

pISA-tree: Standard Directory Tree

a Support for Reproducible Research and Data Reusability

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



December 8, 2017

- 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/localtexmf, R/packages, ...
- personal directory trees for reproducible statistical reports (using R/Sweave/knitr/Rmarkdown)
- **p**roject/**I**nvestigation/**S**tudy/**A**ssay levels, to be compliant with the ISA framework (used to be: Study/Experiment/Analysis)

pISA-tree:

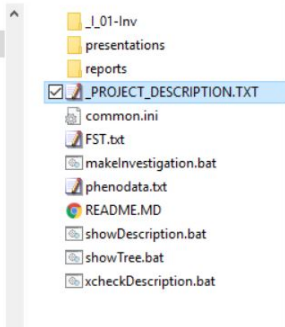
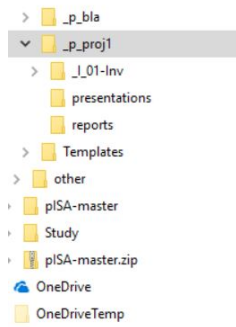
project directory tree

compliant with FAIRDOM/ISA framework

- (local) Windows based systems (extendable to Linux)
- project directory tree with embedded **p**roject/**I**nvestigation/**S**tudy/**A**ssay levels
- '**make...**' batch files are provided at each level to create substructures:
 - **makeproject.bat**
 - **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 - project folder and description



```
project:      _p_proj1
### PROJECT
Short Name:  proj1
Title:      Demo project
Description:  *
Project Path: D:\OMIKE\pISA\main\_p_proj1
Investigator: Name Lastname
##### INVESTIGATIONS

INVESTIGATION:      01-Inv
```

▶ Skip tree representation

pISA tree - Investigation folder and description

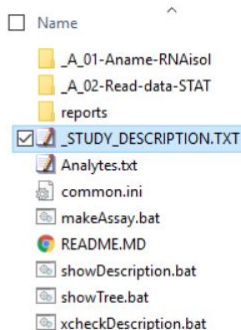
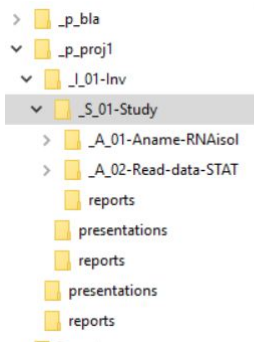
The image shows a file explorer window with a tree view on the left and a file list on the right. The tree view shows a hierarchy starting with `_p_bla`, then `_p_proj1`, and finally `_I_01-Inv` (which is selected). Under `_I_01-Inv`, there is a folder `_S_01-Study` containing subfolders `presentations` and `reports`. Below this are `Templates` and `other` folders, and finally `pISA-master` and `Study` folders. The file list on the right shows the contents of the selected folder, including `_INVESTIGATION_DESCRIPTION.T` (checked), `common.ini`, `FST.txt`, `makeStudy.bat`, `phenodata.txt`, `README.MD`, `showDescription.bat`, `showTree.bat`, and `xcheckDescription.bat`.

```
project:      _p_proj1
Investigation:  _I_01-Inv
### INVESTIGATION
Short Name: 01-Inv
Title:      Demo investigation
Description: Shows an investigation
Investigator: Name Lastname
Phenodata:  ./data/phenodata.txt
Featuredata: ./data/featuredata.txt
#### STUDIES

STUDY:      01-Study
```

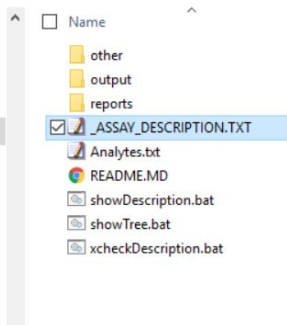
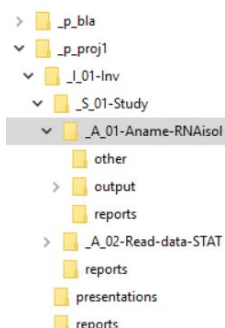
▶ Skip tree representation

pISA tree - Study folder and description



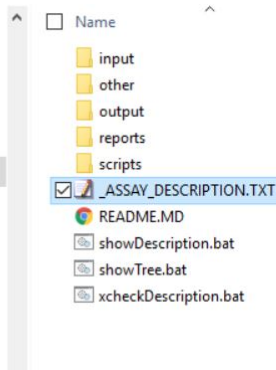
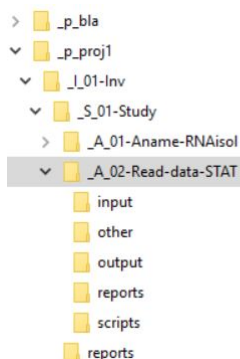
```
project:      _p_proj1
Investigation:  _I_01-Inv
Study:        _S_01-Study
### STUDY
Short Name:   01-study
Title:        Demo study
Description:   *
Investigator:      Name Lastname
Fitobase link:
Raw Data:
#### ASSAYS
ASSAY:         01-Aname-RNAisol
ASSAY:         02-Read-data-STAT
```

pISA tree - Assay folder (Wet) and description



```
project:      _p_proj1
Investigation:  _I_01-Inv
Study:        _S_01-Study
Assay:        _A_01-Aname-RNAisol
### ASSAY
Short Name:   01-Aname-RNAisol
Assay Class:  WET
Assay Type:   RNAisol
Title:        *
Description:  *
RNA ID:       RNA
ng/ul:        ""
Researcher:   *
Comment:      Fields from template
Homogenisation protocol:  slowPrep
Date Homogenisation:      2017-12-08
Another ID:   bla
Data:
Investigator:      Name Lastname
[]
```


pISA tree - Assay folder (Dry) and description



```
project:      _p_proj1
Investigation:  _I_01-Inv
Study:        _S_01-Study
Assay:        _A_02-Read-data-STAT
### ASSAY
Short Name:   02-Read-data-STAT
Assay Class:  DRY
Assay Type:   STAT
Title:        Read raw data
Description:   *
Data:
Investigator: Name Lastname
|
```

../_p_project

makeInvestigation.bat

../_p_project

makeInvestigation.bat

/_I_Investigation

makeStudy.bat

/presentations

/reports

Readme.md

Description.txt

Identifier

Title

Description

Date

PI

Phenodata

..

../_p_project

makeInvestigation.bat

/_I_Investigation

makeStudy.bat

/presentations

/reports

Readme.md
Description.txt
Identifier
Title
Description
Date
PI
Phenodata
...

/_S_Study1

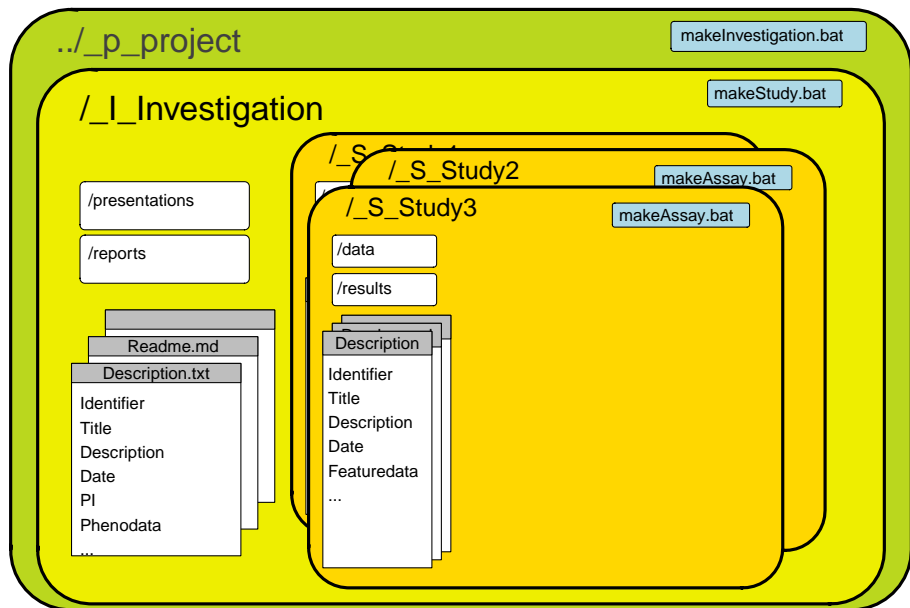
makeAssay.bat

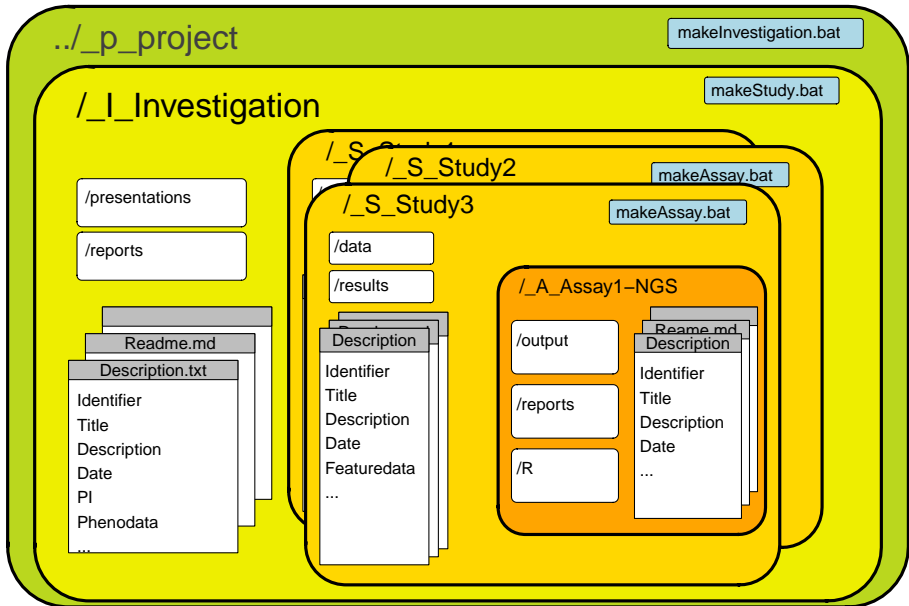
/data

/results

Description
Identifier
Title
Description
Date
Featuredata
...

pISA-tree: more Studies





pISA-tree: another Assay

▶ Tree

../_p_project

makeInvestigation.bat

/_I_Investigation

makeStudy.bat

/presentations

/reports

Readme.md

Description.txt

Identifier
Title
Description
Date
PI
Phenodata
...

/_S_Study2

makeAssay.bat

/_S_Study3

makeAssay.bat

/data

/results

Description

Identifier
Title
Description
Date
Featuredata
...

/_A_Assay1-NGS

/_A_Assay2-Stat

Reame.md
Description

/data

/results

/R

Identifier
Title
Description
Date
...

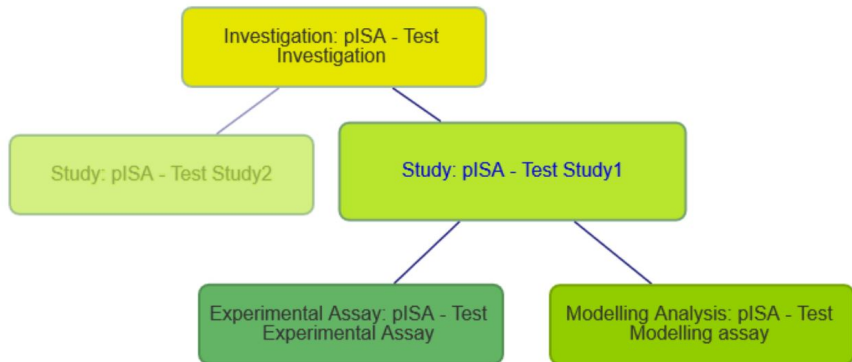
Description files hold metadata

- Description file for each level
- Will follow recommendations of FAIRDOM/SEEK paradigm
- Provide mechanisms for production of ISA-tab files
- Several Assay classes (Wet / Dry / ...)
- Several Assay types (RNAisol, RT, Stat, R, ...)
- Menu-driven Assay metadata entry (to reinforce consistency)

Connection with other solutions

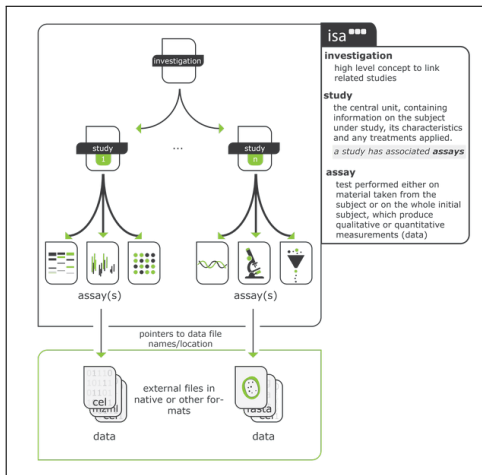
- FAIRdom
- SEEK <https://seek.sysmo-db.org/>
- ISA-tab

Selected item: Study: pISA - Test Study1



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

ISA-tab model structure



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