

SVN repository structure for SANDBOX data

Erik Hendriks

TU Delft, Faculty of Civil Engineering and Geosciences

August 21, 2019

1 Introduction

The subversion repository is used to store the raw data and processing scripts for the SANDBOX project. This document discusses the structure and conventions of the repository.

The structure of the repository should be considered at two levels:

macro The global directory structure of all datasets Figure 1

micro The subdirectory (Figure 2) and file structure per dataset.

This memo summarizes the reasoning behind the macro structure and the conventions to take into account (mainly micro structure).

For general information on repository structure and conventions the reader is referred to the OpenEarth wiki.

2 General structure

2.1 Macro structure

Starting point for the macro structure is to group similar data (same type/instrument/format) as one dataset. Main reason is that this provides the most clear structure and stimulates to develop and maintain dataset specific processing scripts in one place. In postprocessing, these datasets can be split in subsets containing a specific campaign or experiment.

To facilitate the user, at the highest level (trunk/), four main domains are distinguished (see Figure 1). These measurement domains are subsequently filled with the different datasets.

2.2 Micro structure

The micro structure, being the directory structure at dataset level, is supposed to be in line with Figure 2, as commonly used in OpenEarth context. This structure has previously been used in the Kustgenese 2.0 and Zandmotor research projects.

The subdirectories are described here:

raw/ all raw data should be stored here. If many files or various experiments/campaigns are involved, it is highly recommended to create a sensible subdirectory structure.

`scripts/` all scripts should end up here

`doc/` a separate directory for documentation is optional

A `readme.txt` file on the highest level is very helpful. This readme file contains basic information on the dataset and explanation on possible artifacts in the data or essential information for the interpretation of the data (e.g. an instrument was replaced or displaced during the measurement timeseries). In addition a `dataset_details.cfg` file is added automatically (or) by the repository moderator. This file contains version, location, contributors and dataset name as well as information on the availability of `readme.txt`, raw data, scripts. Information of the contact person for the dataset and the owner should be added manually.

3 General structure

To illustrate the macro and micro structure, we have visualised the directory tree for the *hydrodynamics* and *morphology* domains in Figure 3 and Figure 4. Here we see, that for each dataset within these domains, there are subfolders that hold the raw data, scripts, metadata and documentation. Some of these datasets also contain auxiliary data, i.e., calibration curves or other data that are required to interpret the data.

4 Conventions

- English is the primary language at least for directory and file naming.
- Prevent using spaces and capitals in directory names, use underscores instead.
- Informative naming
- Date format: `yyyy_mm_dd` or `yyyymmdd`
- Timezone: UTC

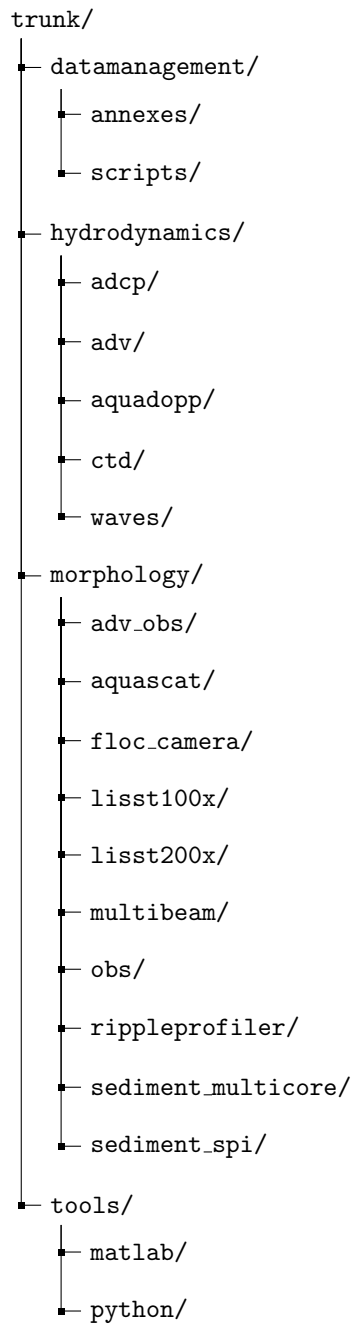


Figure 1: Main directory structure at trunk level

```
<dataset>/
├── raw/
│   ├── yyyy_mm_location_code
│   └── 2017_10_egmond_sbns
├── scripts/
│   ├── raw2nc.py
│   └── ...
├── doc/ (optional)
└── readme.txt
```

Figure 2: Template subdirectory structure for each dataset

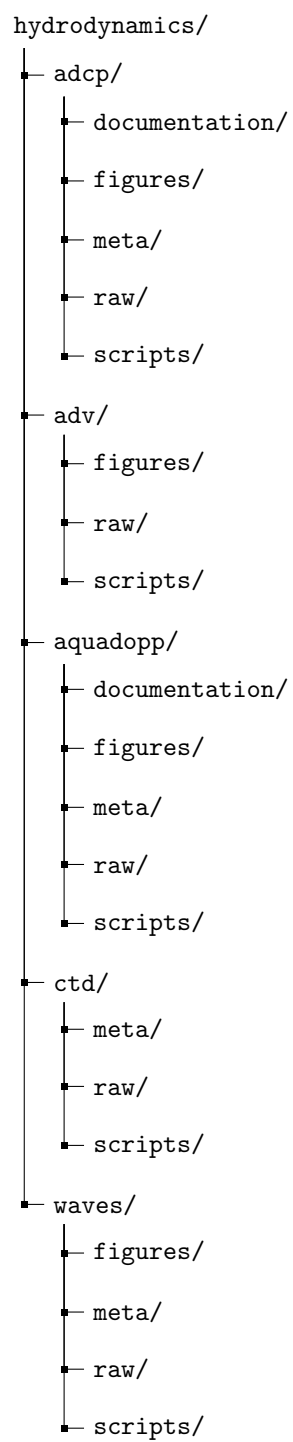


Figure 3: Structure for hydrodynamics domain subdirectory

```
morphology/
├── adv_obs/
│   ├── calibration/
│   ├── documentation/
│   ├── figures/
│   ├── meta/
│   ├── raw/
│   └── scripts/
├── aquascats/
│   ├── raw/
│   └── scripts/
├── floc_camera/
│   └── raw/
├── lisst100x/
│   ├── auxiliary/
│   ├── documentation/
│   ├── figures/
│   ├── meta/
│   ├── raw/
│   └── scripts/
├── lisst200x/
│   ├── raw/
│   └── scripts/
├── multibeam/
│   └── raw/
├── obs/
│   ├── auxiliary/
│   ├── documentation/
│   ├── meta/
│   ├── raw/
│   └── scripts/
├── rippleprofiler/
│   ├── documentation/
│   ├── figures/
│   ├── inter/
│   ├── meta/
│   ├── raw/
│   └── scripts/
├── sediment_multicore/
│   ├── raw/
│   └── scripts/
└── sediment_spi/
    ├── raw/
    └── scripts/
```

Figure 4: Structure for morphology domain subdirectory