

# **MTH5:** A Hierarchical Format for Magnetotelluric Data

J. R. Peacock<sup>1</sup>

<sup>1</sup>U.S. Geological Survey

December 12, 2018

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>General Structure</b>	<b>3</b>
2.1	Description of <b>site</b> . . . . .	3
2.1.1	Example JSON encoded metadata for <b>site</b> . . . . .	4
2.2	Description of <b>field_notes</b> . . . . .	4
2.2.1	Example JSON encoded metadata for <b>field_notes</b> . . . . .	5

# 1 Introduction

The magnetotelluric community is relatively small which has led to various formats for storing the time series data. Some type of ASCII format seems to be the most prevalent because before large data sets that was the easiest method of storage. However, in terms of read/write efficiency, ASCII is the slowest. Various binary formats exist, some proprietary and some open like the Scripps format, though efficient these files lack some critical metadata.

The most widely used format for archiving large data sets is the Hierarchical Data Format (HDF5). The HDF5 Group (<https://www.hdfgroup.org/>) maintains and updates the format as well as the software needed to read and write. The advantages of HDF5 are the metadata can be stored alongside the data, different components or calibration data or different schedules can be stored as separate folders, and the data is stored to the hard drive making reading and writing very efficient. There is also capability to access a single file from multiple different processors making it versatile for parallel computing. The goal for MTH5 is to develop a format where metadata can easily be stored and searched, as well as have a hierarchical structure where a single station can be stored in one HDF5 file. *Note: this is still in the development stage and any comments are welcome.* [jpeacock@usgs.gov](mailto:jpeacock@usgs.gov).

## 2 General Structure

The top level of a MTH5 file, the root directory, stores attributes important to the location of the data, how the data were collected, the provenance of the data, the software used to write the data, and copyright information on how the data can be used. These metadata are stored as JSON encoded strings. The metadata are split into the following headings, where each indentation represents a different layer.

### 2.1 Description of site

- **site:** Information about where the site is
  - ***acquired\_by***: who acquired the data
    - \* ***email***: email of person responsible for the data
    - \* ***name***: name of person responsible for the data
    - \* ***organization***: organization name for person responsible for the data
    - \* ***organization\_url***: organization website for person responsible for the data
  - ***coordinate\_system***: [ Geographic North | Geomagnetic North | something else ]
  - ***datum***: Datum that represents the location coordinates, (WGS84)
  - ***declination***: geomagnetic declination of station location
  - ***declination\_epoch***: epoch from which declination is estimated
  - ***elev\_units***: units of elevation [ meters | feet | something else ]
  - ***elevation***: elevation of station in ***elev\_units***
  - ***end\_date***: date and time of when recording stopped<sup>1</sup>.
  - ***id***: name of the station
  - ***latitude***: latitude of station<sup>2</sup>
  - ***longitude***: longitude of station<sup>2</sup>
  - ***start\_date***: date and time of when recording began<sup>1</sup>
  - ***survey***: survey name and location

---

<sup>1</sup>The preferred format is YYYY-MM-DDThh:mm:ss.ms UTC

<sup>2</sup>Preferred format is decimal degrees

### 2.1.1 Example JSON encoded metadata for site

```
{
  "acquired_by": {
    "email": "generic@email.com",
    "name": "John Doe",
    "organization": "Company Name",
    "organization_url": "www.company_name.com"
  },
  "coordinate_system": "Geomagnetic North",
  "datum": "WGS84",
  "declination": 15.5,
  "declination_epoch": 1995,
  "elev_units": "meters",
  "elevation": 1111.72,
  "end_date": "2015-08-17T14:19:38.000000 UTC",
  "id": "mshH020",
  "latitude": 46.655559999999994,
  "longitude": -121.48472,
  "start_date": "2015-08-14T13:59:55.000000 UTC",
  "survey": "MT survey Washington, USA"
}
```

## 2.2 Description of field\_notes

### 2.2.1 Example JSON encoded metadata for field\_notes

```
{
  "data_logger": {
    "id": "ZEN18",
    "manufacturer": "Zonge",
    "type": "32-Bit 5-channel GPS synced"
  },
  "data_quality": {
    "author": "C. Cagniard",
    "comments": "testing",
    "rating": 0,
    "warnings_comments": "bad data at 2018-06-07T20:10:00.00",
    "warnings_flag": 1
  },
  "electrode_ex": {
    "azimuth": 15.5,
    "chn_num": 4,
    "contact_resistance": 1,
    "gain": 1,
    "id": 1,
    "length": 100.0,
    "manufacturer": "Borin",
    "type": "Fat Cat Ag-AgCl",
    "units": "mV"
  },
  "electrode_ey": {
    "azimuth": 105.5,
    "chn_num": 5,
    "contact_resistance": 1,
    "gain": 1,
    "id": 2,
    "length": 92.0,
    "manufacturer": "Borin",
    "type": "Fat Cat Ag-AgCl",
    "units": "mV"
  },
  "magnetometer_hx": {
    "azimuth": 15.5,
    "chn_num": 1,
    "gain": 1,
    "id": 2374,
    "manufacturer": "Geotell",
    "type": "Ant 4 Induction Coil",
    "units": "mV"
  },
  "magnetometer_hy": {
    "azimuth": 105.5,
    "chn_num": 2,
    "gain": 1,
    "id": 2384,
    "manufacturer": "Geotell",
    "type": "Ant 4 Induction Coil",
    "units": "mV"
  },
  "magnetometer_hz": {
    "azimuth": 90,
    "chn_num": 3,
    "gain": 1,
    "id": 2514,
    "manufacturer": "Geotell",
    "type": "Ant 4 Induction Coil",
    "units": "mV"
  }
}
```