pISA-tree: Standard project directory tree

A. Blejec, ... , K. Gruden

2018-03-29

# What is the pISA-tree?

This set of batch files is used to create standard directory tree for research projects.

## Layers

At the top layer, you have to create a directory (pISA\_projects root directory) which will contain future projects. The root directory is the place for initial batch files. The first step is to copy/download the batch files from the [pISA GitHub](https://github.com/ablejec/pISA/) into the root directory.

A **Project** is organized as a collection of one or more **investigations**. An **investigation** is similarly organized as a collection of one or more **studies**. Each **study** has it's own collection of one or more **assays**. Assays can be of specific type (e.g. connected to MicroArray, NGS, Modeling, Statistical analysis, ...) and are structured accordingly.

## Batch files

* makeproject.bat - makes a new **project** directory tree
* makeInvestigation.bat - makes a new **investigation** directory tree (subdirectory tree within the **project**)
* makeStudy.bat - makes a new **study** (subdirectory tree within the **investigation**)
* makeAssay.bat - makes a new **assay** (subdirectory tree within the **study**)

# Creation of the directory tree

The directory tree is a way to enforce the subordination of pISA layers. To emphasize the layer type, directory names are constructed automatically using the standard prefix and short layer ID. Standard prefixes are:

* \_p\_for project
* \_I\_for investigation
* \_S\_for study
* \_A\_for assay

## project

To create a new project, run (double click) the file makeProject.bat and enter the project short name (ID). This will make a direcory tree, metadata files and a local copy of makeInvestigation.bat. Short project name (ID), automatically prefixed with \_p\_, is used as the name of the directory (no spaces or special characters, file name conventions apply!!). If you set the project ID as bla the project directory name will automatically become \_p\_bla.

## Investigation

To create a new investigation, run (double click) the file makeInvestigation.bat and enter the study short name (ID). This will make a directory tree, metadata files and a local copy of makeStudy.bat. Short investigation name (ID), automatically prefixed with \_I\_, is used as the name of the directory (no spaces or special characters, file name conventions apply!!). The investigation directory name for the investigation ble will be \_I\_ble.

## Study

To create a new study, run the makeStudy.bat and enter the study ID (Short name). This will make a direcory tree with several standard folders, metadata and auxiliary files and a local copy of makeAssay.bat. The study folder name will be \_S\_blu for a study with the short name (ID) blu.

## Assay

Analyses for each study are collected in the folder of that study. To make a new assay, run the makeAssay.bat file.

You will be asked to enter assay **Class** [Wet/Dry]:

* Wet: measurements on the biological material (MicroArray, NGS, PCR, ...)
* Dry: process data (Statistcs, Modelling, ... )

Then you will enter the assay **Type** (see documentation for covered types: NGS, MA, STat ...) and assay ID (Short name, for example ASSAY1). Short assay name and type (separated by '-' and prefixed by \_A\_) are used as the name of the assay directory tree (for example: \_A\_ASSAY1-NGS).

Folders in assay directory trees for different Classes slightly differ, according to the need of the specific Class.

Assay classes and types are defined as subfolders of the Templates subfolder. An example is ../Templates/Wet/RT. To add another type of *Wet* assay (named *mytype*) , one can create a new subfolder with the name *mytype*: ../Templates/Wet/mytype. To add another class, create subfolder *myclass*: ../Templates/myclass.

## Metadata files

Each level has a *metadata* file, listing the informative items for that level. Metadata files are tab delimited files with two columns:

1. item name (ended by a colon)
2. item metadata or value

Item metadata can be some text (for example investigator's name or a longer metadata of the study, study, and analysis) or a value (for example the path to *phenodata* file). Each item should be typed in one line. Special escaped characters for line break (\n) and tab (\t) are allowed. Be careful if the metadata contains prime symbol (' ,as in 5'), it is safer to spell it, like 5-prime.

An example of the metadata entry is:

Investigator: Miha Mihav  
Phenodata: ./data/phenodata.txt

The project level directory contains the file common.ini, which is organized in the same way as metadata files. This file contains fields and metadatas that are fixed for the project, for example principal investigator name. Information in this file will be appended to the metadata files for all investigations, studies, and assays within a project.

## Analytes files

In addition to the basic items, one can use also assay specific items. The assay specific items are prespecified in the Analytes.ini files, placed in the *Class/Type* subfolder of subfolder Templates.

The Analytes.ini files are plain text files. You can open and edit them by Notepad, WOrdpad or Excel (Use 'Open with ...'). Each line represents one Item-Value pair, separated by tab character (\t). The first part (Item name) will apear as the assay specific question during the assay creation. The second part (Item value) will be offered as the default (or first) value for information entry.

An example of the Analytes.ini file:

Item name  
Isolation Protocol Rneasy\_Plant  
Operator John/Bob  
Date Homogenisation %today%

# Auxiliary batch files

## showMetadata.bat

Collects all metadata files in a tree below the current level. Metadata are typeset in the file **METADATA.MD**. (plain text file in a markdown format)

## xcheckMetadata.bat

Checks all metadata files for missing required information (\*) in a tree below the current level. Produces the file named **xCheckMetadata.md** which is similar to the one produced by showMetadata.bat but lists only lines with asterisks (\*).

## showTree.bat

List a directory tree below the current level in the file **TREE.TXT**.

## update.bat

Replaces batch files in existing tree with the updated versions in the project directory. Use after update from [pISA GitHub](https://github.com/ablejec/pISA/)