LIT-Dataset - Setup Guide

*Laboratory for Innovation and Technology in
Embedded Systems and Energy (LIT)
Universidade Tecnologica Federal do Paraná - UTFPR*

*Bruna M. Mulinari - brunamachadomulinari@gmail.com
Hellen C. Ancelmo - hellen@alunos.utfpr.edu.br
Adriano R. de A. Gamba - adrianogamba@alunos.utfpr.edu.br
Lucas Tokarski - tokarski.lima@gmail.com
Luiz A. Bernardi - lbernardi@alunos.utfpr.edu.br
Lucas da S. Nolasco - lucasnolasco@alunos.utfpr.edu.br
Douglas P. B. Renaux - douglasrenaux@utfpr.edu.br
André E. Lazzaretti - lazzaretti@utfpr.edu.br
Fabiana Pottker - fpottker@utfpr.edu.br
Robson R. Linhares - robson@dainf.ct.utfpr.edu.br
Carlos R. E. Lima - erig@utfpr.edu.br
Elder Oroski - oroski@utfpr.edu.br*

Abstract

The LIT-Dataset is composed by three Subsets of voltage and current data collected as electrical loads were switched on and off. It was conceived and engineered to provide data for evaluation of NILM (Non-Intrusive Load Monitoring) Systems.This guide presents the information to setup and use the LIT-Dataset, which is available for download. Also, it describes how the binary data in the dataset can be converted into MATLAB readable data.

Contents

1 Overview

The LIT-Dataset consists of three subsets, each with its own purpose:

1. *Synthetic:* waveforms were collected with a jig that precisely controls the on and off times in a scenario of controlled load shaping. The jig controls each individual load, up to eight, using TRIACs and relays, hence, allowing for on-times at specified angles of the mains sine wave. The on and off times are recorded in the waveform with a resolution better that 5 ms, allowing the identification of the mains semi-cycle where the "on" or "off" load-event occurred. Typically waveforms are 30 seconds long at sampling rates above 15 KHz, achieving 256 samples per cycle.

2. *Simulated:* several classes of loads were simulated in a MATLAB/SIMULINK environment, and validated against actual loads. The use of simulated loads allows us to modify the load characteristics so to create scenarios that would be very difficult/complex to create in the real world.

3. *Natural:* waveforms were collected in an actual environment (residential, research, commercial, industrial) over longer periods of time. Sensors are attached to each load so to precisely detect and record the on and off times of each load, while the waveforms record the aggregate current and voltage of all monitored loads.

2 Getting Started

To use the LIT-Dataset properly, follow the steps:

1. Access the website were further information is available on this dataset. Before downloading the dataset please register using the link available on the website. After completing the registration form an e-mail is sent with he password to unlock the dataset.

2. Download this .zip file with the file structure for the LIT-Dataset. Select a local folder where the LIT-Dataset will be stored in your computer; expand this .zip file there.
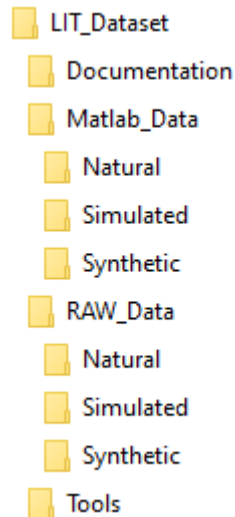

Figure 1: LIT˙Dataset file structure as downloaded in this step.

3. Three links are available on the website to download each one of the Subsets. Each subset may be downloaded as a single .zip file or in parts (Synthetic has 4 parts and Simulated has 6 parts).
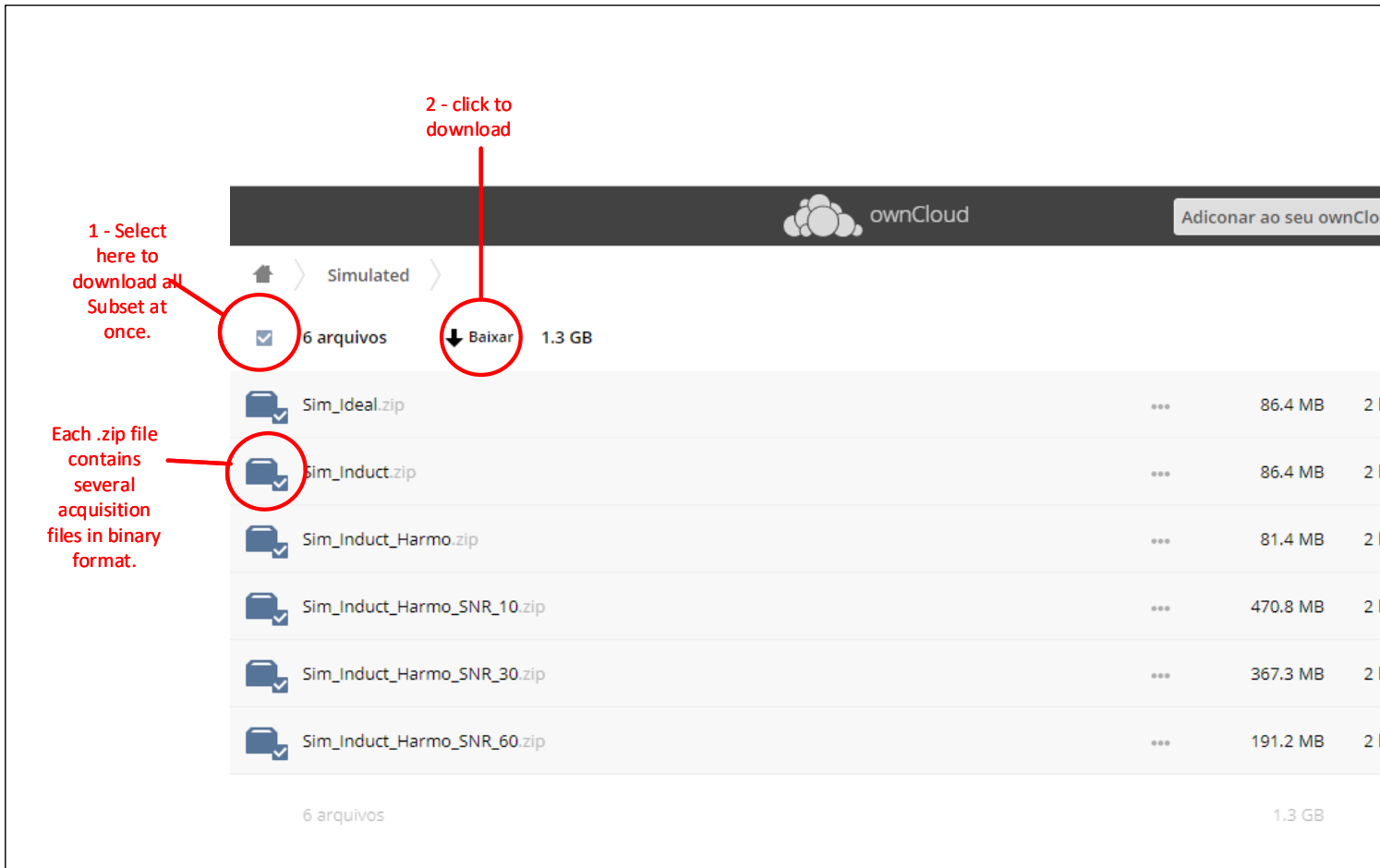

Figure 2: Downloading a Subset

4. Extract the .zip file (using the password sent by e-mail) in the corresponding folder in the RAW_Data folder. Since (in this example) all 6 .zip files were downloaded together, first you unzip the Simulated.zip file to obtain the six .zip files which can then be expanded (as a group) using the password. The waveforms in binary format (as described in format specification document available on the website) are then accessible.
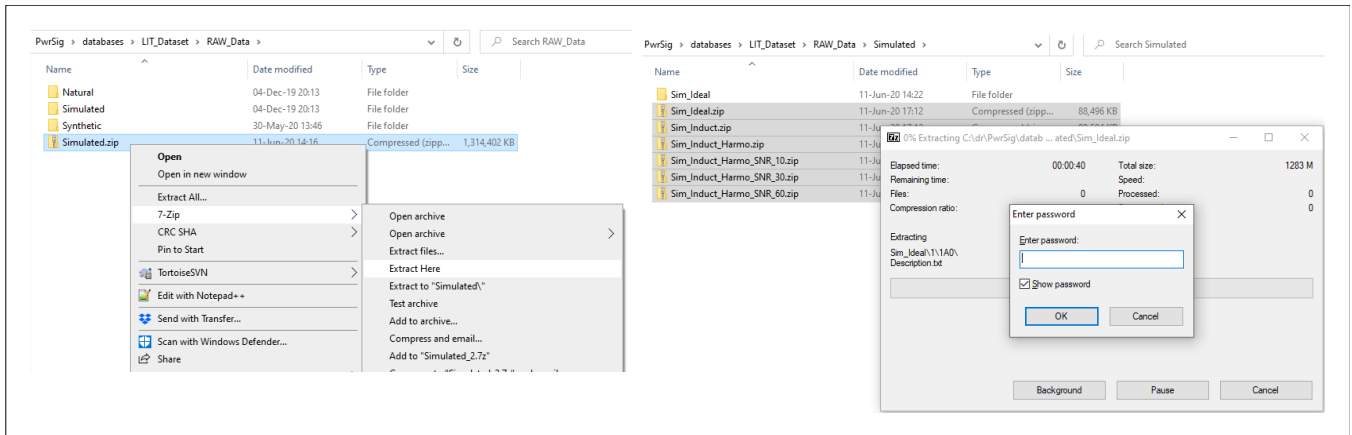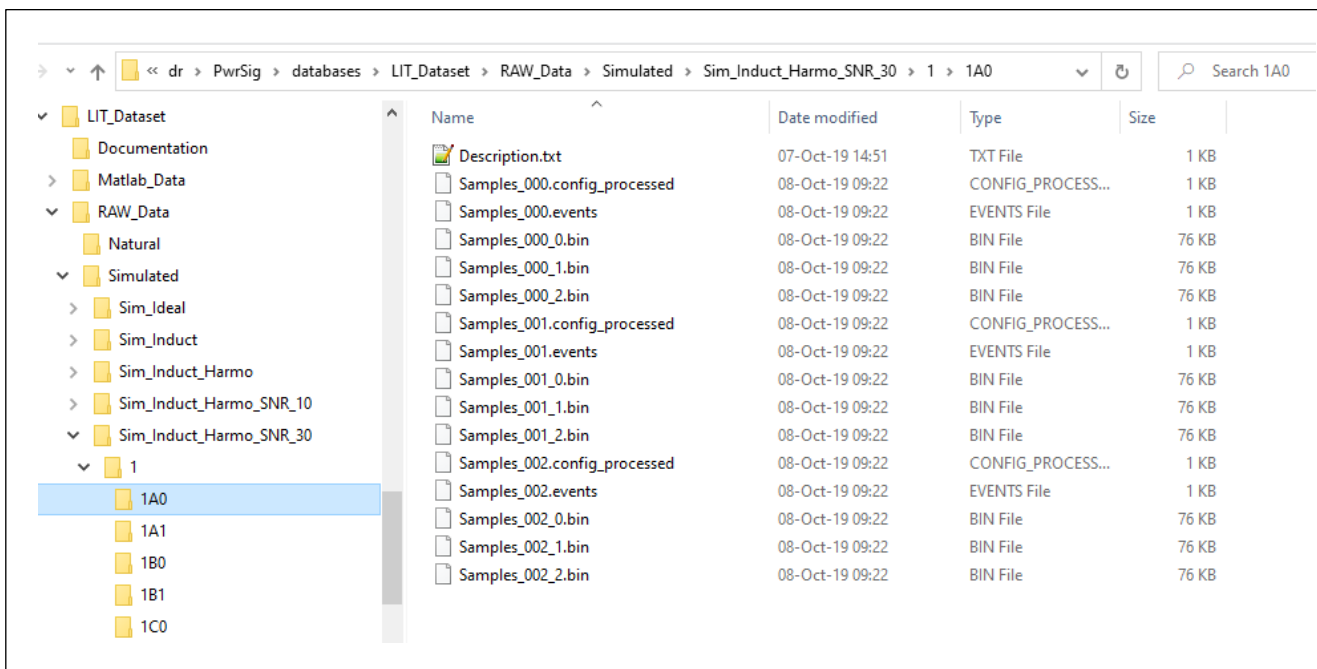
Figure 3: Dataset files expansion

Figure 4: View of one of the expanded folders

5. (optional) Run the MATLAB script (as described in Section section 5 to convert the binary files into MATLAB readable data. This step takes about 2 hours for the Synthetic subset and 3 hours for the Simulated subset.
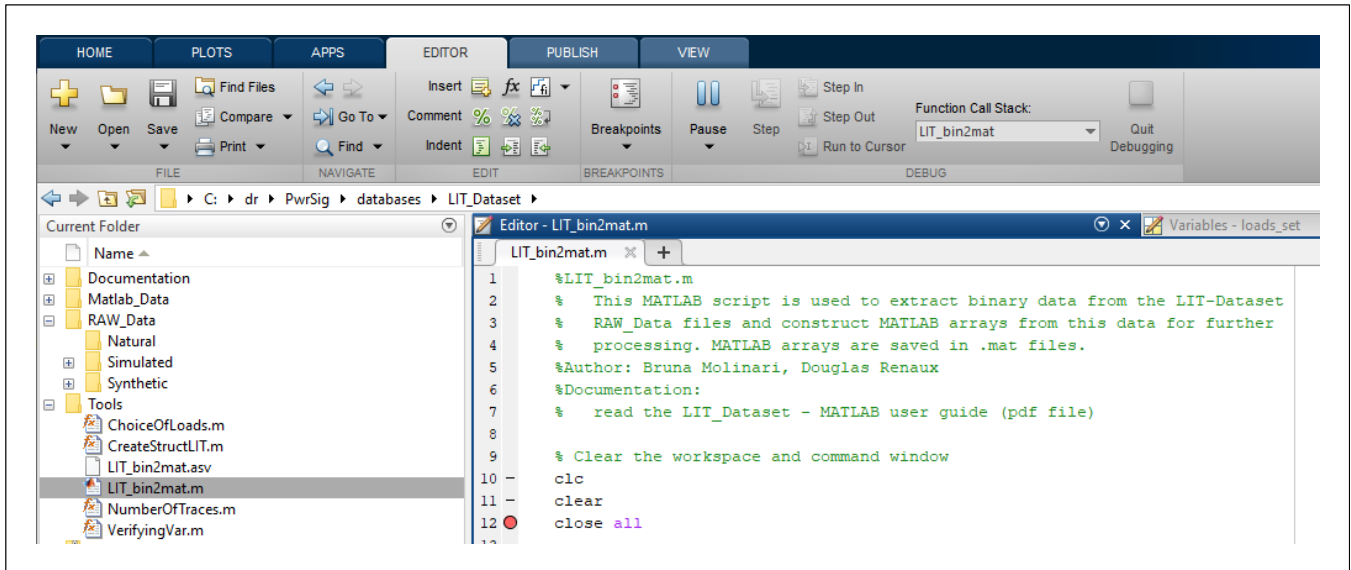

Figure 5: View of one of the expanded folders

! → Remember: check out at our website for available updates. New files are added over time.

3   LIT-Dataset File System Structure

The file system structure presented here is required for proper use of the binary-to-matlab conversion as well as for future updates of the dataset files.

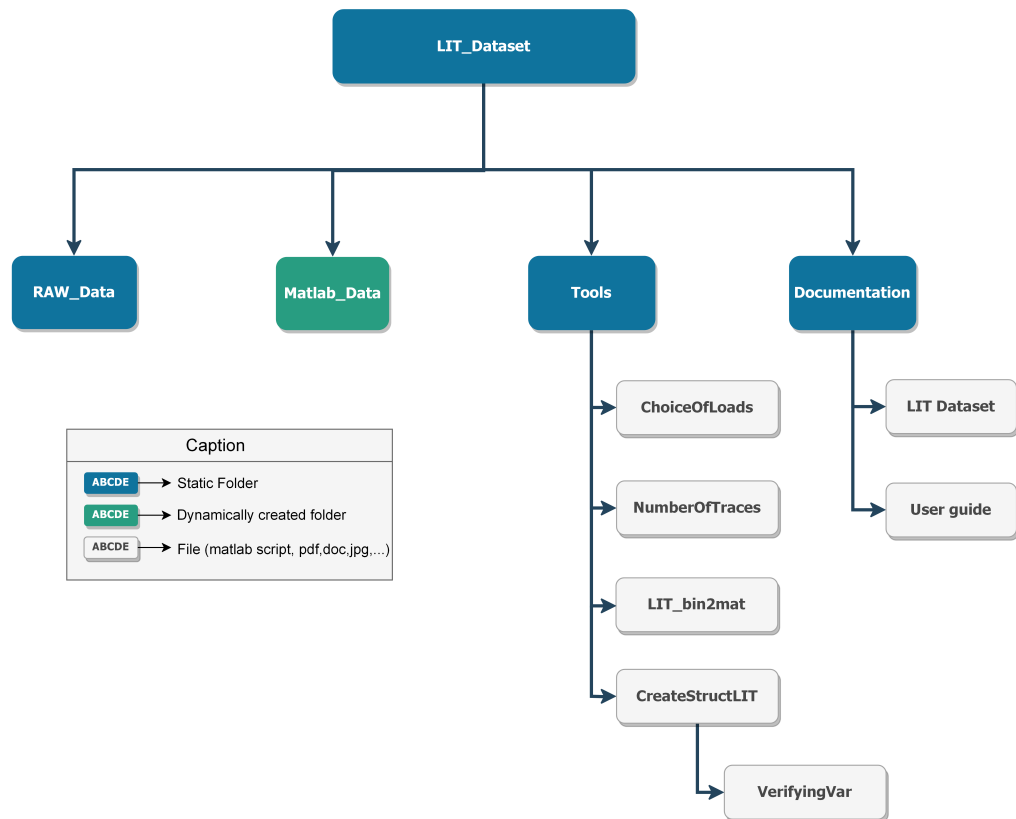The file system structure is presented in Figure 6 and described below:

Figure 6: File system structure (top level).

RAW_Data: This folder holds the raw acquisition files (in binary format) as downloaded from the website. Its contents is detailed in Figure 7. Note that inside the *RAW_Data* folder there is one folder for each subset (Synthetic, Simulated and Natural). Each subset is further decomposed in folders with single load acquisitions (1), two concurrent load acquisitions (2) and so on. Further documentation is available in the pdf files in the Documentation folder.
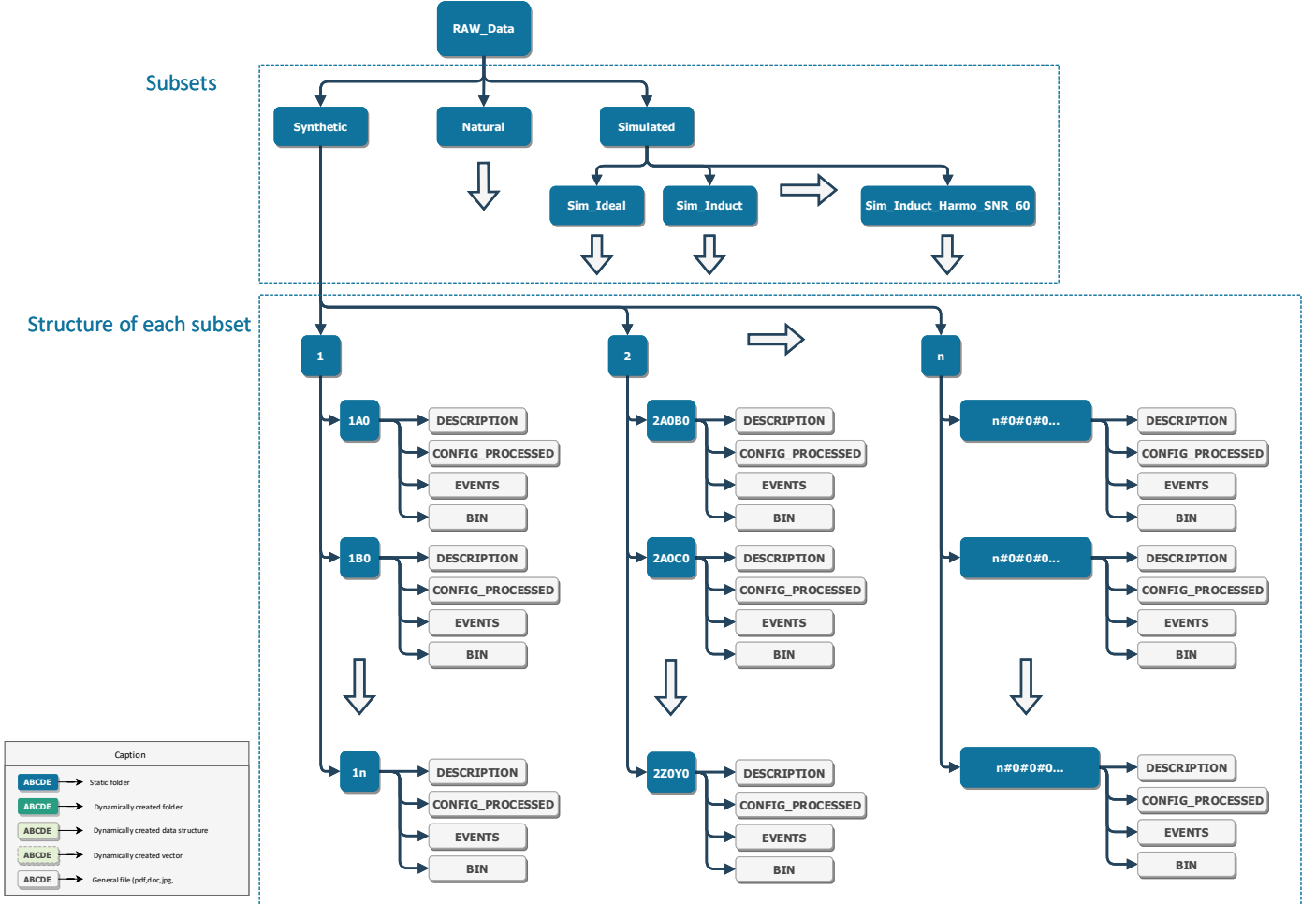

Figure 7: RAW_Data folder structure.

Matlab_Data: This folder is created by the MATLAB script (*LIT_bin2mat.m*) to store the generated data structures (MAT-File Format).
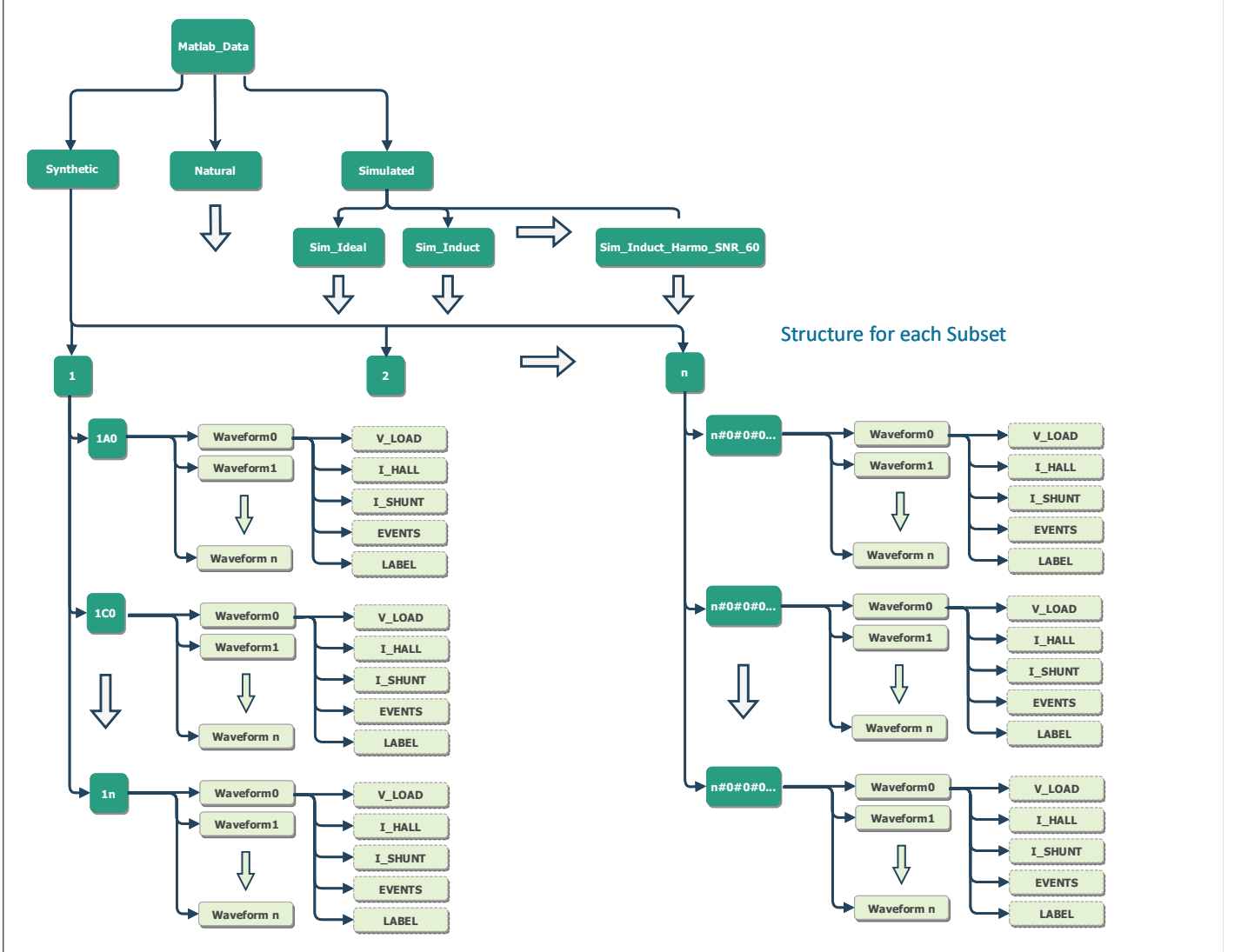

Figure 8: Matlab_Data folder structure.

Tools: This folder contains the Matlab functions that convert the binary files into .mat files. *LIT_bin2mat.m* is the main function (see Section 5) and calls the others. The function *ChoiceOfLoads* is used to select which acquisitions are converted from binary files to .mat files. The *NumberOfTraces* function identifies the number of waveforms that the selected acquisition contains. *CreateStructLIT* performs the conversion of the binary format into a .mat file.
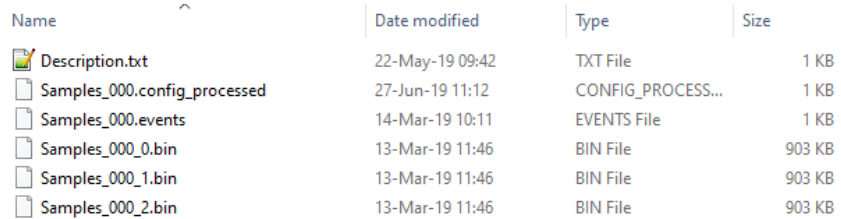
Documentation: This folder holds the documentation related to the LIT dataset as pdf files. It is an auxiliary folder that does not affect the behavior of either the dataset or the tools.

4   LIT-Dataset Contents

In the context of the LIT-Dataset, the following concepts apply:

Waveforms: A waveform is a sequence of voltage and aggregated current samples collected over a specific time. These samples are stored in a file in binary format to reduce file size. In literature, such a file is also called a trace.

The files presented in the figure below belong to the same waveform.

Name	Date modified	Type	Size
Description.txt	22-May-19 09:42	TXT File	1 KB
Samples_000.config_processed	27-Jun-19 11:12	CONFIG_PROCESS...	1 KB
Samples_000.events	14-Mar-19 10:11	EVENTS File	1 KB
Samples_000_0.bin	13-Mar-19 11:46	BIN File	903 KB
Samples_000_1.bin	13-Mar-19 11:46	BIN File	903 KB
Samples_000_2.bin	13-Mar-19 11:46	BIN File	903 KB

Figure 9: Files that compose a waveform.

Description.txt: is an ASCII file that describes the load(s) being monitored and the time where powr events occur.

Samples_000.config_processed: describes the conversion values required to interpret the binary values in the next files.

Samples_000.events: has the timestamps where the load events occurred.

Samples_000_x.bin: has the samples (in binary) for sensor x. Sensors are either voltage sensors or current sensors. The number of sensors and their characteristics is described in Samples_000.config_processed.

Load Configuration: A load configuration is a description of what loads compose a specific waveform recording and the times when each of these loads are turned on or off.

Acquisition File Set: For each load configuration, the LIT-Dataset has stores several waveforms, each recorded with a different power-on mains angle. For the Synthetic Subset, since the loads are energized by TRIACS, precise power-on timings are possible. Hence, each load configuration has 16 recordings, correspondig to power-on when mains are at angle 0°, 22.5°, 45°, and so on.

Each .zip file in the LIT-Dataset corresponds to a single Acquisition File Set. For the Synthetic Subset, each .zip file contains 81 files, including descriptions, events and binary samples.

5   Using the function LIT_bin2mat

The *LIT_bin2mat.m* is a MATLAB script that reads the raw (binary) files in the *RAW—˙Data* folder and creates the corresponding data structures (stored as MAT files) in the *Matlab˙Data* folder .

In the following sections we describe how the script works and how it can be configured.

5.1 Input parameters description

These are the configuration parameters that can be selected by the user.

Subset: This parameter selects a subset of the LIT-Dataset.

Parameter Type: char array;

Options:

- 'Synthetic';
- 'Sim_Ideal';
- 'Sim_Induct_Harmo';
- 'Sim_Induct_Harmo_SNR_10';
- 'Sim_Induct_Harmo_SNR_30';
- 'Sim_Induct_Harmo_SNR_60';
- 'Natural';

  The variations of the Simulated Subset are described in the report: "LIT-Dataset Description of Loads and Subsets".

5.2 Output parameters description

The output are waveform files at a given power-on angle (triggering angle) selected by the user. The file name is composed by:

$$Waveform_- + ID_{ang} \tag{1}$$

Where the ID_{ang} refers to the ID Angle chosen, see the document "LIT-Dataset Description of Loads and Subsets".

The waveform files include:

- iShunt: A vector containing the current signal collected by shunt sensor.
- iHall: A vector containing the current signal acquired by hall sensor.
- vGrid: A vector containing the voltage grid signal acquired.
- events: A numbered vector with the turn-on and turn-off events information. The value '1' is set at the exact sample of a turn-on event and its complementary '-1' is set for a turn-off event. All the other samples are set with a null value.
- labels: A string valued vector to correlate each turn-on and turn-off event indicated at the r_events and r_vector with the correpondent load ID.
- duration: This is a single integer to identify the acquisition duration in seconds.

Each waveform is generated inside the configuration folder of loads chosen by the user.

! → If the folder structure, described in Section 3 is kept. The Output folder will be generated automatically.