliant with HED and BIDS, we provide examples of both ways to add HED-SCORE tags using the HED column in the events file (example with modulators in sub-ieegModulator) and the JSON events sidecar (example with seizures in sub-eegSeizureTUH and artifacts in sub-eegArtifactTUH).

BIDS example with seizures
This HED-SCORE BIDS example set includes one subject where seizures are tagged (sub-eegSeizureTUH). This example is based on one subject (41, male) from the TUH EEG Seizure Corpus [47]. This sub-dataset was annotated using twenty-seven labels [48]

described in Supplemental Table 1. The TUH labels were matched to their corresponding HED-SCORE tags for incorporation in the example dataset (Supplemental Table 1). Figure 4 shows the columns of the BIDS events file (..._*events.tsv*) with a temporal representation of the annotations on selected channels for the seizures example (also available for viewing on GitHub)
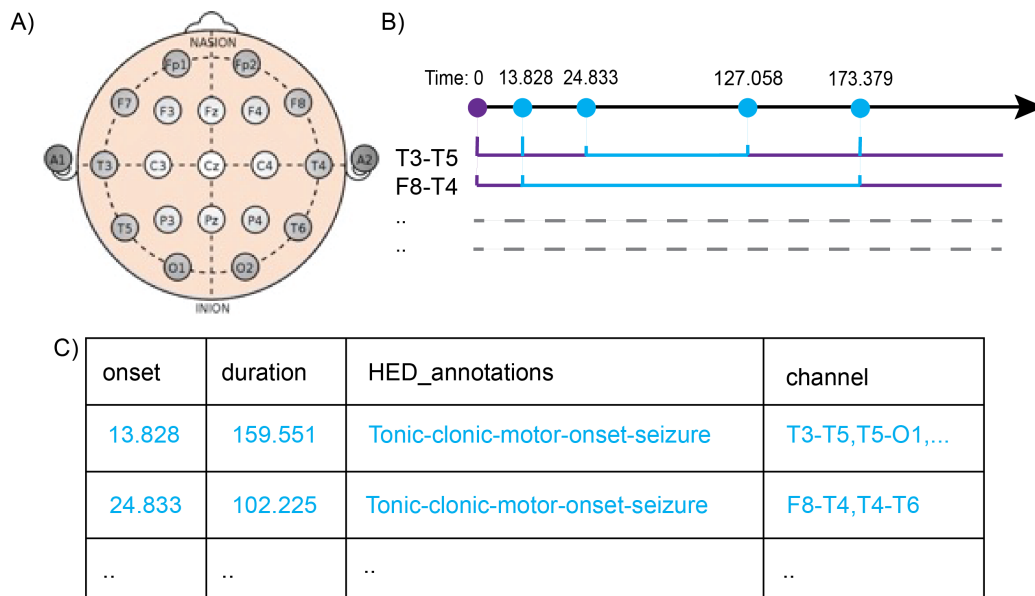

Figure 4 – Schematic example of a section of a BIDS events file (..._events.tsv) with HED-SCORE annotations. A) Electrode location and channel labels of the TUH EEG Seizure Corpus. B) Temporal representation of the example annotations on a subset of selected channels. C) Example annotations as they appear in the BIDS tab-separated value events file with columns for onset and duration in seconds and a HED column to annotate the graphoelements. The column 'channel' is added to indicate which channels in the BIDS _channels.tsv file the annotation corresponds to. The corresponding _channels.tsv file (not visible here, but can be seen in the online example) contains the montage in the dataset.

BIDS example with artifacts

The HED-SCORE BIDS example set includes another subject where artifacts are tagged (sub-eegArtifactTUH). This example is based on one subject (55, female) from the TUH EEG Artifact Corpus[49], a subset of TUEG that contains different artifacts. The artifacts were matched to HED-SCORE tags (Supplemental Table 1). The dataset includes annotations of 3 different types of artifacts: eye movement, electrode artifacts, and muscle artifacts. An example of how these artifact annotations are described in the JSON sidecar for the BIDS events file is shown in Figure 5 (also available for viewing on GitHub).

```
{
    "onset": {
        "Description": "REQUIRED. Onset (in seconds) of the event, measured from
         the beginning of the acquisition of the first data point stored in the
         corresponding task data file. Negative onsets are allowed, to account for events
         that occur prior to the first stored data point."
    },
    "duration": {
        "Description": "REQUIRED. Duration of the event (measured from onset) in
         seconds"
    },
    "annotation_type": {
        "LongName": "Hierarchical Event Descriptors annotations",
        "Description": "EEG interpretation Hierarchical Event Descriptors
         annotations",
        "Levels": {
            "eyem": "A spike-like waveform created during patient eye movement.
                     This artifact is usually found on all of the frontal polar electrodes with
                     occasional echoing on the frontal electrodes.",
            "elec": "An electrode artifact encompasses various electrode related
                     artifacts. Electrode pop is an artifact characterized by channels using the same
                     electrode 'spiking' with an electrographic phase reversal. ",
            "musc": "Muscle artifact. A very common, high frequency, sharp artifact
                     that corresponds with agitation/nervousness in a patient.",
            "eyem+musc": "Eye and Muscle"
        },
        "HED": {
            "eyem": "Eye-movement-horizontal-artifact",
            "elec": "Non-biological-artifact",
            "musc": "EMG-artifact",
            "eyem+musc": "(Eye-movement-horizontal-artifact , EMG-artifact)"
        }
    },
    "channel": {
        "LongName": "Annotated channel",
        "Description": "Comma-separated list of annotated channels corresponding to
         *_channels.tsv."
    }
```

Figure 5 - example of an accompanying events JSON sidecar that lists the HED tags used in annotating the artifacts example (sub-eegArtifactTUH). Required BIDS columns are onset and duration. When annotating HED-SCORE events, the column annotation_type is added to describe with levels and HED tags are used. The column 'channel' can be added to describe which channels in the BIDS _channels.tsv file the annotation corresponds to. This _channels.tsv file contains the montage in the dataset.

BIDS example with modulators

The HED-SCORE BIDS example set includes another subject where modulators are used (sub-ieegModulator). This example is based on one subject (46, male) monitored at Mayo Clinic (Rochester, MN) in the Epilepsy Monitoring Unit. The subject provided informed consent, and the research was performed in accordance with the Mayo Clinic Institutional Review Board, which also authorizes sharing of the data. It is important to note that the SCORE terminology is intended for scalp EEG, not intracranial EEG (iEEG). Many terms in SCORE don't translate to iEEG due to, e.g., differences in signal properties and standard scalp electrode locations with EEG (i.e., 10-20 system). However, some HED-SCORE tags may be used cautiously in iEEG settings, such as the modulators shown in this example (also available for viewing on GitHub). This example includes photic stimulus procedure annotations. Individual

events were annotated directly using a HED column in the BIDS events file
(…_*events.tsv*) (see Figure 6).

```
onset    duration  HED_annotations
0.656    88.625    Open-eyes
4.016    7.742     Intermittent-photic-stimulation/1 Hz
14.954   8.16      Intermittent-photic-stimulation/3 Hz
26.012   8.1       Intermittent-photic-stimulation/8 Hz
37.128   7.453     Intermittent-photic-stimulation/10 Hz
48.024   8.066     Intermittent-photic-stimulation/12 Hz
59.035   8.022     Intermittent-photic-stimulation/15 Hz
70.164   7.887     Intermittent-photic-stimulation/20 Hz
81.123   8.003     Intermittent-photic-stimulation/30 Hz
95.027   92.095    Close-eyes
102.031  7.905     Intermittent-photic-stimulation/1 Hz
112.887  8.042     Intermittent-photic-stimulation/3 Hz
123.866  7.992     Intermittent-photic-stimulation/8 Hz
134.945  8.015     Intermittent-photic-stimulation/10 Hz
145.902  8.019     Intermittent-photic-stimulation/12 Hz
156.821  7.951     Intermittent-photic-stimulation/15 Hz
167.902  8.079     Intermittent-photic-stimulation/20 Hz
178.954  7.995     Intermittent-photic-stimulation/30 Hz
```

Figure 6 - example of the BIDS events file (…_events.tsv) with the HED tags used in annotating the modulators example (sub-ieegModulator). This example shows required BIDS onset and duration columns in seconds and a HED column which describes when the subject opened and closed their eyes and when photic stimulation occurred. The HED column also shows how photic stimulation requires a numeric value to be added that indicates the frequency of the stimulation in units of Hz.

Discussion

We implemented SCORE in a HED library schema, such that the SCORE terms are described in structured fields in a human and machine-readable format. This implementation adheres to the HED design principles of uniqueness, clarity, structural sparsity, and orthogonality. The HED-SCORE library schema is compatible with annotating events in BIDS format, and we show several examples of HED-SCORE annotations integrated with the BIDS metadata standards. The HED-SCORE library schema is available in several formats on GitHub and can be viewed in an expandable HTML viewer.

Demonstrating BIDS compatibility is essential for several large efforts where EEG data are shared in BIDS, such as the EEGManyLabs project[50], and open repositories, including the Cuban Human Brain Mapping Project[51] and data sharing platforms such as LORIS[52]. The HED-SCORE library schema includes not only clinical EEG terms but also terminology that can be used for EEG signal preprocessing, such as artifact annotations. A dataset enriched with the HED-SCORE library schema can be validated using the BIDS (https://github.com/bids-standard/bids-validator) and HED validators (https://hedtools.org/hed/). This opens an excellent opportunity for researchers to develop BIDS apps[53] and efficiently run automated analysis pipelines with BIDS datasets annotated using the HED-SCORE library schema.

The development of the HED-SCORE library schema is the first step towards developing annotation tools in response to various clinical and scientific needs. For example, 'CTagger'[6] is a user-friendly user interface for easily annotating datasets with HED and, specifically, the HED-SCORE library schema. Integration of 'suggestedTag' schema attributes as suggestions within the 'CTagger' is currently under development. EEGLAB[54] tools already incorporate HED support, and several initiatives are piloting these integrations, including EEGNet.org (http://eegnet.org/) and the Global Brain Consortium (https://globalbrainconsortium.org/).

Annotating large datasets will be necessary to benefit from the open-source implementation of SCORE. Here, we showed examples of the HED- SCORE library schema applied to existing datasets. Labels used in two datasets from the TUEG archive (https://isip.piconepress.com/projects/tuh_eeg/) were matched to their corresponding HED-SCORE library schema tags. This mapping allows the conversion of the entire TUH data labels to HED-SCORE tags. Moreover, it demonstrates the ease of matching labels in existing annotated databases to their corresponding HED-SCORE library schema tags. The developed HED-SCORE library schema thus provides a base for future tagging efforts by other groups.

SCORE is designed and intended to report on EEG signals. However, to diagnose and study seizures in individuals with neurological conditions, various methods such as EEG, iEEG, or MEG can be used to capture and record electromagnetic brain activity.

The main distinction between these methods is the spatial resolution of the recordings. Where EEG and MEG give a global view, they have a relatively lower spatial resolution, while iEEG has a much higher spatial resolution at the cost of sparse coverage [55,56]. The signals, therefore, have very different features. While a few of the scalp EEG graphoelements and findings may apply to iEEG, many do not. Still, specific terms from the HED-SCORE library schema may be used cautiously in iEEG and MEG data. There are ongoing efforts[56-58] to review and characterize iEEG activity in a standardized manner, but these have not yet achieved broad consensus.

There are several limitations to this work. The HED-SCORE library schema only describes normal and abnormal EEG graphoelements. Therefore, this work does not detail the description of patient information, information related to referral and recording conditions, and administrative data, which are discussed in SCORE papers[12,13]. Moreover, we do not yet include neonatal SCORE templates. The American Clinical Neurophysiology Society standardized EEG terminology and categorization to describe continuous EEG monitoring in neonates[30] is part of SCORE. This has partially overlapping, and some additional terminology and terms change with age. Given the structure of HED, this can be included in future extensions of the HED-SCORE library schema. Lastly, tagging events in a large dataset will be necessary to further show the utility of the HED-SCORE library schema.

The HED-SCORE library schema implementation facilitates the sharing of well-annotated data for large-scale analysis. Furthermore, the HED-SCORE library schema can be integrated with existing tools to create new automated analyses, including approaches implemented in future BIDS apps.

Methods

The HED-SCORE library schema follows HED design principles

To extend the HED library schema, it is crucial to understand the cornerstones of third-generation HED, described in detail in the HED specification (https://hed-specification.readthedocs.io/). The HED-SCORE library schema adheres to the following HED design principles, basic rules, and form requirements.

HED design principles include:

- Preserve the orthogonality of concepts in specifying vocabularies.
- Abstract functionality into layers (e.g., more general vs. more specific).
- Separate content from presentation.
- Independent implementation from the interface (for flexibility).

Specifically, the HED library schema must conform to basic rules:

- *Uniqueness:* Every node name must be unique within a library schema.
- *Clarity:* Node names should be meaningful and readily understood by most users.
- *Structural sparsity:* Schema nodes should have at most seven child nodes if possible.
- *Orthogonality:* Terms used independently of one another should be in different sub-trees.

Moreover, there are requirements and style suggestions for the form of individual terms in the HED schema. An individual schema term should begin with a capital letter followed by lowercase letters, and spaces are not allowed. Accordingly, in the case of multiple words, hyphens should be used.

Hierarchical implementation of SCORE in HED

The HED standard schema is organized into subtrees representing various aspects of an event, such as its type, the agents and action involved, various items and properties, and relationships between these elements. In contrast, the HED-SCORE library schema has a top-level organization focused on the identification of and their morphology, which can be followed by location, features related to time, and the effect of modulators. The top levels of the HED-SCORE library schema correspond to the main types of events described in the SCORE papers [12,13]. These top levels represent modulators that pertain to external stimuli and interventions that change graphoelements, background EEG activity, sleep and drowsiness, interictal findings, clinical episodes and electrographic seizures, physiological patterns and patterns of uncertain significance, artifacts and findings in polygraphic EMG, EOG, and ECG

channels. The last top level is *'Finding-property'*, which is analogous to the HED standard schema *'-Property'*. This includes descriptive elements to describe "something that pertains to a thing, a characteristic of some entity or a quality or feature regarded as a characteristic or inherent part of someone or something." See the section on "Library schema design decisions" below for more detail.

Library schema design decisions

Several design decisions were made during the development of the HED-SCORE library schema to ensure adherence to the HED design principles.

First, the "Finding-property" top-level was added to the HED-SCORE library schema, as described above. This permits adherence to the HED's orthogonality rule by avoiding the repetition of general descriptors across the schema that can be used along with several different elements, for example, to indicate the laterality of a graphoelements or a seizure classification of an epileptic seizure.

Additionally, substantial effort was dedicated to the meaningful and unique naming of the HED-SCORE library schema individual terms so that each term could be understood independently. Further, HED requires that child nodes in the schema hierarchy must satisfy an "is-a" relationship with the parent level. This constraint significantly complicates the HED library schema design, but it allows users to use 'short-form' annotations and to search general terms. This improves usability and, subsequently, user acceptability. Because *Aware-focal-onset-epileptic-seizure* is-a *Epileptic-seizure*, a search for *Epileptic-seizure* we will return events annotated with *Aware-focal-onset-epileptic-seizure* as well as all other annotated seizure events. HED tools can convert all short-form tags (e.g., *Aware-focal-onset-epileptic-seizure*) to long-form full paths tags (e.g., */Episode/Epileptic-seizure/Focal-onset-epileptic-seizure/Aware-focal-onset-epileptic-seizure*).

Finally, since some of the terms used in SCORE terminology were already present in the standard HED schema (i.e., 'increasing' and 'decreasing'), these terms were not included in the HED-SCORE library schema. Hence, the HED-SCORE library schema is designed to not overlap with the standard HED schema.

HED-SCORE library schema development

The commit history reflects the development process of the HED-SCORE library schema. The initial pull request reflected the exact terminology used in the SCORE papers, and this library schema was reviewed with the HED Working Group team. During this review, it was pointed out that the elements should be refined for use in short form. This required making elements longer such that the terms were meaningful on their own and independent of their supervening schema term hierarchies. After the schema was refined, it was verified against the SCORE papers [12,13], and the library schema was pre-released in HED-SCORE version 0.0.1.

Then, we asked for input on the HED-SCORE prerelease version 0.0.1 from clinical collaborators and the BIDS community. The main concern was the clinical usability of the HED-SCORE library schema, which require linking graphoelements to relevant properties such as morphology, location, features related to time, and modulators. To maintain this interrelated structure of SCORE while adhering to the HED orthogonality principle (whereby each schema term can be used in only one place) a '*suggestedTag*' schema attribute was added in certain tags in the HED-SCORE library schema. Suggested tags indicate other HED tags that may be included when a given term is used. For example, *Epileptiform-interictal-activity* has several suggested tags regarding morphology, like *Spike-morphology* and *Sharp-wave-morphology*. Similarly, *Generalized-onset-epileptic-seizure* has several suggested tags regarding seizure classification such as *Tonic-motor-seizure* and *Myoclonic-motor-seizure*. This guides users to consider only those suggested tags that are relevant.

HED-SCORE library schema validation

HED tools can be applied to any library schema. According to the standard HED development process for (library) schema and using online HED tools, the HED-SCORE library schema was first developed in the MediaWiki format, a line-oriented markdown language that is easy to read and edit. It was validated to ensure its compliance with current HED (8.0.0+) requirements. As further HED tools for event validation, search and analysis use XML format, HED tools were then used to convert the MediaWiki schema library to XML format.

HED-SCORE example dataset validation

Several HED resources are available to validate BIDS datasets with HED tags (https://www.hed-resources.org). The example HED-SCORE dataset was validated using the HED tools python package (https://pypi.org/project/hedtools/, version: 0.2.0) and the provided Jupyter notebook (https://github.com/hed-standard/hed-examples/blob/main/hedcode/jupyter_notebooks/bids/bids_validate_dataset_with_libraries.ipynb) which validates HED annotations incorporated in a BIDS dataset. The tools gather all the HED information from the BIDS dataset, schemas used in the dataset, as found in the '*dataset_description.json*' file, the events (…*_events.tsv)* files, and accompanying JSON (…*_events.json)* files. Then, validation is performed on each row of each events (…*_events.tsv)* file.

Data Availability

The BIDS examples without electrophysiology data will be shared on the BIDS-examples page on GitHub (https://github.com/bids-standard/bids-examples) and is currently in a pull request (https://github.com/bids-standard/bids-examples/pull/324). The example dataset, including the electrophysiology data, will be shared on OpenNeuro.org with this publication.

Code Availability

The HED-SCORE library schema is available on GitHub (https://github.com/hed-standard/hed-schemas/tree/main/library_schemas/score); documentation is available on the HED webpage https://hed-schemas.readthedocs.io/en/latest/hed_score_schema.html.

Acknowledgments

Competing interests

Abbreviations

HED	Hierarchical Event Descriptors
SCORE	Standardized Computer-based Organized Reporting of EEG
BIDS	Brain Imaging Data Structure
EEG	Electroencephalography
ILAE	International League Against Epilepsy
IFCN	International Federation of Clinical Neurophysiology
FAIR	Findable, Accessible, Interoperable, and Reusable
MRI	Magnetic Resonance Imaging
MEG	Magnetoencephalography
iEEG	Intracranial Electroencephalography
HTML	HyperText Markup Language
JSON	JavaScript Object Notation
TSV	Tab-separated values
XML	Extensible Markup Language
TUEG	Temple University Hospital EEG Corpus
EMG	Electromyography
EOG	Electroocoulogram
ECG	Electrocardiogram

References

1   Benbadis, S. R. & Tatum, W. O. Overintepretation of EEGs and misdiagnosis of epilepsy. *Journal of clinical neurophysiology* **20**, 42-44 (2003).

2   Gerber, P. A. *et al.* Interobserver agreement in the interpretation of EEG patterns in critically ill adults. *Journal of Clinical Neurophysiology* **25**, 241-249 (2008).

3   Stroink, H. *et al.* Interobserver reliability of visual interpretation of electroencephalograms in children with newly diagnosed seizures. *Developmental medicine and child neurology* **48**, 374-377 (2006).

4   Robbins, K., Truong, D., Jones, A., Callanan, I. & Makeig, S. Building FAIR functionality: Annotating event-related imaging data using Hierarchical Event Descriptors (HED). (2020).

5   Bigdely-Shamlo, N. *et al.* in *2013 IEEE Global Conference on Signal and Information Processing.* 1-4 (IEEE).

6   Rognon, T. *et al.* in *2013 IEEE Global Conference on Signal and Information Processing.* 5-8 (IEEE).

7   Bigdely-Shamlo, N. *et al.* Hierarchical Event Descriptors (HED): Semi-Structured Tagging for Real-World Events in Large-Scale EEG. *Front Neuroinform* **10**, 42, doi:10.3389/fninf.2016.00042 (2016).

8   Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data* **3**, 1-9 (2016).

9   Gorgolewski, K. J. *et al.* The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific data* **3**, 1-9 (2016).

10   Niso, G. *et al.* MEG-BIDS, the brain imaging data structure extended to magnetoencephalography. *Scientific data* **5**, 1-5 (2018).

11   Pernet, C. R. *et al.* EEG-BIDS, an extension to the brain imaging data structure for electroencephalography. *Scientific data* **6**, 1-5 (2019).

12   Beniczky, S. *et al.* Standardized computer-based organized reporting of EEG: SCORE. *Epilepsia* **54**, 1112-1124 (2013).

13   Beniczky, S. *et al.* Standardized computer-based organized reporting of EEG: SCORE–second version. *Clinical Neurophysiology* **128**, 2334-2346 (2017).

14   Chatrian, G. A glossary of terms most commonly used by clinical electroencephalographers. *Electroenceph. Clin. Neurophysiol.* **37**, 538-548 (1974).

15   Angeles, D. Proposal for revised clinical and electroencephalographic classification of epileptic seizures. *Epilepsia* **22**, 489-501 (1981).

16   Noachter, S. A glossary of terms most commonly used by clinical electroencephalographers and proposal for the report form for the EEG findings. *Electroenceph clin Neurophysiol* **52**, 21-41 (1999).

17   Blume, W. T. *et al.* Glossary of descriptive terminology for ictal semiology: report of the ILAE task force on classification and terminology. *Epilepsia* **42**, 1212-1218 (2001).

18   Flink, R. *et al.* Guidelines for the use of EEG methodology in the diagnosis of epilepsy: International league against epilepsy: Commission report commission on European Affairs: Subcommission on European Guidelines. *Acta Neurologica Scandinavica* **106**, 1-7 (2002).

19   Berry, R. B. *et al.* The AASM manual for the scoring of sleep and associated events. *Rules, Terminology and Technical Specifications, Darien, Illinois, American Academy of Sleep Medicine* **176**, 2012 (2012).

20   Society, A. C. N. Guideline twelve: guidelines for long-term monitoring for epilepsy. *Journal of clinical neurophysiology: official publication of the American Electroencephalographic Society* **25**, 170-180 (2008).

21   Daube, J. R., Low, P. A., Litchy, W. J. & Sharbrough, F. W. Standard specification for transferring digital neurophysiological data between independent computer systems (ASTM E1467-92). *J Clin Neurophysiol* **10**, 397 (1993).

22   Guideline seven: a proposal for standard montages to be used in clinical EEG. American Electroencephalographic Society. *J Clin Neurophysiol* **11**, 30-36 (1994).

23   Ebersole, J. S. & Pedley, T. A. *Current practice of clinical electroencephalography*. (Lippincott Williams & Wilkins, 2003).

24   Niedermeyer, E. & Da Silva, F. L. *Electroencephalography--Basic principles, clinical applications, and related fields*. (Urban & Schwarzenberg, 2020).

25   Kane, N. *et al.* A revised glossary of terms most commonly used by clinical electroencephalographers and updated proposal for the report format of the EEG findings. Revision 2017. *Clinical neurophysiology practice* **2**, 170 (2017).

26   Fisher, R. S. *et al.* Operational classification of seizure types by the International League Against Epilepsy: Position Paper of the ILAE Commission for Classification and Terminology. *Epilepsia* **58**, 522-530 (2017).

27   Scheffer, I. E. *et al.* ILAE classification of the epilepsies: position paper of the ILAE Commission for Classification and Terminology. *Epilepsia* **58**, 512-521 (2017).

28   Trinka, E. *et al.* A definition and classification of status epilepticus–Report of the ILAE Task Force on Classification of Status Epilepticus. *Epilepsia* **56**, 1515-1523 (2015).

29   Hirsch, L. *et al.* American clinical neurophysiology society's standardized critical care EEG terminology: 2012 version. *Journal of clinical neurophysiology* **30**, 1-27 (2013).

30   Tsuchida, T. N. *et al.* American clinical neurophysiology society standardized EEG terminology and categorization for the description of continuous EEG monitoring in neonates: report of the American Clinical Neurophysiology Society critical care monitoring committee. *Journal of Clinical Neurophysiology* **30**, 161-173 (2013).

31   Wüstenhagen, S. *et al.* EEG normal variants: A prospective study using the SCORE system. *Clinical Neurophysiology Practice* **7**, 183-200 (2022).

32   Masnada, S. *et al.* EEG at onset and MRI predict long-term clinical outcome in Aicardi syndrome. *Clinical Neurophysiology* **142**, 112-124 (2022).

33   Benbadis, S. R., Beniczky, S., Bertram, E., MacIver, S. & Moshé, S. L. The role of EEG in patients with suspected epilepsy. *Epileptic Disorders* **22**, 143-155 (2020).

34   Aanestad, E., Gilhus, N. E. & Brogger, J. Interictal epileptiform discharges vary across age groups. *Clinical Neurophysiology* **131**, 25-33 (2020).

35   Meritam, P. *et al.* Diagnostic yield of standard-wake and sleep EEG recordings. *Clinical Neurophysiology* **129**, 713-716 (2018).

36   Beniczky, S., Rubboli, G., Covanis, A. & Sperling, M. R. Absence-to-bilateral-tonic-clonic seizure: A generalized seizure type. *Neurology* **95**, e2009-e2015 (2020).

37   Arntsen, V., Strandheim, J., Helland, I. B., Sand, T. & Brodtkorb, E. Epileptological aspects of juvenile neuronal ceroid lipofuscinosis (CLN3 disease) through the lifespan. *Epilepsy & Behavior* **94**, 59-64 (2019).

38   Larsen, P. M. *et al.* Photoparoxysmal response and its characteristics in a large EEG database using the SCORE system. *Clinical Neurophysiology* **132**, 365-371 (2021).

39   Kidokoro, H. *et al.* High-amplitude fast activity in EEG: An early diagnostic marker in children with beta-propeller protein-associated neurodegeneration (BPAN). *Clinical Neurophysiology* **131**, 2100-2104 (2020).

40   Reus, E. E., Visser, G. H. & Cox, F. M. Using sampled visual EEG review in combination with automated detection software at the EMU. *Seizure* **80**, 96-99 (2020).

41   Wüstenhagen, S., Terney, D., Gardella, E., Aurlien, H. & Beniczky, S. Duration of epileptic seizure types: A data-driven approach. *Epilepsia* (2022).

42   Brogger, J. *et al.* Visual EEG reviewing times with SCORE EEG. *Clinical neurophysiology practice* **3**, 59-64 (2018).

43   Japaridze, G., Kasradze, S., Aurlien, H. & Beniczky, S. Implementing the SCORE system improves the quality of clinical EEG reading. *Clinical neurophysiology practice* **7**, 260-263 (2022).

44   Guerrero-Aranda, A., Friman-Guillen, H. & González-Garrido, A. A. Acceptability by End-users of a Standardized Structured Format for Reporting EEG. *Clinical EEG and Neuroscience*, 15500594221091527 (2022).

45   Beniczky, S. *et al.* Interrater agreement of classification of photoparoxysmal electroencephalographic response. *Epilepsia* **61**, e124-e128 (2020).

46   Obeid, I. & Picone, J. The temple university hospital EEG data corpus. *Frontiers in neuroscience* **10**, 196 (2016).

47   Shah, V. *et al.* The temple university hospital seizure detection corpus. *Frontiers in neuroinformatics* **12**, 83 (2018).

48   Ochal, D. *et al.* The temple university hospital eeg corpus: Annotation guidelines. *Institute for Signal and Information Processing Report* **1** (2020).

49   Hamid, A. *et al.* in *2020 IEEE Signal Processing in Medicine and Biology Symposium (SPMB).* 1-4 (IEEE).

50   Pavlov, Y. G. *et al.* # EEGManyLabs: Investigating the replicability of influential EEG experiments. *cortex* **144**, 213-229 (2021).

51   Valdes-Sosa, P. A. *et al.* The Cuban Human Brain Mapping Project, a young and middle age population-based EEG, MRI, and cognition dataset. *Scientific Data* **8**, 45 (2021).

52   Das, S., Zijdenbos, A. P., Harlap, J., Vins, D. & Evans, A. C. LORIS: a web-based data management system for multi-center studies. *Frontiers in neuroinformatics* **5**, 37 (2012).

53   Gorgolewski, K. J. *et al.* BIDS apps: Improving ease of use, accessibility, and reproducibility of neuroimaging data analysis methods. *PLoS computational biology* **13**, e1005209 (2017).

54   Delorme, A. & Makeig, S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods* **134**, 9-21 (2004).

55   Grinvald, A. & Hildesheim, R. VSDI: a new era in functional imaging of cortical dynamics. *Nature Reviews Neuroscience* **5**, 874-885 (2004).

56   Mercier, M. R. *et al.* Advances in human intracranial electroencephalography research, guidelines and good practices. *NeuroImage*, 119438 (2022).

57   Frauscher, B. *et al.* Atlas of the normal intracranial electroencephalogram: neurophysiological awake activity in different cortical areas. *Brain* **141**, 1130-1144 (2018).

58   Flanary, J. *et al.* Reliability of Visual Review of Intracranial EEG in Identifying the Seizure Onset Zone: A Systematic Review and Implications for the Accuracy of Automated Methods. *Epilepsia* (2022).

Supplementary materials

Index	TEUG Label	Description (adapted from TUEG [46,47,48,49])	SCORE HED library schema (short-form)
1	spsw	Spike and/or slow wave. Patterns of EEGs observed during epileptic seizures. A short duration epileptiform even involving an electrographic spike in activity and/or a slow wave (low frequency wave). Usually no more than 1 sec in duration.	Spike-and-slow-wave-morphology
2	gped	Generalized periodic epileptiform discharge. Periodic diffuse spike/sharp wave discharges across multiple regions or hemispheres.	Generalized-periodic-discharges
3	pled	Periodic lateral epileptiform discharge. A regular, periodically occurring spike/sharp wave seen in a certain locality of the scalp.	Lateralized-periodic-discharges
4	eybl	Eyeblink. A specific, sharp, high amplitude eye movement artifact corresponding to blinking of the eye.	Eye-blink-artifact
5	artf	Artifact. Any non-brain activity electrical signal, such as those due to equipment or environmental factors.	Non-biological-artifact
6	bckg	Baseline/non-interesting events, all other non-seizure cerebral signals	Not applicable, baseline is the default
7	seiz	Seizure. A basic annotation for seizures.	Epileptic-seizure
8	fnsz	Focal nonspecific seizure. A large category of seizures occurring in a specific focality.	Focal-onset-epileptic-seizure
9	gnsz	Generalized seizure. A large category of seizures occurring in most if not all of the brain.	Generalized-onset-epileptic-seizure
10	spsz	Simple partial seizure. Brief seizures that start in one location of the brain (and may spread) where the patient is fully aware and able to interact.	Aware-focal-onset-epileptic-seizure
11	cpsz	Complex partial seizure. Same as simple partial seizure (spsz) but with impaired awareness.	Impaired-awareness-focal-onset-epileptic-seizure
12	absz	Absence seizure. Brief, sudden seizure involving lapse in attention. Usually lasts no more than 5 seconds and commonly seen in children.	Absence-seizure
13	tnsz	Tonic seizure. A seizure involving the stiffening of the muscles. Usually associated with and annotated as tonic-clonic seizures, but not always (rarely there is no clonic phase).	Tonic-motor-seizure
14	cnsz	Clonic seizure. A seizure involving sustained, rhythmic jerking.	Clonic-motor-seizure
15	tcsz	Tonic-clonic seizure. A seizure involving loss of consciousness and violent muscle contractions.	Tonic-clonic-motor-seizure
16	atsz	Atonic seizure. A seizure involving the loss of tone of muscles in the body.	Atonic-motor-seizure
17	mysz	Myoclonic seizure. A seizure associated with brief involuntary twitching or myoclonus.	Myoclonic-motor-seizure

18	nesz	Non-epileptic seizure. Any non-epileptic seizure observed. Contains no electrographic signs.	Seizure-PNES
19	intr	Interesting patterns. Any unusual or interesting patterns observed that don't fit into the above classes.	Uncertain-significant-pattern
20	slow	Slowing. A brief decrease in frequency	Translation in HED-SCORE not found
21	eyem	Eye movement. A very common frontal/prefrontal artifact seen when the eyes move.	Eye-movement-horizontal-artifact
22	chew	Chewing. A specific artifact involving multiple channels that corresponds with patient chewing, "bursty".	Chewing-artifact
23	shiv	Shivers. A specific, sustained sharp artifact that corresponds with patient shivering.	Movement-artifact
24	musc	Muscle artifact. A very common, high frequency, sharp artifact that corresponds with agitation/nervousness in a patient.	EMG-artifact
25	elpp	Electrode pop. A short artifact characterized by channels using the same electrode "spiking" with perfect symmetry.	Electrode-pops-artifact
26	elst	Electrostatic artifact. Artifact caused by movement or interference on the electrodes, variety of morphologies.	Induction-artifact
27	calb	Artifact caused by calibration of the electrodes. Appears as a flattening of the signal in the beginning of files.	Other-artifact
28	hphs	Hypnagogic hypersynchrony, a brief period of high amplitude slow waves	Hypnagogic-hypersynchrony
29	trip	Large, three-phase waves frequently caused by an underlying metabolic condition.	Translation in HED-SCORE not found
30	elec	This tag was developed for the TUAR artifact corpus an indicates electrode artifacts that encompass three electrode related artifacts. 1) Electrode pop is an artifact characterized by channels using the same electrode "spiking" with an electrographic phase reversal. 2) Electrostatic is an artifact caused by movement or interference of electrodes and or the presence of dissimilar metals. 3) A lead artifact is caused by the movement of electrodes from the patient's head and or poor connection of electrodes. This results in disorganized and high amplitude slow waves.	Non-biological-artifact

Table 1 – Labels and descriptions used to annotate the Temple University Hospital EEG Corpus (TUEG)[46,47,48,49] and their corresponding SCORE HED library schema annotations. In the TUEG Corpus all channels with baseline activity were tagged as 'bckg'. In the BIDS examples, however, background activity is not tagged, as we assume that the default is that channels show baseline activity and the TUEG 'bckg' label is not translated to an HED-SCORE tag. Direct translations for TUEG labels 'slow' and 'trip' were not found in HED-SCORE.