HED schemas

Release 0.0.1

HED Working Group

Jul 13, 2023

CONTENTS:

_static/images/croppedWideLogo.png

Downloads

- HED schemas (PDF
- `**HED resources (<https://www.hed-resources.org/>`_**

Note: this is a work in progress.

INTRODUCTION TO HED SCHEMAS

HED (Hierarchical Event Descriptors) is an evolving framework for the description and formal annotation of events and other types of data. The HED ecosystem includes a structured vocabulary (specified by a HED schema) together with tools for validation and for using HED annotations in data search, extraction, and analysis.

A **HED schema** is a hierarchically-structured specification of a vocabulary. The HED ecosystem includes a **standard schema** containing the basic vocabulary needed for annotation of experimental data as well as specialized **library schemas** for the additional field-specific terms needed to complete an annotation.

1.1 Scope of HED

HED allows researchers to annotate what happened during an experiment, including experimental stimuli and other sensory events, participant responses and actions, experimental design, the role of events in the task, and the temporal structure of the experiment. The resulting annotation is machine-actionable, meaning that it can be used as input to algorithms without manual intervention. HED facilitates detailed comparisons of data across studies.

As the name HED implies, much of the HED framework focuses on associating metadata with the experimental timeline to make datasets analysis-ready and machine-actionable. However, HED annotations and framework can be used to incorporate other types of metadata into analysis by providing a common API (Application Programming Interface) for building inter-operable tools.

1.2 Role of library schemas

To avoid uncontrolled expansion of the base HED vocabulary with specialized terminology, HED supports the creation of library schemas, which are specialized vocabularies that can be used in conjunction with the base schema to analyze specific aspects of interest.

To use a programming analogy, when programmers write a Python module, the resulting code does not become part of the Python language or core library. Instead, the module becomes part of a library used in conjunction with core modules of the programming language. HED annotations may contain any combination of tags from the standard vocabulary and/or HED library vocabularies.

Several library schemas are currently under development including the SCORE library for describing data features of clinical interest (e.g., seizure, sleep stage IV) as well as schemas for describing features in language structure and video.

Each library schema has its own directory under in the **hed-schemas** GitHub repository.

1.3 The HED community and resources

All HED-related source and documentation repositories are housed on the HED-standard organization GitHub site, https://github.com/hed-standard, which is maintained by the HED Working Group. HED development is open-source and community-based. The official HED website https://www.hedtags.org.

The HED Working Group invites those interested in HED to contribute to the HED ecosystem and development process.

HED schemas are community-driven. Users can contribute to existing schema or propose the development of new schema by posting an **issue** to the **hed-schemas** GitHub repository.

1.4 HED schemas in BIDS

BIDS, which stands for Brain Imaging Data Structure, is a widely-used standard that specifies how neuroimaging data should be organized. HED is well-integrated into the BIDS standard.

The most common use case (for 99.9% of the HED users) is to use the standard HED schema available on GitHub in the `standard_schema` directory of the (https://github.com/hed-standard/hed-schemas/tree/main/standard_schema/hedxml).

Starting with BIDS version 1.8.0, BIDS allows the value associated with the `"HEDVersion"` key in the `dataset_description.json` file to be a list rather than a string expressing the HED version.

The following example specifies that the annotations in this dataset use HED standard schema version 8.1.0, along with library schema `testlib` version 1.0.2. Tags from the `testlib` schema library are to be prefixed with `la:`.

Example: Proposed specification of library schema in BIDS.

```
{
    "Name": "A wonderful experiment",
    "BIDSVersion": "1.8.0",
    "HEDVersion": ["8.1.0", "la:testlib_1.0.2"]
}
```

The `"la"` library schema is the `./library_schemas/testlib/hedxml/HED_testlib_1.0.2.xml` file found in the **hed-schemas** GitHub repository. The specification indicates that annotations using HED tags from this library have the `la:` prefix (e.g., `la:XXX`).

TWO

HED LISA SCHEMA

The cognitive neuroscience of language is a large subdomain that investigates the neural basis of language processing. Language is central to human cognition and the majority of neuroimaging experiments make use of language in their experimental designs. Analysis of language experiments can be based on the low-level orthographic or phonetic characteristics of stimuli, or higher level syntactic and grammatical properties.

LISA is a Hierarchical Event Descriptors Library Schema for **Li**nguistic **S**timuli **A**nnotation. The schema allows for detailed annotation of neuroimaging experiments that involve language events, from carefully controlled experiments in the domain of language processing, to more complex naturalistic paradigms involving written or spoken language.

LISA allows for annotation of language stimuli on different levels through the orthogonal definition of Language-items and Language-item-properties. Full sentences can be annotated with sentence-level characteristics, while the individual words in the sentence can be associated with word-level characteristics, and so on. Annotation possibilities are extensive and cover characteristics found across languages to allow for between language comparisons.
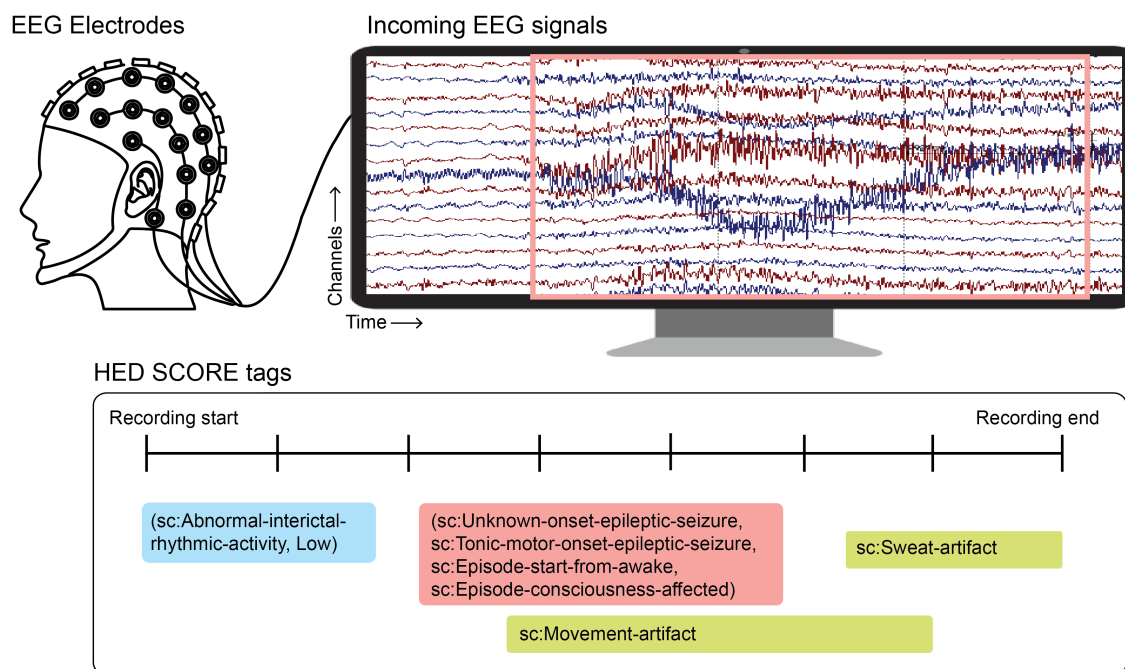
2.1 Development

LISA is currently under development and only available in prerelease. The current version of the schema is primarily centered around written language and further development focusses on adding grammatical aspect characteristics and spoken word characteristics into the vocabulary. If you are interested in participating in this effort or have any comments or suggestions, please post an **issue** to this repository.

HED SCORE SCHEMA

Sharing data in standardized, reproducible formats enables powerful mega-analyses that advance neuroscience. In the clinical setting of epilepsy research, the standardization of clinical terminology of electrophysiological events holds great potential for large-scale computation. Machine readability of electrophysiological event annotations is key to allow analyses across various tools and packages. The Standardized Computer-based Organized Reporting of EEG (SCORE)[1],[2] standard is a textual description for annotating EEG and ictal clinical events using standardized terms.

In this study, we make the SCORE standard machine-readable using the Hierarchical Event Descriptor (HED) library schema. HED library schemas allow researchers to extend the standard HED schema vocabulary by supporting specialized vocabularies. Our SCORE standard implementation in HED tackles the SCORE textual reports' lack of machine readability and makes the SCORE standard available and machine-readable by open-source software.

We show several examples of annotations using the HED-SCORE library schema in the Brain Imaging Data Structure (BIDS). The HED-SCORE library schema can be used by many researchers worldwide to annotate electrophysiology measurements from the human brain.

3.1 Development

The HED-SCORE library schema maintains the hierarchy as presented in SCORE papers [^1,2]. With the GitHub commit history reflecting the development process of the HED-SCORE library schema.

In the HED standard schema, top levels identify events of interest. Annotating events includes identifying graphoelements 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.

The SCORE HED schema library is intended to describe all normal and abnormal EEG features. Therefore, description of patient information, referral and recording condition information, administrative data, and continuous EEG monitoring in neonates is beyond this scope.

3.2 Validation

The HED schema library for SCORE was converted and validated using the HED tools. See more here.

3.3 Brain imaging data structure (BIDS)

HED schema library for SCORE is compatible with the BIDS human and machine-readable events annotations .tsv files, see more here. An implementation example using HED schema library for SCORE annotations is available here.

3.4 References

[1]: Beniczky, Sándor, et al. "Standardized computer-based organized reporting of EEG: SCORE." Epilepsia 54.6 (2013): 1112-1124.

[2]: Beniczky, Sándor, et al. "Standardized computer-based organized reporting of EEG: SCORE–second version." Clinical Neurophysiology 128.11 (2017): 2334-2346.

[3]: *Manuscript: Tal Pal Attia et al., (in prep). "Hierarchical Event Descriptor library schema for clinical EEG data annotation".*

FOUR

INDICES AND TABLES

- genindex
- modindex
- search