

pISA-tree: Standard Directory Tree as a Support for Reproducible Research

Andrej Blejec, Spela Baebler, Anna Coll, Marko Petek,
Živa Ramsak, Kristina Gruden

andrej.blejec@nib.si

Dept. of Biotechnology and Systems Biology
National Institute of Biology, Ljubljana, Slovenia



October 25, 2016

- need to **locally organize data** within small/moderate size research units
- **find, exchange and reuse data** from different experiments
- **use** and enhance what **researchers are already used to**
- provide **enough metadata to enable upload** of data to public repositories
- promote use of **versioning systems** (e.g. git and GitHub) where applicable

- directory tree structures: TeX/locattexmf, R/packages, ...
- personal directory trees for reproducible statistical reports (using R/Sweave/knitr/Rmarkdown)
- **I**nvestigation/**S**tudy/**A**ssay levels, to be compliant with the ISA framework (at first we recognized levels as Study/Experiment/Analysis)

plSA-tree:

project directory tree

compliant with FAIRDOM/ISA framework

- (local) Windows based systems (extendable to Linux)
- project directory tree with embedded **I**nvestigation/**S**tudy/**A**ssay levels
- '**make...**' batch files are provided at each level to create substructures:
 - **makeInvestigation.bat**
 - **makeStudy.bat**
 - **makeAssay.bat**
- template 'description' files at each level to **encourage metadata entry** (text files with key/data pairs)

▶ Skip tree representation

pISA tree - Investigation folder and description

The image shows a file explorer window with the following structure:

- Investigation1
 - _STUDIES
 - Study1
 - _ASSAYS
 - data
 - presentations
 - results
 - Study2
 - _ASSAYS
 - Assay1-NGS
 - Assay2-Stat
 - data
 - presentations
 - results
 - presentations

The selected file, `_INVESTIGATION_DESCRIPTION.TXT`, contains the following text:

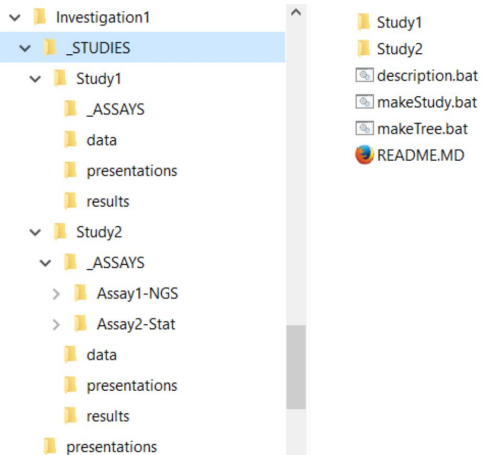
```
INVESTIGATION
SHORT NAME: Investigation
INVESTIGATION DESCRIPTION
INVESTIGATOR:      Miha Mi
PHENODATA:        ./data/phenod
FEATUREDATA:      ./data/

STUDY:           Study1

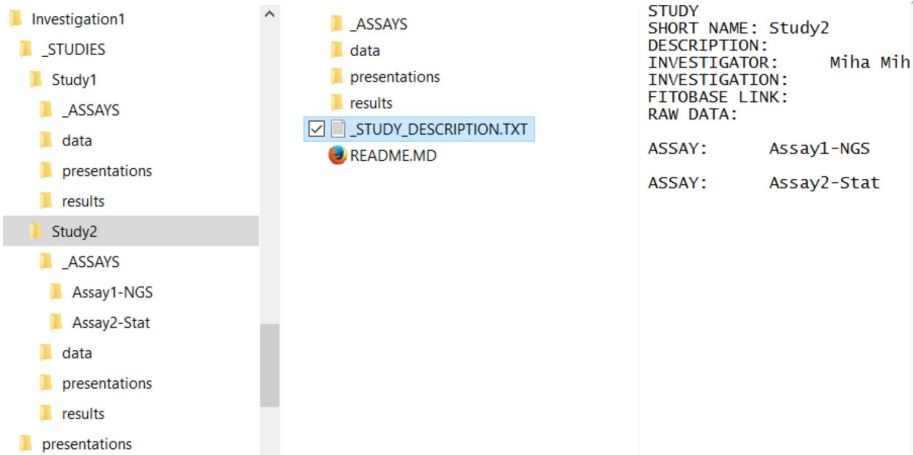
STUDY:           Study2
```

At the bottom left, there is a button labeled "Skip tree representation".

pISA tree - Studies are collected in /_STUDIES folder



pISA tree - Study folder and description

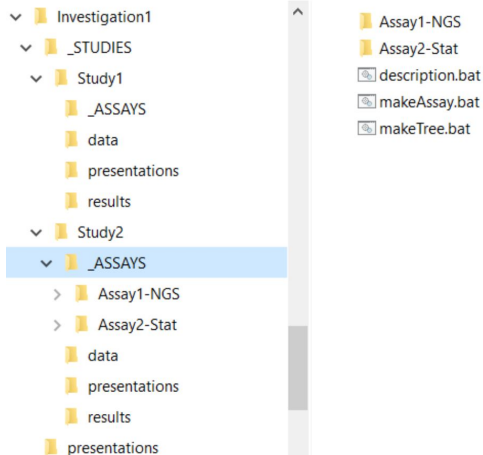


The image displays a file explorer interface for a pISA tree. On the left, a tree view shows a hierarchy of folders: Investigation1, _STUDIES, Study1, and Study2 (which is selected and highlighted). Under Study2, there are sub-folders: _ASSAYS, Assay1-NGS, Assay2-Stat, data, presentations, and results. The right pane shows the contents of the selected folder, including: _ASSAYS, data, presentations, results, a checked file named _STUDY_DESCRIPTION.TXT, and a README.MD file. To the right of the file explorer, a preview window displays the content of the selected file:

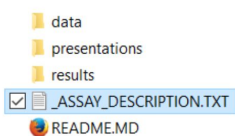
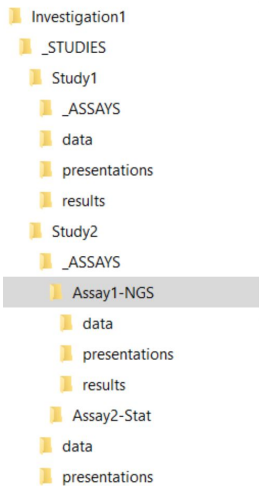
```
STUDY
SHORT NAME: Study2
DESCRIPTION:
INVESTIGATOR:      Miha Mih
INVESTIGATION:
FITOBASE LINK:
RAW DATA:

ASSAY:      Assay1-NGS
ASSAY:      Assay2-Stat
```

pISA tree - Assays are collected in /_ASSAYS folder



pISA tree - Assay folder and description



```
ASSAY
SHORT NAME: Assay1-NGS
DESCRIPTION:
INVESTIGATOR:      Miha Mil
STUDY:
DATA:
```

pISA-tree: project root directory

../project

makeInvestigation.bat

pISA-tree: add Investigation

../project

makeInvestigation.bat

/presentations

Readme.md

Description.txt

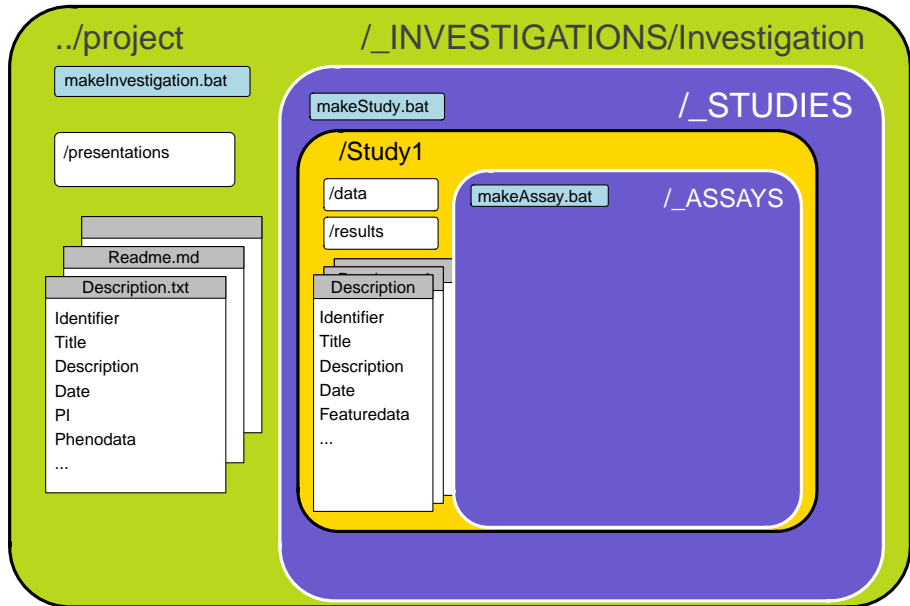
Identifier
Title
Description
Date
PI
Phenodata
...

/_INVESTIGATIONS/Investigation

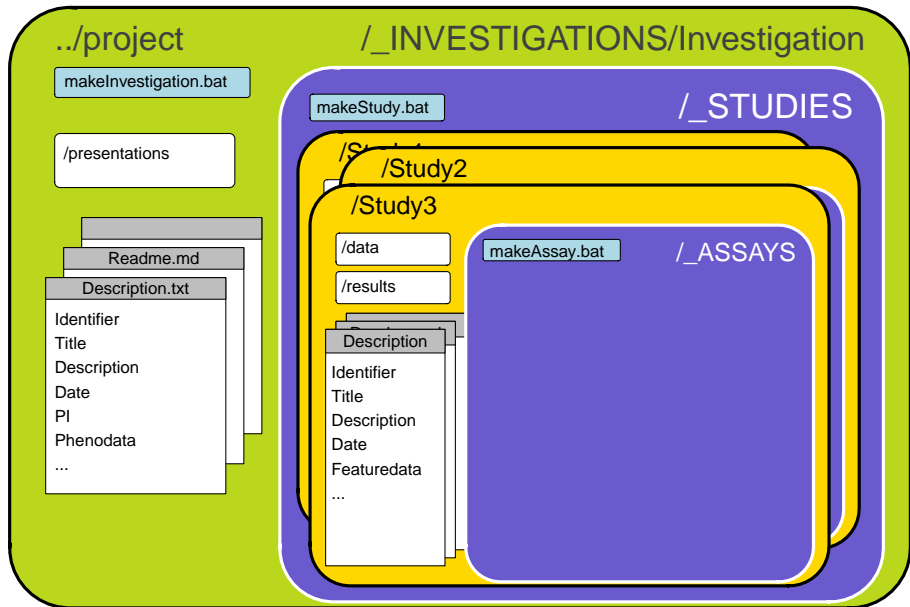
makeStudy.bat

/_STUDIES

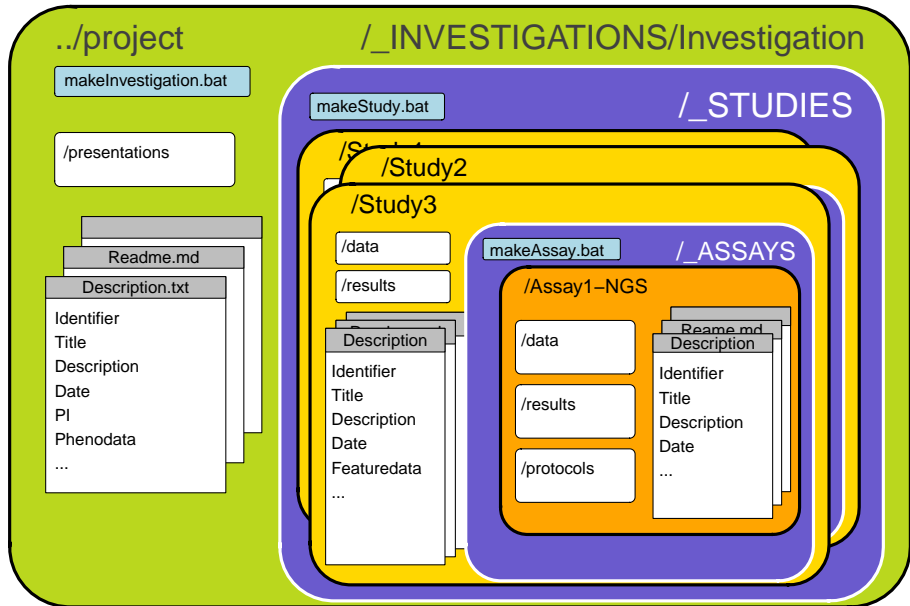
pISA-tree: add a Study within /_STUDIES folder



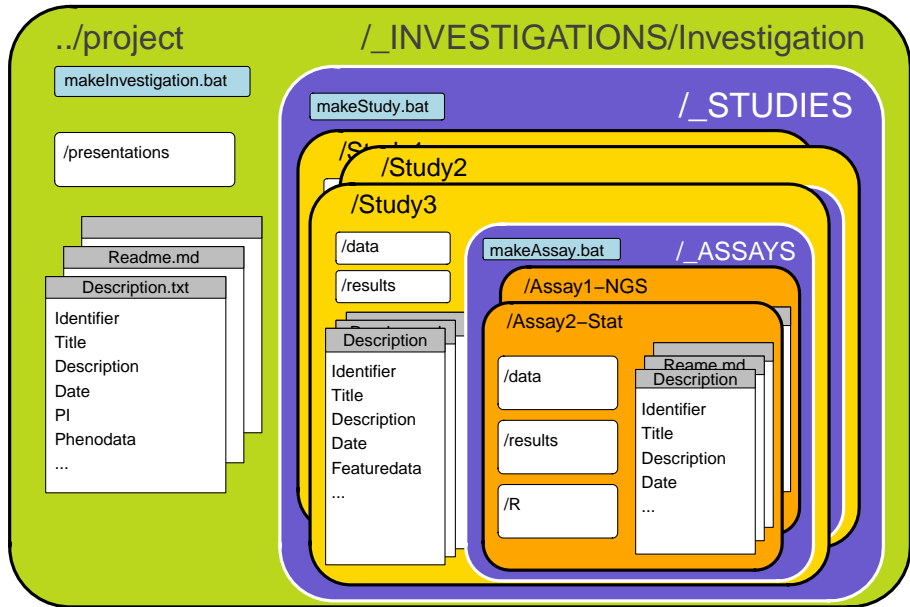
pISA-tree: more Studies



pISA-tree: add an Assay within /_ASSAYS folder



pISA-tree: another Assay



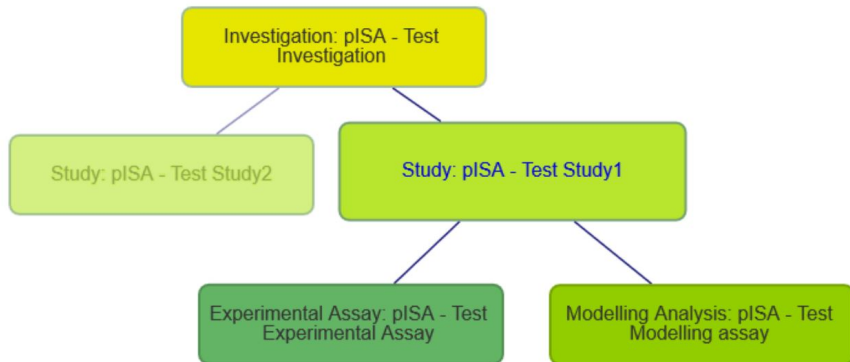
Description files hold metadata

- Description file for each level
- Will follow recommendations of FAIRDOM/SEEK paradigm
- Provide mechanisms for production of ISA-tab files

Connection with other solutions

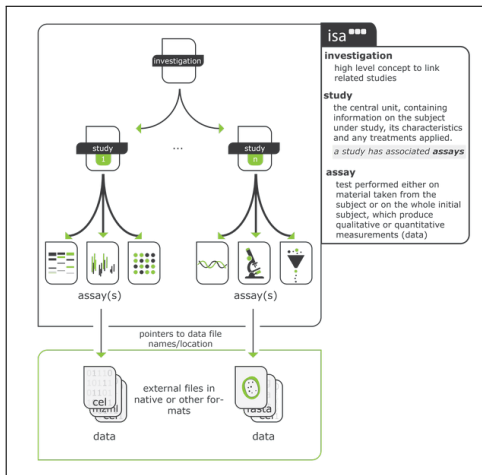
- FAIRdom
- SEEK <https://seek.sysmo-db.org/>
- ISA-tab

Selected item: Study: pISA - Test Study1



<http://demo.seek4science.org/investigations/18>

ISA-tab model structure



<http://isa-tools.org/format/specification/>


Thank you

To learn more and discuss, visit us at our poster ...

P16



pISA-Tree: Standard Directory Tree as a Support for Reproducible Research

Wednesday October 26, 2016: 10.20-11.20

pISA-tree: Standard Directory Tree as a Support for Reproducible Research 

Andrej Blejec, Spela Baebler, Anna Coll, Marko Petek, Ziva Ramsak, Kristina Gruden

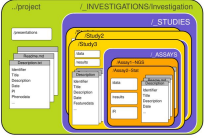
Department of Biotechnology and Systems Biology,
National Institute of Biology, Ljubljana, Slovenia.
andrej.blejec@nib.si



Motivation
Basic idea of science is reproducibility of phenomena and experiments. Reproducibility of data analyses and reports is becoming more and more important. Our aim is to provide a system to store data in a way that can be used for small and moderate size projects and fulfill minimal requirements of ISA-tab and FAIR paradigm.

Methods
Standard directory trees are applicable to research data storage. The main condition is that information is organized in files and we are not interested in the system that enables access to individual line (row) or columns (variable) in a tabular data structure in the file. The tree structure is generated on the fly by use of batch files (on Windows platform) that generate necessary folders and meta-data template files.

pISA directory tree structure



Conclusion
pISA-tree structure relies on directories that are readily available on any computer platform and familiar to use by the researchers. It can support small to moderate projects. Since the translation of meta-data into ISA-tab standard format is not too complex, pISA-tree is a step towards the FAIR paradigm.

Results
We implemented the system of standard directory trees for support of the research in our research test. Most often, our research project can be hierarchically structured into what we call **Investigation** which is composed of several **Studies**. Each individual study can have one or more **Assays**. To reflect this hierarchy, we named such directory tree the **pISA-tree**. To make new levels, we provided three batch files: makeInvestigation, makeStudy, and makeAssay. Special attention is given to Description files that contain meta-information about research, protocol, samples, features or omics. They are in line with the standards accepted for particular assays (e.g. MIQE, MIAM, MIRIAM, ...) which allow easy change of data with other data management services.

<http://andrej.blejec.nib.si/doc>

[\[View poster\]](#)