Guideline for documenting and archiving measurement projects in structural engineering research

Gunnstein T. Frøseth, Ole Øiseth,
Knut A. Kvåle, and Øyvind W. Petersen

Department of Structural Engineering, NTNU

December 13, 2021

Abstract

This document gives a brief guideline on how to properly document and archive measurements such that the data acquired in the project can be used by yourself and others in the future. Essential metadata for the project, trials and sensors are presented. The HDF5 file format is suggested for implementation of the guideline.

1 Introduction

Researchers in structural engineering have to conduct a measurement project at some point in their career. The objective of a measurement project is to obtain data that supports a hypothesis or explores some idea. Typically, after aquiring, analyzing and reporting on the hypothesis, the data and hypothensis is then abandoned for the next hypothesis and measurement project. Many times, data obtained in one measurement project is useful in another project, e.g. if peak acceleration in a pedestrian bridge is obtained in project A then project B which seeks to determine the distribution of peak acceleration in all pedestrian bridges will find the data from project A very useful. It is also common that other researchers can exploit the data acquired in a project to support their work.

Data produced in measurement projects are typically acquired, analyzed and reported within a relatively short timeframe in which the researcher has control (fresh memory) of the measurement set up and environment that the project was performed in. The project is typically documented on various hand written notes and short files on the current working directory. When the researcher continues on to the next project, the data is stored on some directory on some harddrive, while the essential metadata is left lying around on handwritten notes and/or on a different directory on another harddrive.

A prerequisite for reusing data for other projects or for data to be useful to other researchers is that the data is stored and the measurement project is properly documented, i.e. both the data and metadata is archived. The distinction between data and metadata is made clear below:

- *Data* is the actual reading from a sensor, e.g. the acceleration from an accelerometer, the strain from a strain gage or the temperature from a thermometer. This is the first component of the data of a measurement project that most people think of.

- *Metadata* is relevant data about the data. Type, sensitivitY, position and orientation of an accelerometer that is used to obtain accelerations are all examples of metadata. The leader and participants of the project, the location where the measurements were conducted and the objective of the measurement project are also examples of metadata. Metadata is the component of the measurement project that is most often missing when you retrieve data from a fellow researcher or from a previous measurement project.

Generally, it is easy to define the data in a measurement project because it is closely related to a finite number of easily identifiable sensors. On the other hand, it is much more difficult to define the metadata more precisely since all data about data is technically metadata. Another complicating factor is that metadata which is necessary and sufficient to one project is not sufficient to another project.

Ideally, this document should have provided a overview of all relevant metadata to any measurement project. However, it is impossible and impractical to gather and store metadata about a project which is complete and sufficient for all future measurement projects. This document therefore only provides a recommendation of the most important and common pieces of metadata, i.e. essential metadata, found in measurement projects within structural engineering.

Furthermore, this document also provides recommendations on file format and storage format since these aspects are essential to be able to access and share measurement projects with other researchers today and in the future.

2 Metadata - data about data

The metadata of a measurement project can be categorized into metadata about the *project*, metadata about each *trial* and metadata about each *sensor*.

2.1 Project metadata

The essential metadata about the measurement project aims to put the data into context such that prospective users can understand and evaluate whether the data is useful for their project. Contact information to the leader and

other essential personel should be included to ensure that the user knows where more information about the project can be requested and proper attribution and credit is given. A project coordinate system should also be established to enable precise description of the placement and orientation of sensors and other spatial information in the project.

2.2 Trial metadata

Several trials[1] are necessary or desireable within a measurement project to explore the operational range of the system, determine the influence of different excitation, test the structure under different operational conditions or change the sensor setup to improve the observability of the system.

One trial may differ significantly from another and these differences may not be measured or "observed" in the data. If possible, it is essential to properly describe and document the distinguising features of the trails for posteriority. Another important aspect is to be able to distinguish between different trials such that unique names should be given to each trial.

2.3 Sensor metadata

The essential metadata for the sensors is information that enables conversion from electrical units to physical units, e.g. sensitivity for an accelerometer or gage factor for a strain gage, and information about the spatial location and orientation of the sensor to ensure that the measured response can be associated with the correct structural member and structural behaviour, see table 1. Please note that information about the temporal information must also be stored, this is can either be done by adding a stand alone dataset with sampling times to the group (a time vector) or by the start time and sampling interval for each data point in the data.

---

[1]A trial may also be known as a run, try, experiment, test or an attempt.

Table 1: Essential metadata for documenting a measurement project.

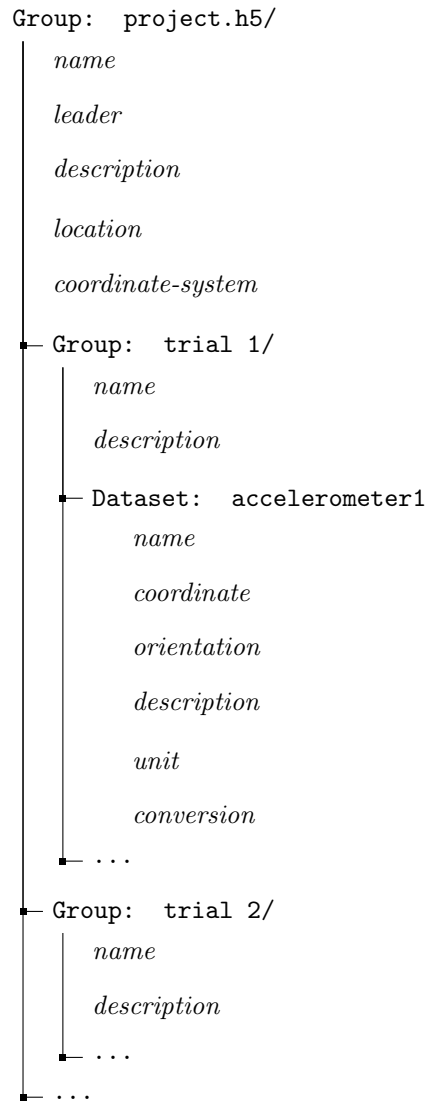| Metadata	Key	Description	Example
Project	name	Name of the project	MeTinT - Measurement from trains in traffic.
	contact	Name of the main contact person for the project. Include contact info, e.g. email-address.	Gunnstein T. Frøseth gunnstein.t.froseth@ntnu.no, Department of Structural Engineering, NTNU, Trondheim, Norway.
	description	Description of the project. Include information about the over purpose, time frame and other relevant information	MeTinT is a measurement project between 2021 and 2026 where a Type 73 multiple unit, in daily service at Dovrebanen between Trondheim and Oslo, is instrumented for monitoring of the train itself and the railway infrastructure. The overall purpose is to provide data for damage detection and structural health monitorin of railway infrastructure.
	location	The location where the measurement system is installed.	Type 73 multiple unit operating between Trondheim and Oslo.
	coordinate-system	Coordinate system for the project, which is used to define the position and orientation of all other data in the project. Define the origin, type of coordinate system and the (positive) direction of the axis.	Cartesian coordinate system with origin in the center of mass of the wagon with the pantograph, x-axis defined positive towards the closest leading wagon, z-axis positive upwards and y-axis defined positive by right hand rule.
Trial	name	Name of trial.	lundamo-bridge-1
	description	Description of the trial. Focus in particular on the distinguising features of the trial. If the trial is a product of an event, then describe what triggered the acquisition.	Passage of Lundamo railway bridge northbound on 2021-02-24 12:01:09.
Sensor	name	Name of the sensor for identification and visualization.	A10x
	coordinate	Coordinate of the sensor in the project coordinate system.	(13.3, 0., 3.4)
	orientation	Orientation of the *positive direction* of the sensor in project coordinate system.	(0, 0, 1)
	description	General description of sensor installation and make and serial number. Include information about how it was installed (magnet, glue, bolt) and any specification on location.	X-axis of Dytran3538A accelerometer S/N: 381. Installed on roof of train on pantograph mounting bracket with a magnet.
	unit	Unit of the data points that are stored.	Volt (V)
	conversion	How to convert the data reading, which is typically in an electrical unit (e.g. volt) into a "spatial" unit (e.g. g or m/s2).	1002 mV/g

3   Practical implementation

Metadata and data must be stored together to ensure that both components are properly preserved/archived for future use. There are endless ways to achieve this, but in this section a recommended approach is presented to facilitate consistency in the way that project data and metadata are archived, allow efficient use of the data and ensure a robust approach to archiving.

Ideally, an approach where metadata and data can be stored and coordinated together should be used. *Hierarchical Data Format Version 5* (HDF5) is a open source binary data format designed for persistant storage of big, complex and heterogeneous data. The file format is available on all operating systems and a large number of programming languages, including Fortran 90, Java, C, C++, Python and Matlab.

Most importantly, the format allows for elegant handling and coordination of data and metadata. The data model for the file format is simple and revolves around two objects `Groups` and `Datasets`. The datasets can be arbitrary python objects, e.g. arrays, images, graphs or even documents, such as pdf or Excel. Groups are containers that can hold a variety of different datasets and other groups. Both groups and datasets have `attributes`, which can be used to store metadata about the object and thereby directly link up the data and the metadata.

An illustration of documentation of measurement program with the HDF5 file format is given below where the file `project.h5` is a HDF5 file with the data and metadata for some project.

```
Group:  project.h5/
│   name
│
│   leader
│
│   description
│
│   location
│
│   coordinate-system
│
├─ Group:  trial 1/
│  │   name
│  │
│  │   description
│  │
│  ├─ Dataset:  accelerometer1
│  │      name
│  │
│  │      coordinate
│  │
│  │      orientation
│  │
│  │      description
│  │
│  │      unit
│  │
│  │      conversion
│  ├─ ...
│
├─ Group:  trial 2/
│  │   name
│  │
│  │   description
│  ├─ ...
├─ ...
```

A HDF5 file is also a group object and is refered to as the root with address /. In most cases, a file is only associated with a single measurement project and it is therefore appropriate to add the project metadata (*name*, *leader*, *description*, *coordinate-system*) to the root object as attributes as shown in the figure. In cases where multiple measurement projects are included in a single HDF5 file, it is fully possible to add a project group with (metadata stored in attributes) as a subgroup of the root.

Groups for trials `trial 1` and `trail 2` are shown in the figure, and each of these groups have the trial attributes (*name*, *description*). Within the trial groups, a dataset for the sensor `accelerometer1` is shown with the sensor meta-

data as attributes on the dataset.

4 Archiving data

Once the measurement project has been properly documented and organized, e.g. in an HDF5 file, it should be archived on a permanent storage solution. This means that the HDF5 should *not* be left on a USB stick, external HDD or on a personal computer, but sent to a server designed and dedicated to archiving data. The server should as a minimum have hardware and software redundancy, and also a plan to maintain the server.

There exists a plethora of commercial and free solutions for this, for those project associated with a university or research insitution, the institutional library will likely have their own arching solution or may help you in the process of archivhing the data.

If you are associated with the Dynamics group (Department of Structural Engineering, NTNU), contact Gunnstein T. Frøseth about the `Munin` project which is the groups own server.

5 Conclusion

This document has presented the essential metadata for documenting a measurement project for future use by yourself and others. The HDF5 file format has been recommended for practical implementation of the documentation of the measurement project.