Bulk uploading from pISA-tree to FAIRDOM using SEEK API and R package seekr

Andrej Blejec
Špela Baebler, Anna Coll, Marko Petek, Živa Ramšak,
Maja Zagorščak, Kristina Gruden

National Institute of Biology
Ljubljana, Slovenia
andrej.blejec@nib.si

May 9, 2022

Motivation

Goal

Automate upload of possibly large number of files to FAIRDOMHub.

Situation

- Data and files are managed using the pISA-tree.
- FAIRDOMHub and pISA-tree share the ISA structure.
- Data generating (WET) and data analysis (DRY) phases generate many files.
- Manual upload of large number of files can be tedious.

Solution

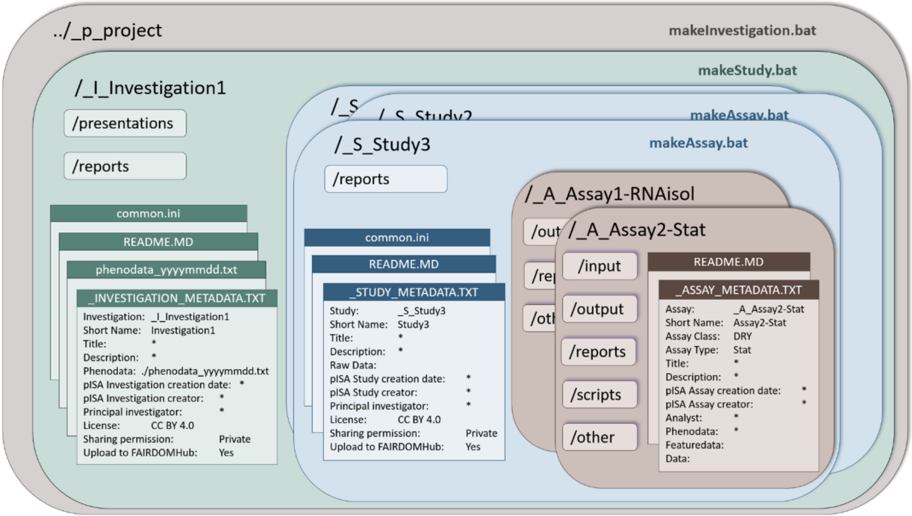
Desired solution

1. Researchers prepare the list of files for upload.
2. Files from the list are automatically uploaded.

Methods

- Employ mapping of ISA structure from pISA-tree to FAIRDOMHub.
- Use SEEK API for upload and information exchange.
- We developed R based package **seekr** on top of the SEEK API.
- Above two steps are performed with two functions:
    1. `skFilesToUpload()`
       Prepare list of files to upload.
    2. `skUploadFiles()`
       Bulk upload of files in the file list.

pISA-tree structure



../_p_project

makeInvestigation.bat

/_I_Investigation1

makeStudy.bat

/presentations

/reports

common.ini

README.MD

phenodata_yyyymmdd.txt

_INVESTIGATION_METADATA.TXT

Investigation: _I_Investigation1
Short Name: Investigation1
Title: *
Description: *
Phenodata: ./phenodata_yyyymmdd.txt
pISA Investigation creation date: *
pISA Investigation creator: *
Principal investigator: *
License: CC BY 4.0
Sharing permission: Private
Upload to FAIRDOMHub: Yes

/_S

/_S_Study2

makeAssay.bat

/_S_Study3

makeAssay.bat

/reports

common.ini

README.MD

_STUDY_METADATA.TXT

Study: _S_Study3
Short Name: Study3
Title: *
Description: *
Raw Data:
pISA Study creation date: *
pISA Study creator: *
Principal investigator: *
License: CC BY 4.0
Sharing permission: Private
Upload to FAIRDOMHub: Yes

/_A_Assay1-RNAisol

/out

/re

/oth

/_A_Assay2-Stat

/input

/output

/reports

/scripts

/other

README.MD

_ASSAY_METADATA.TXT

Assay: _A_Assay2-Stat
Short Name: Assay2-Stat
Assay Class: DRY
Assay Type: Stat
Title: *
Description: *
pISA Assay creation date: *
pISA Assay creator: *
Analyst: *
Phenodata: *
Featuredata:
Data:

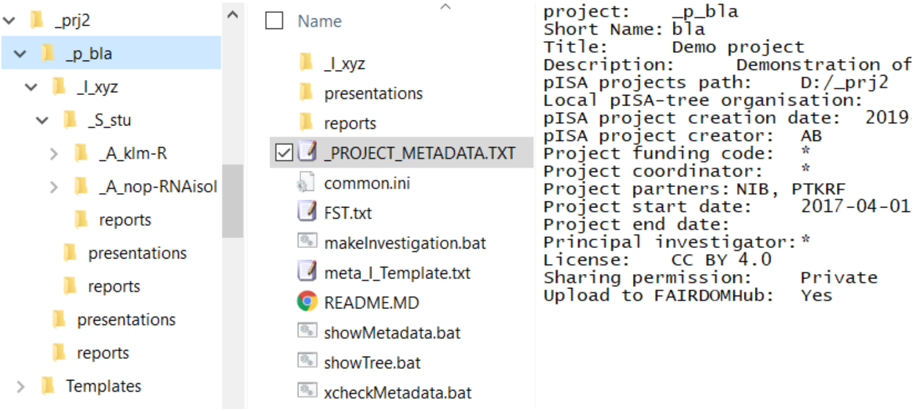
pISA-tree:

project directory tree
compliant with FAIRDOM/ISA framework

- File system based (Windows, extendable to Linux).
- Project directory tree with nested folders for **p**roject/**I**nvestigation/**S**tudy/**A**ssay layers.
- '**make...**' batch files are provided to create layers
  - `makeProject.bat`
  - `makeInvestigation.bat`
  - `makeStudy.bat`
  - `makeAssay.bat`
- Template 'metadata' files at each layer to **encourage metadata entry** (text files with Key/Value pairs).
- Allows **files in all ISA layers**.

▸ To **seekr** features

pISA tree - project folder and metadata
Researcher's view



- _prj2
  - _p_bla
    - _I_xyz
      - _S_stu
        - _A_klm-R
        - _A_nop-RNAisol
        - reports
      - presentations
      - reports
    - presentations
    - reports
  - Templates

☐ Name

- _I_xyz
- presentations
- reports
- ☑ _PROJECT_METADATA.TXT
- common.ini
- FST.txt
- makeInvestigation.bat
- meta_I_Template.txt
- README.MD
- showMetadata.bat
- showTree.bat
- xcheckMetadata.bat

```
project:    _p_bla
Short Name: bla
Title:      Demo project
Description:      Demonstration of
pISA projects path:     D:/_prj2
Local pISA-tree organisation:
pISA project creation date:   2019
pISA project creator:   AB
Project funding code:   *
Project coordinator:    *
Project partners: NIB, PTKRF
Project start date:     2017-04-01
Project end date:
Principal investigator: *
License:    CC BY 4.0
Sharing permission:     Private
Upload to FAIRDOMHub:   Yes
```

▸ To **seekr** features

▸ Skip tree representation

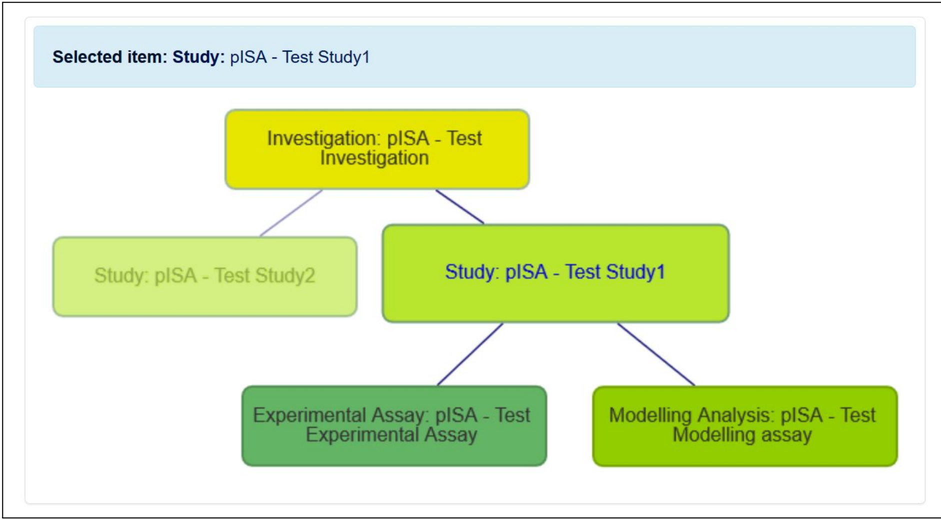
Motivation

- 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.
- Use metadata to **augment reproducibility of data analyses**.

Ideas

- Directory tree structures: LaTeX/`texmf`, R/packages, ...
- Personal directory trees for reproducible statistical reports (`R/Sweave/knitr/Rmarkdown/LaTeX` and `RStudio/WinEdt` projects).
- Structure research stages into **p**roject/**I**nvestigation/**S**tudy/**A**ssay layers, to be compliant with the ISA framework (we used to call the layers: Study/Experiment/Analysis).
- Low overhead, minimum installation of tools.

FAIRDOM / SEEK



Selected item: Study: pISA - Test Study1

Investigation: pISA - Test Investigation

Study: pISA - Test Study2

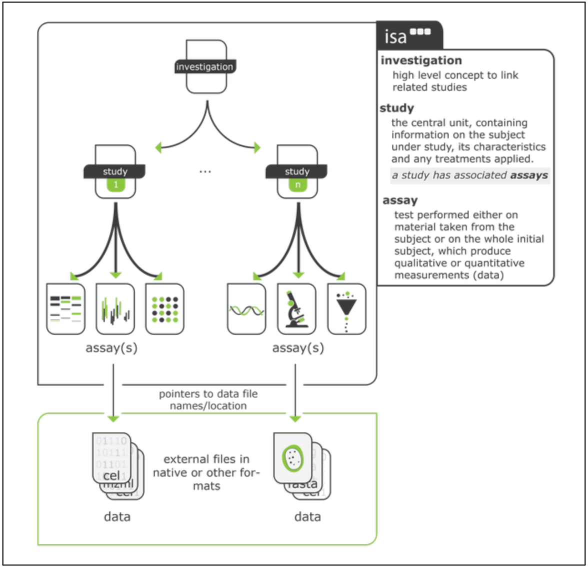
Study: pISA - Test Study1

Experimental Assay: pISA - Test Experimental Assay

Modelling Analysis: pISA - Test Modelling assay

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

ISA-tab model structure



investigation
high level concept to link related studies

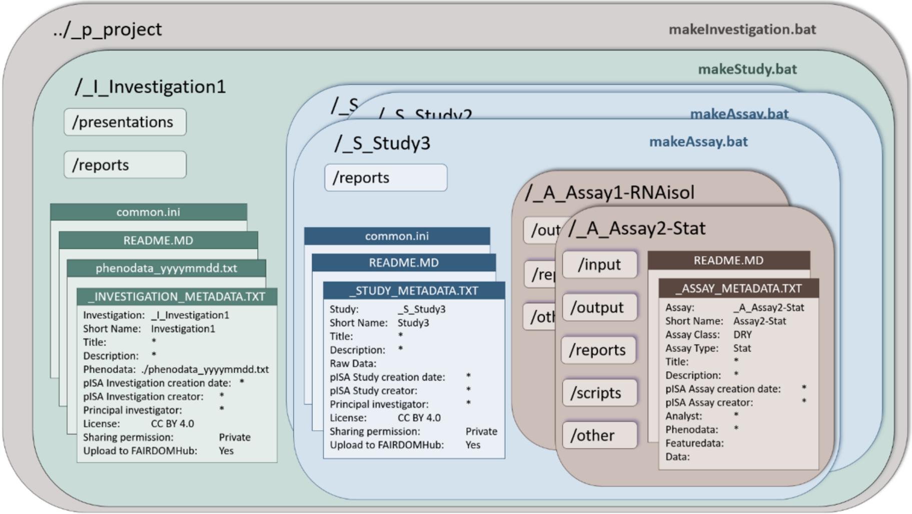
study
the central unit, containing information on the subject under study, its characteristics and any treatments applied.

a study has associated ***assays***

assay
test performed either on material taken from the subject or on the whole initial subject, which produce qualitative or quantitative measurements (data)

`http://isa-tools.org/format/specification/`

pISA-tree structure



Investigation: _I_Investigation1
Short Name: Investigation1
Title: *
Description: *
Phenodata: ./phenodata_yyyymmdd.txt
pISA Investigation creation date: *
pISA Investigation creator: *
Principal investigator: *
License: CC BY 4.0
Sharing permission: Private
Upload to FAIRDOMHub: Yes

Study: _S_Study3
Short Name: Study3
Title: *
Description: *
Raw Data:
pISA Study creation date: *
pISA Study creator: *
Principal investigator: *
License: CC BY 4.0
Sharing permission: Private
Upload to FAIRDOMHub: Yes

Assay: _A_Assay2-Stat
Short Name: Assay2-Stat
Assay Class: DRY
Assay Type: Stat
Title: *
Description: *
pISA Assay creation date: *
pISA Assay creator: *
Analyst: *
Phenodata: *
Featuredata:
Data:

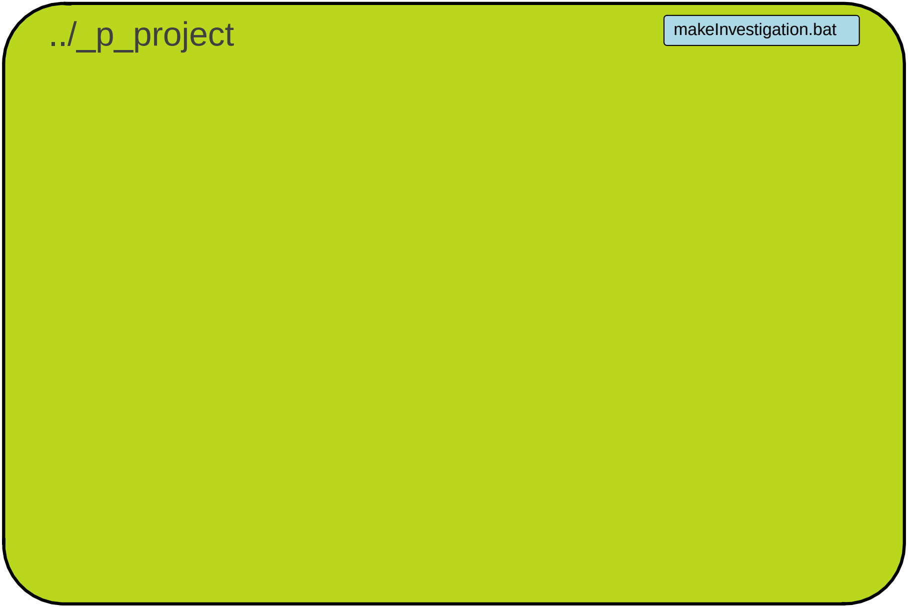
pISA-tree:

project directory tree
compliant with FAIRDOM/ISA framework

- File system based (Windows, extendable to Linux).
- Project directory tree with nested folders for **p**roject/**I**nvestigation/**S**tudy/**A**ssay layers.
- '**make...**' batch files are provided to create layers
  - `makeProject.bat`
  - `makeInvestigation.bat`
  - `makeStudy.bat`
  - `makeAssay.bat`
- Template 'metadata' files at each layer to **encourage metadata entry** (text files with Key/Value pairs).
- Allows **files in all ISA layers**.

▸ Skip tree representation

pISA-tree: project root directory `makeproject.bat`



../_p_project

makeInvestigation.bat

pISA-tree: add Investigation `makeInvestigation.bat`

../_p_project

makeInvestigation.bat

/_I_Investigation

makeStudy.bat

/presentations

/reports

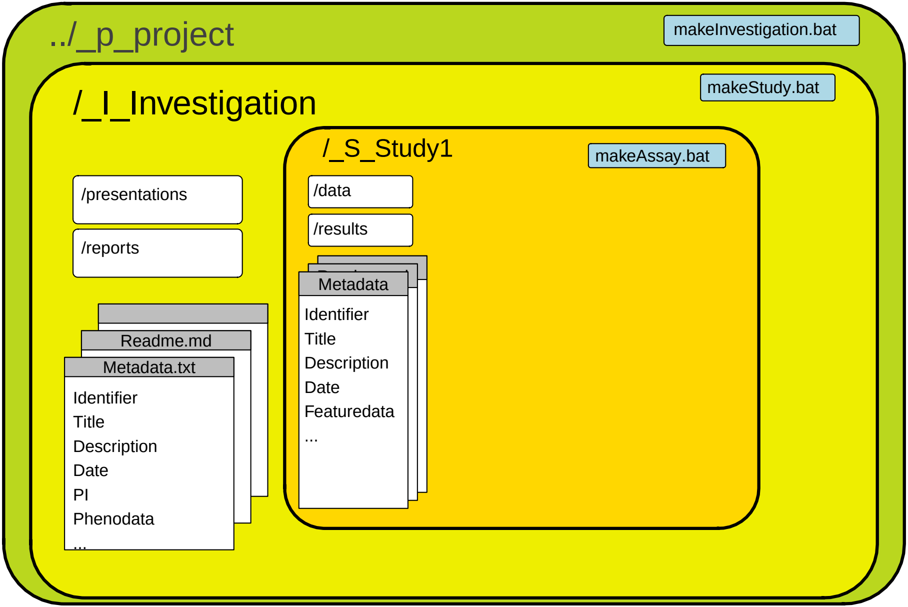
Readme.md

Metadata.txt

Identifier
Title
Description
Date
PI
Phenodata
...

pISA-tree: add a Study

`makeStudy.bat`



../_p_project

makeInvestigation.bat

/_I_Investigation

makeStudy.bat

/presentations

/reports

Readme.md

Metadata.txt

Identifier
Title
Description
Date
PI
Phenodata
...

/_S_Study1

makeAssay.bat

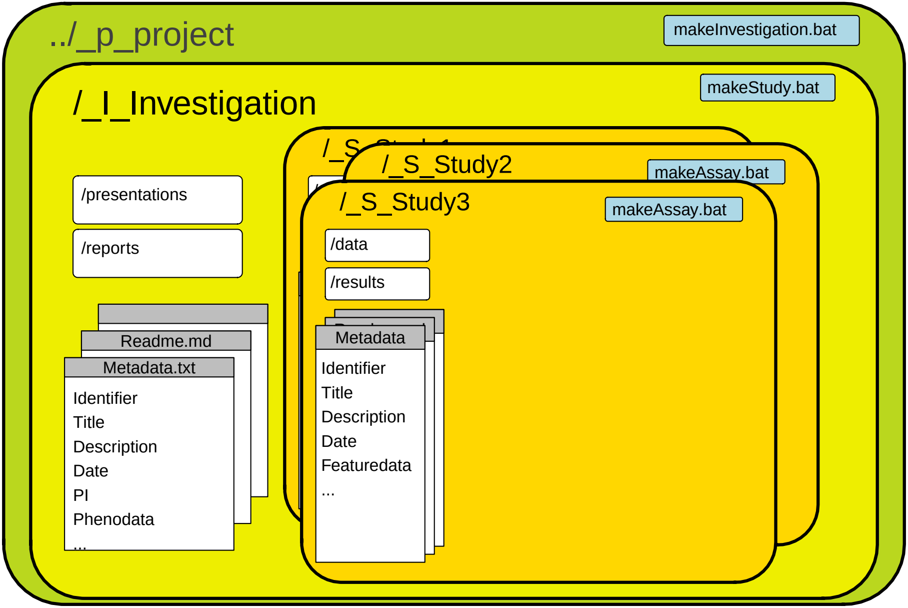
/data

/results

Metadata

Identifier
Title
Description
Date
Featuredata
...

pISA-tree: more Studies



../_p_project

makeInvestigation.bat

/_I_Investigation

makeStudy.bat

/presentations

/reports

Readme.md

Metadata.txt

Identifier
Title
Description
Date
PI
Phenodata
...

/_S

/_S_Study2

makeAssay.bat

/_S_Study3

makeAssay.bat

/data

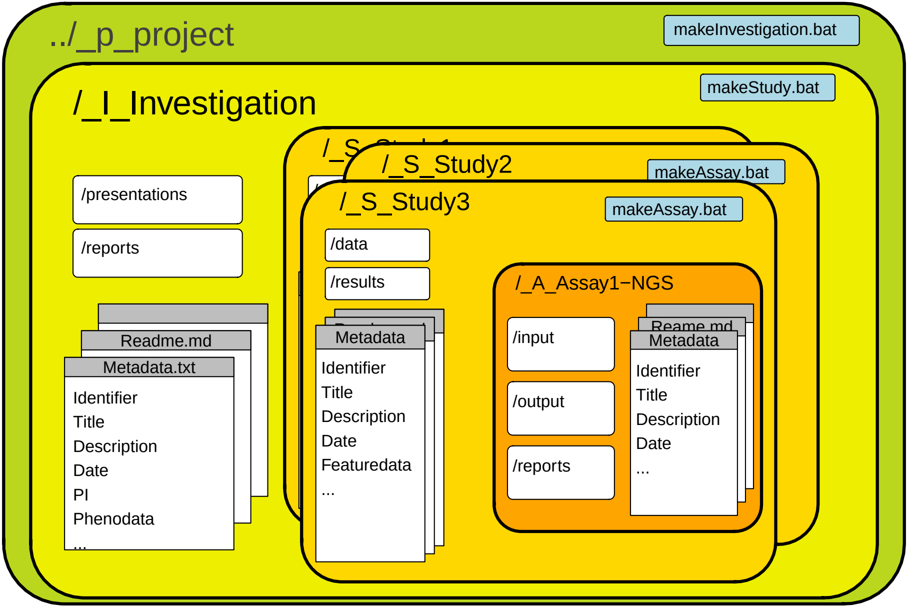
/results

Metadata

Identifier
Title
Description
Date
Featuredata
...

pISA-tree: add an Assay

`makeAssay.bat`



../_p_project

makeInvestigation.bat

/_I_Investigation

makeStudy.bat

/presentations

/reports

Readme.md

Metadata.txt

Identifier
Title
Description
Date
PI
Phenodata
...

/_S_Study2

makeAssay.bat

/_S_Study3

makeAssay.bat

/data

/results

Metadata

Identifier
Title
Description
Date
Featuredata
...

/_A_Assay1−NGS

/input

/output

/reports

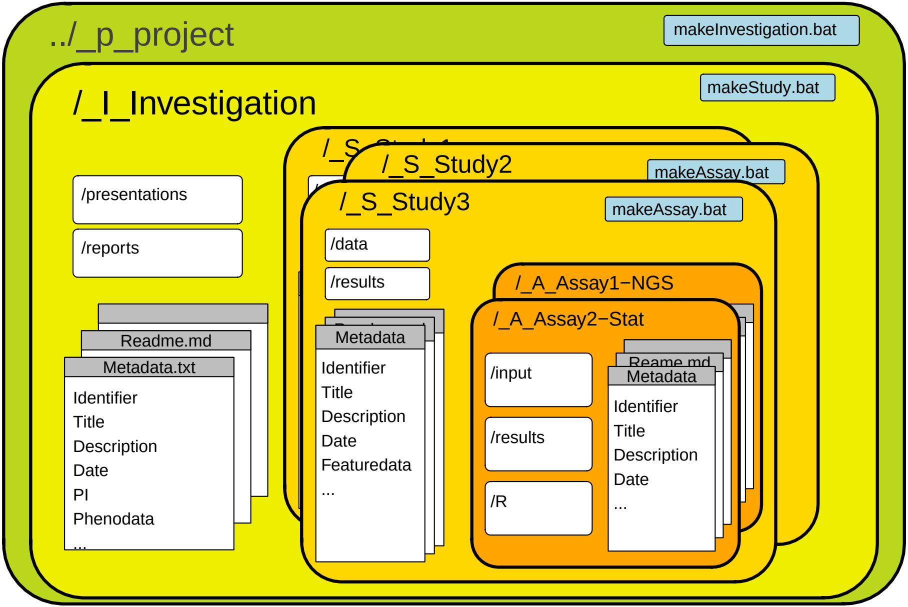
Reame.md

Metadata

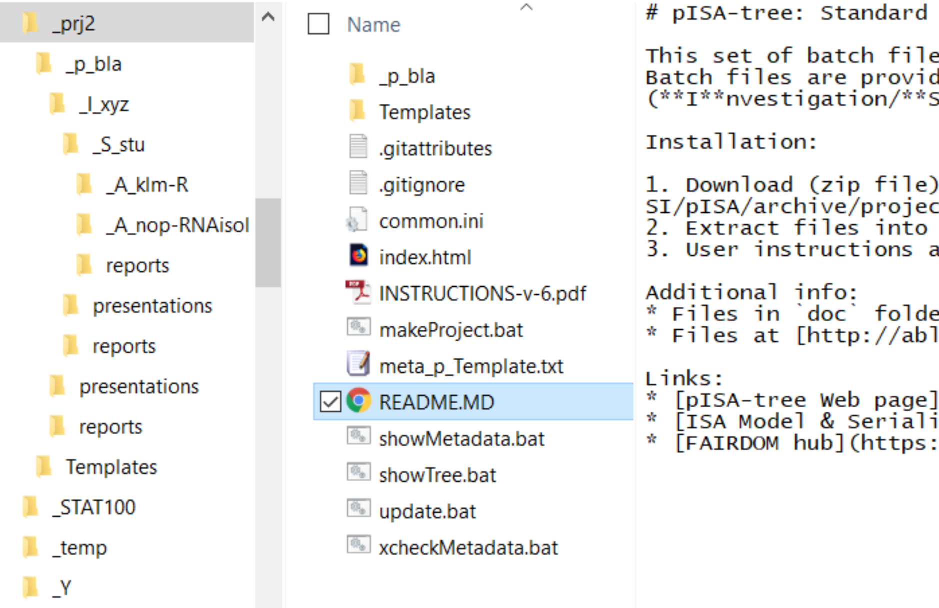
Identifier
Title
Description
Date
...

pISA-tree: another Assay

▸ Tree

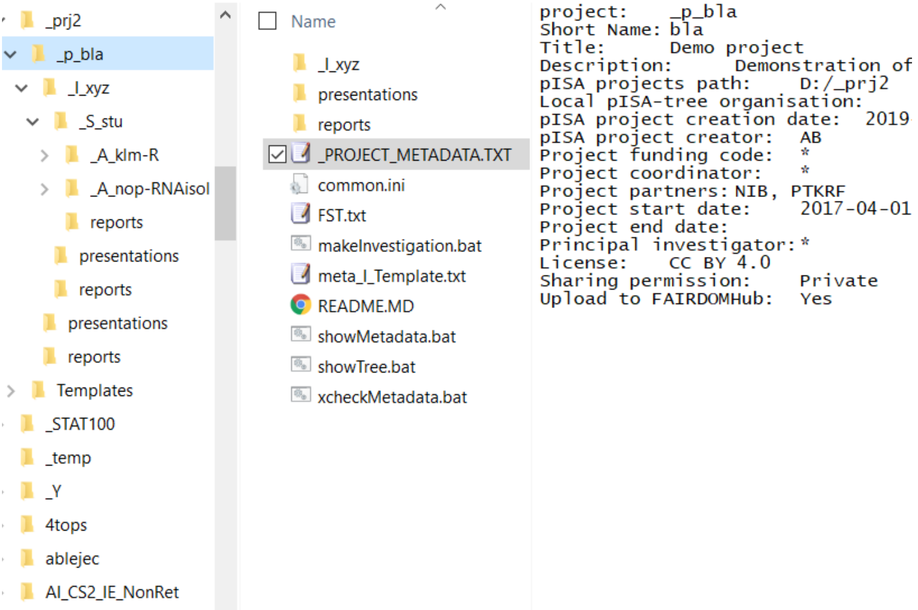


pISA-tree: the user view



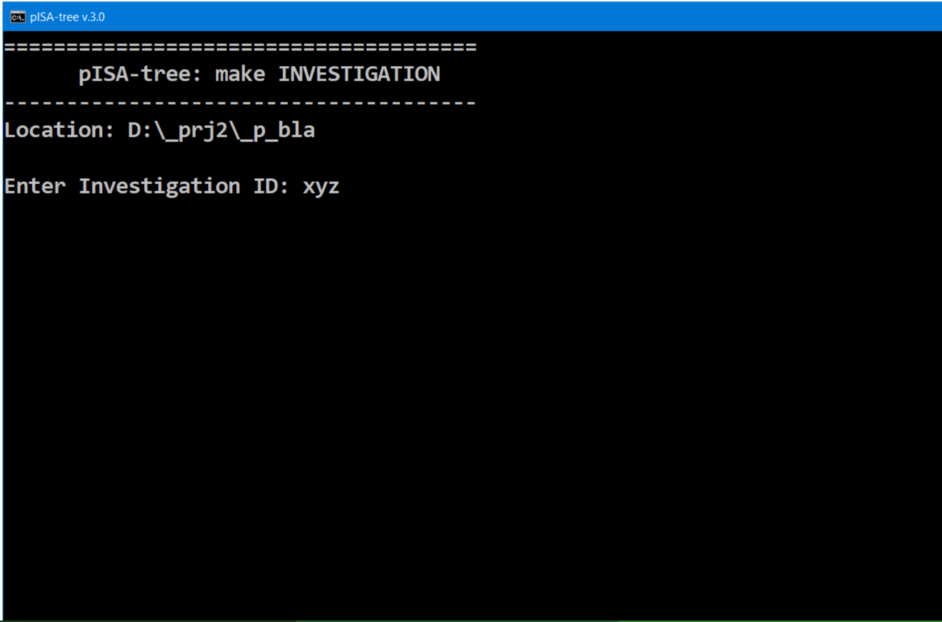
```
# pISA-tree: Standard

This set of batch file
Batch files are provid
(**I**nvestigation/**S

Installation:

1. Download (zip file)
SI/pISA/archive/projec
2. Extract files into
3. User instructions a

Additional info:
* Files in `doc` folde
* Files at [http://abl

Links:
* [pISA-tree Web page]
* [ISA Model & Seriali
* [FAIRDOM hub](https:
```

pISA tree - project folder and metadata



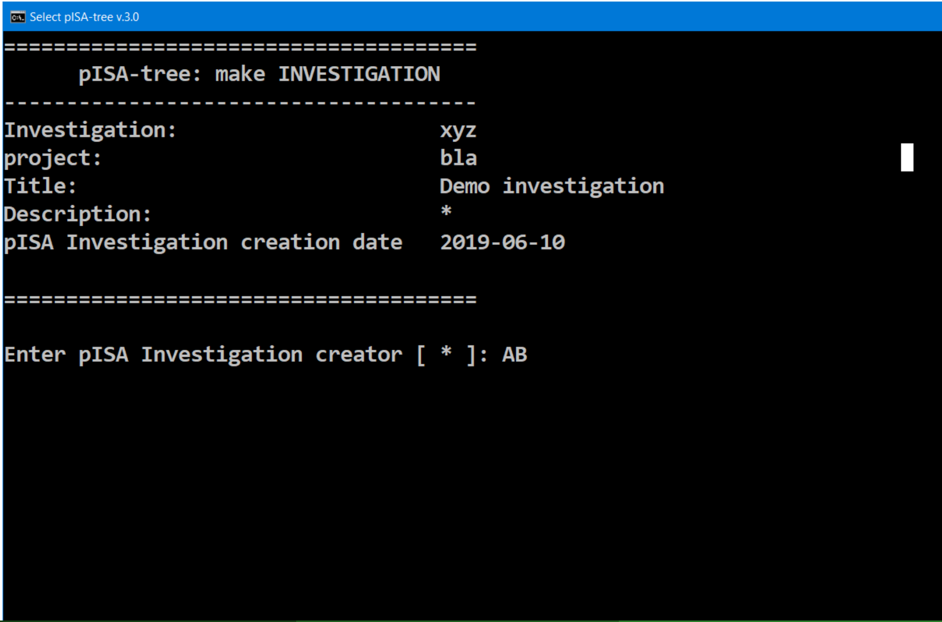
```
project:     _p_bla
Short Name: bla
Title:       Demo project
Description:      Demonstration of
pISA projects path:     D:/_prj2
Local pISA-tree organisation:
pISA project creation date:   2019
pISA project creator:    AB
Project funding code:    *
Project coordinator:     *
Project partners: NIB, PTKRF
Project start date:      2017-04-01
Project end date:
Principal investigator: *
License:     CC BY 4.0
Sharing permission:      Private
Upload to FAIRDOMHub:    Yes
```

pISA tree - `makeInvestigation.bat`



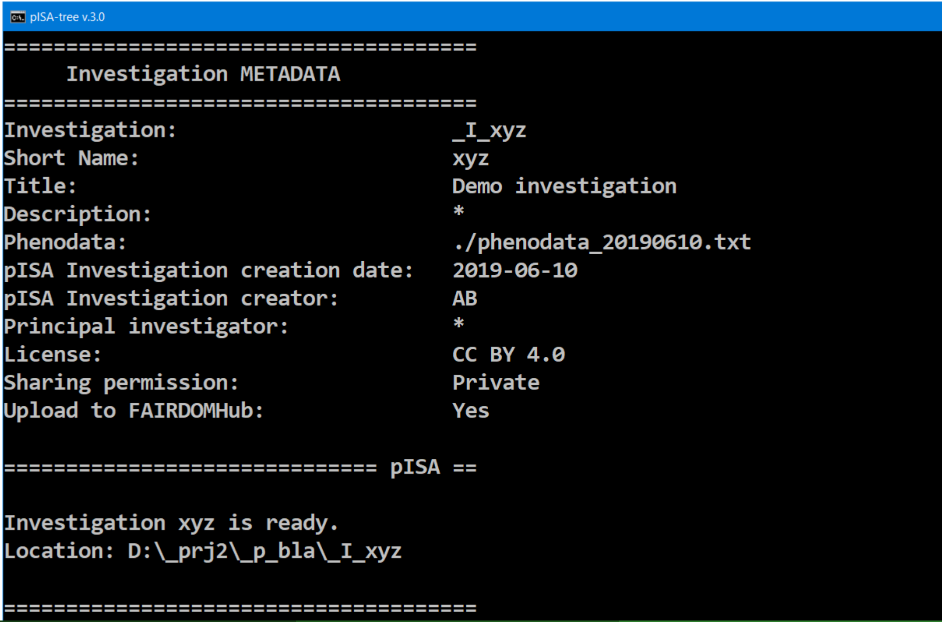
```
======================================
      pISA-tree: make INVESTIGATION
--------------------------------------
Location: D:\_prj2\_p_bla

Enter Investigation ID: xyz
```

pISA tree - `makeInvestigation.bat`



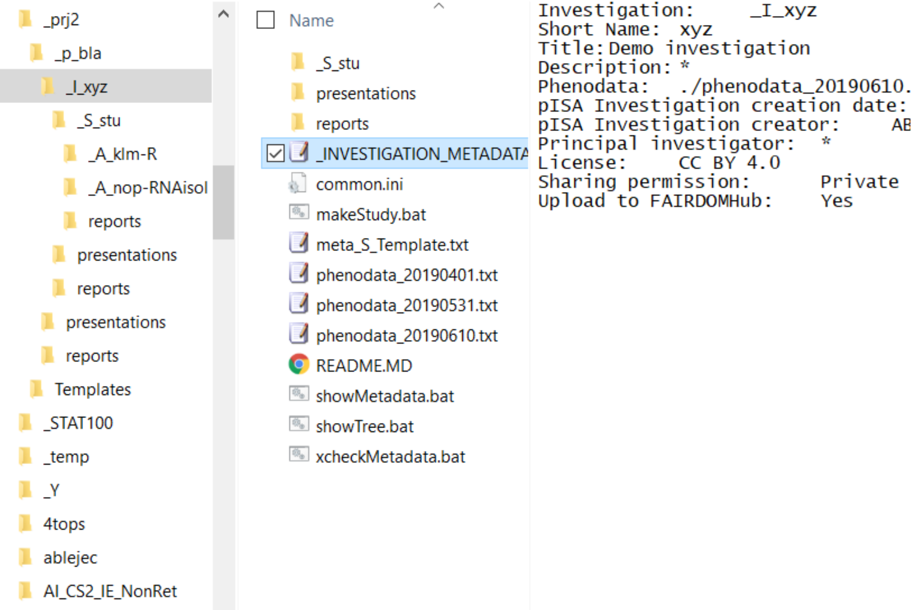
```
======================================
      pISA-tree: make INVESTIGATION
--------------------------------------
Investigation:                     xyz
project:                           bla
Title:                             Demo investigation
Description:                       *
pISA Investigation creation date   2019-06-10

======================================

Enter pISA Investigation creator [ * ]: AB
```

pISA tree - `makeInvestigation.bat`



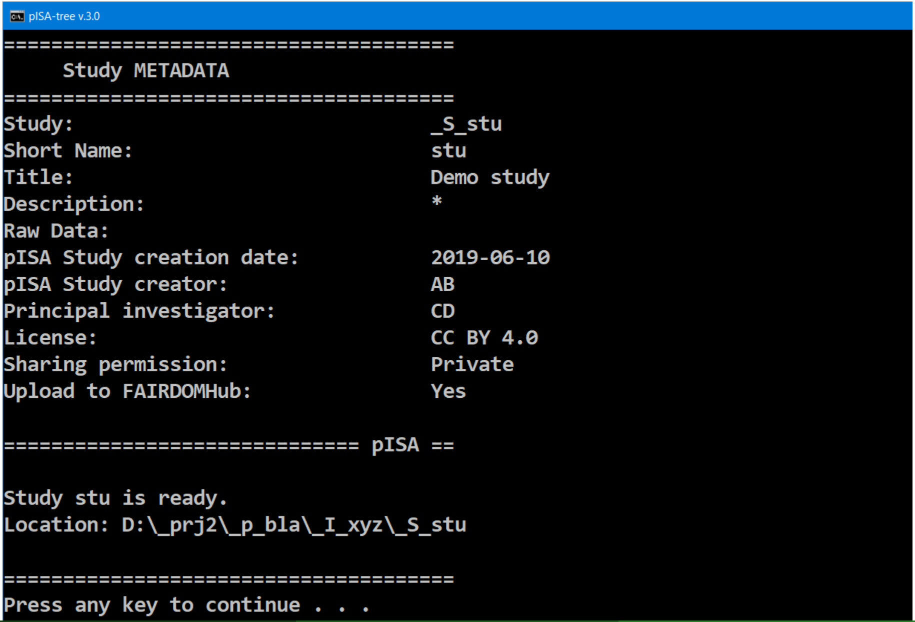
```
======================================
     Investigation METADATA
======================================
Investigation:                    _I_xyz
Short Name:                       xyz
Title:                            Demo investigation
Description:                      *
Phenodata:                        ./phenodata_20190610.txt
pISA Investigation creation date: 2019-06-10
pISA Investigation creator:       AB
Principal investigator:           *
License:                          CC BY 4.0
Sharing permission:               Private
Upload to FAIRDOMHub:             Yes

============================== pISA ==

Investigation xyz is ready.
Location: D:\_prj2\_p_bla\_I_xyz


======================================
```

pISA tree - Investigation folder and metadata



```
Investigation:      _I_xyz
Short Name:  xyz
Title:Demo investigation
Description: *
Phenodata:   ./phenodata_20190610.
pISA Investigation creation date:
pISA Investigation creator:      AB
Principal investigator:   *
License:     CC BY 4.0
Sharing permission:       Private
Upload to FAIRDOMHub:     Yes
```

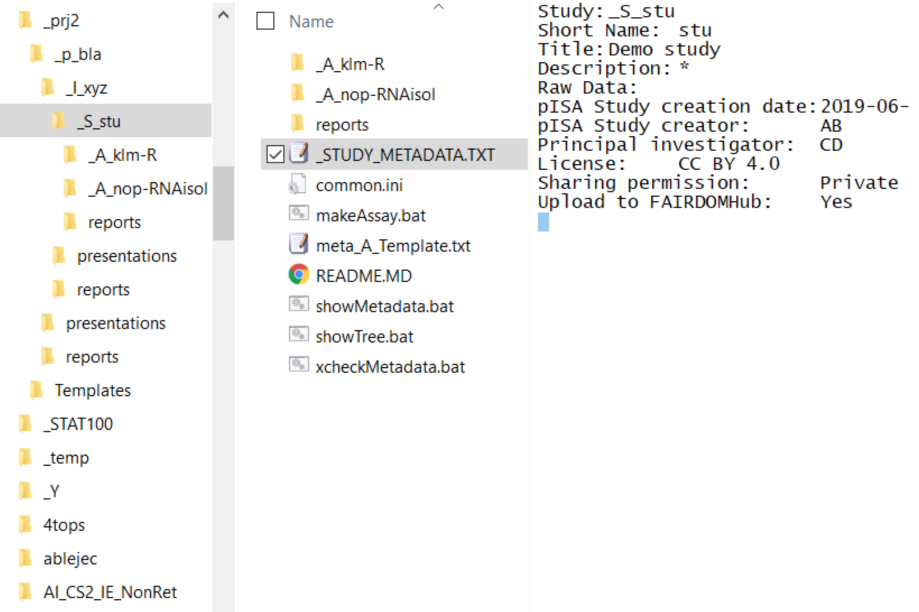
pISA tree - `makeStudy.bat`



pISA-tree v.3.0

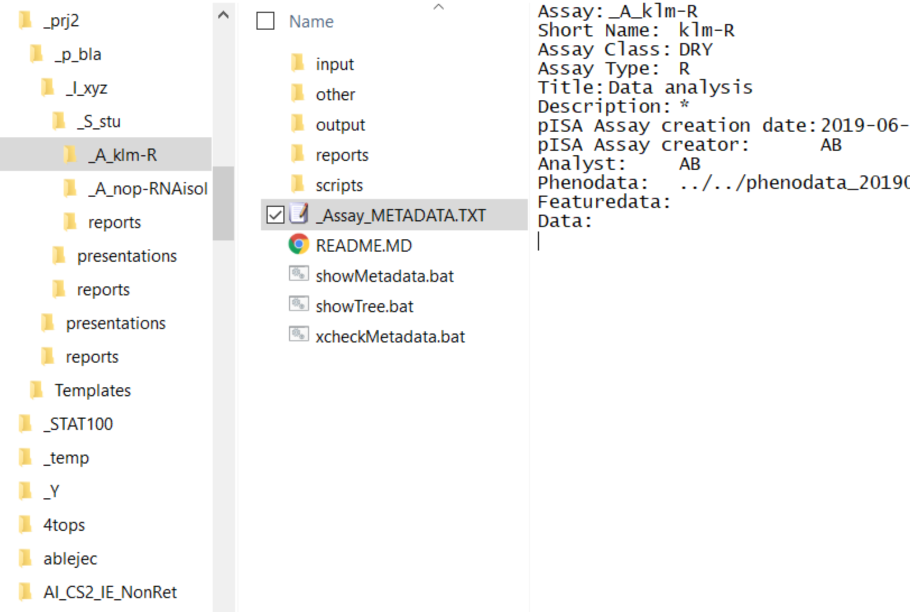
```
======================================
     Study METADATA
======================================
Study:                                _S_stu
Short Name:                           stu
Title:                                Demo study
Description:                          *
Raw Data:
pISA Study creation date:             2019-06-10
pISA Study creator:                   AB
Principal investigator:               CD
License:                              CC BY 4.0
Sharing permission:                   Private
Upload to FAIRDOMHub:                 Yes

============================== pISA ==

Study stu is ready.
Location: D:\_prj2\_p_bla\_I_xyz\_S_stu

======================================
Press any key to continue . . .
```

pISA tree - Study folder and metadata



```
Study:_S_stu
Short Name:  stu
Title:Demo study
Description: *
Raw Data:
pISA Study creation date:2019-06-
pISA Study creator:       AB
Principal investigator:   CD
License:     CC BY 4.0
Sharing permission:       Private
Upload to FAIRDOMHub:     Yes
```

pISA tree - Assay folder (DRY) and metadata

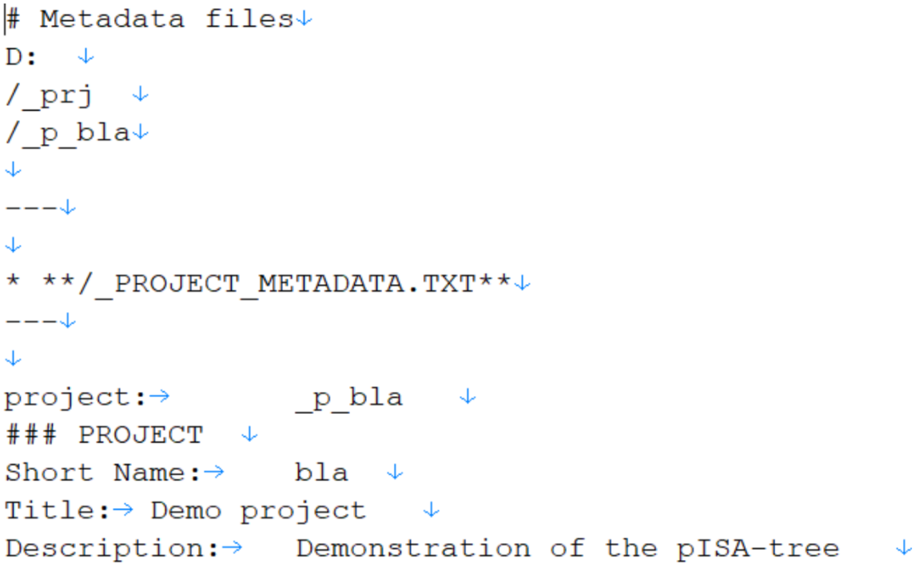


```
Assay:_A_klm-R
Short Name:  klm-R
Assay Class: DRY
Assay Type:  R
Title:Data analysis
Description: *
pISA Assay creation date:2019-06-
pISA Assay creator:       AB
Analyst:     AB
Phenodata:   ../../phenodata_2019
Featuredata:
Data:
```

Auxiliary batch files

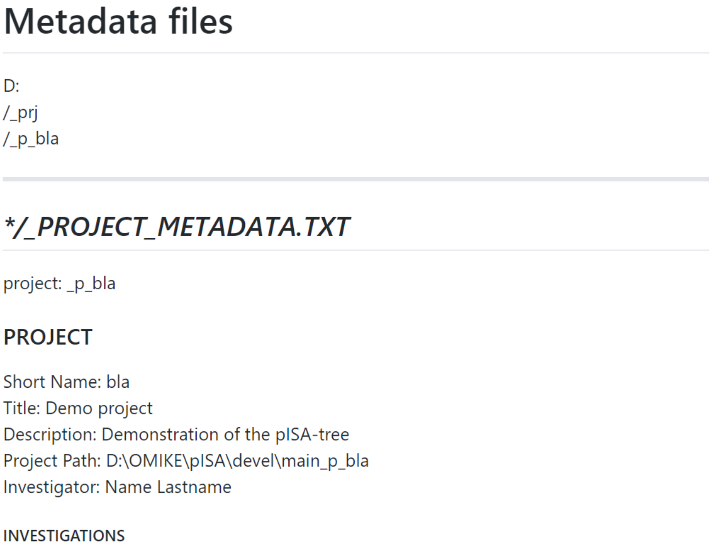
- `showMetadata.bat`
- `xcheckMetadata.bat`
- `showTree.bat`

`showMetadata.bat`
Encourage use of markdown



```
# Metadata files↓
D:   ↓
/_prj  ↓
/_p_bla↓
↓
---↓
↓
* **/_PROJECT_METADATA.TXT**↓
---↓
↓
project:→        _p_bla   ↓
### PROJECT  ↓
Short Name:→    bla  ↓
Title:→ Demo project   ↓
Description:→   Demonstration of the pISA-tree   ↓
```

showMetadata.bat
Encourage use of markdown



Metadata files

D:
/_prj
/_p_bla

*/_PROJECT_METADATA.TXT

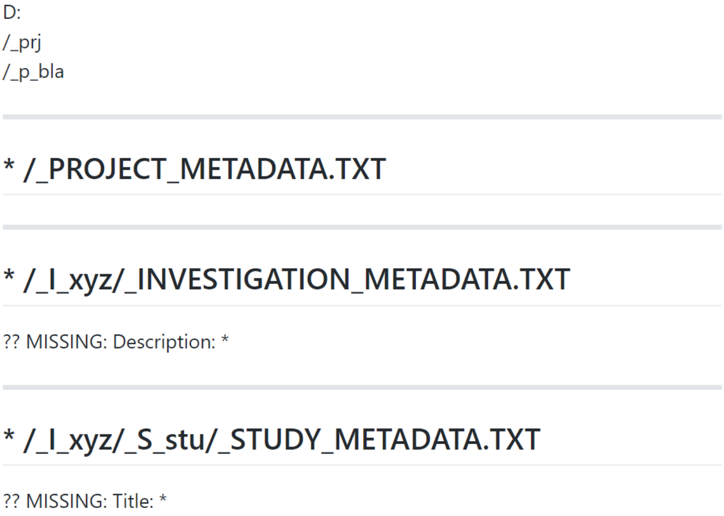
project: _p_bla

PROJECT

Short Name: bla
Title: Demo project
Description: Demonstration of the pISA-tree
Project Path: D:\OMIKE\pISA\devel\main_p_bla
Investigator: Name Lastname

INVESTIGATIONS

`xcheckMetadata.bat`
Shows what is missing



D:
/_prj
/_p_bla

* /_PROJECT_METADATA.TXT

* /_I_xyz/_INVESTIGATION_METADATA.TXT

?? MISSING: Description: *

* /_I_xyz/_S_stu/_STUDY_METADATA.TXT

?? MISSING: Title: *

showTree.bat

Show the directory tree

```
D:\_prj2\_p_bla\_I_xyz\_S_stu\_A_nop-RNAisol
Folder PATH listing for volume DATA
Volume serial number is 5859-00FF
D:.
|   Analytes.txt
|   d.tmp
|   README.MD
|   showMetadata.bat
|   showTree.bat
|   t.tmp
|   xcheckMetadata.bat
|   _Assay_METADATA.TXT
|
+---other
|       README.MD
|
+---output
|   |   README.MD
|   |
|   \---raw
|           README.MD
|
```

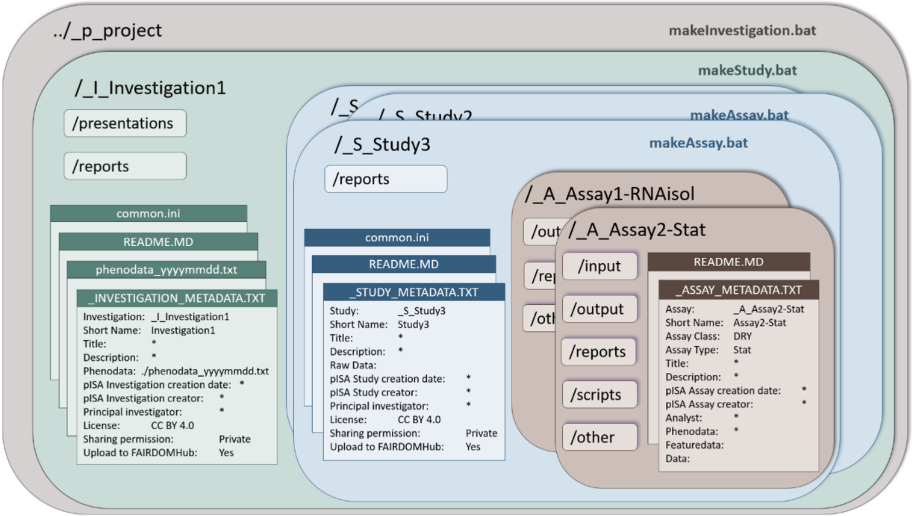
pISA-tree features

- A system with **low overhead**.
- **Metadata file** for each layer.
- Several **Assay classes** (Wet / Dry / ...).
- Several **Assay types** (RNAisol, RT, Stat, R, ...).
- Menu-driven metadata entry (to **reinforce consistency**).
- Follows recommendations of **ISA/SEEK paradigm**.
- Allows **files in all ISA levels**.
- Availability: `https://github.com/NIB-SI/pISA-tree`.

Connection with other solutions

- FAIRDOM
  https://fairdomhub.org/
- SEEK
  https://seek.sysmo-db.org/
- ISA-tab
  https://isa-tools.github.io/
  function `createISAtab()` in R package **pisar**
- FAIRDOMhub API
  http://docs.seek4science.org/tech/api/
  R package **seekr**
  https://github.com/NIB-SI/seekr
- reproducible statistical analyses in R
  R package **pisar**
  https://github.com/NIB-SI/pisar

pISA-tree structure



common.ini

README.MD

phenodata_yyyymmdd.txt

_INVESTIGATION_METADATA.TXT

Investigation: _I_Investigation1
Short Name: Investigation1
Title: *
Description: *
Phenodata: ./phenodata_yyyymmdd.txt
pISA Investigation creation date: *
pISA Investigation creator: *
Principal investigator: *
License: CC BY 4.0
Sharing permission: Private
Upload to FAIRDOMHub: Yes

common.ini

README.MD

_STUDY_METADATA.TXT

Study: _S_Study3
Short Name: Study3
Title: *
Description: *
Raw Data:
pISA Study creation date: *
pISA Study creator: *
Principal investigator: *
License: CC BY 4.0
Sharing permission: Private
Upload to FAIRDOMHub: Yes

README.MD

_ASSAY_METADATA.TXT

Assay: _A_Assay2-Stat
Short Name: Assay2-Stat
Assay Class: DRY
Assay Type: Stat
Title: *
Description: *
pISA Assay creation date: *
pISA Assay creator: *
Analyst: *
Phenodata: *
Featuredata:
Data:

From pISA-tree to FAIRDOMHub

- They share **ISA structure**.
- pISA-tree is practical at the (inter)**institutional level**.
- FAIRDOMHub provides **FAIR-ness and wider visibility**.
- FAIRDOM-SEEK provides browsing and manual/interactive creation, access of items, including the file upload.
- FAIRDOM-SEEK includes a JSON API that allows computer driven searching, listing, reading, updating, creating and upload of items in SEEK.

Why API and R package **seekr**?

- Researchers produce large number of files.
- Manual/interactive upload of files and many ISA layers can become tedious.
- To support computer driven upload we developed R package **seekr** based on FAIRDOM-SEEK API functionalities.

R package **seekr**

Functions in **seekr** support:

- Searching: `skFindId()`, `skFindTitle()`, `skExists()`
- Listing: `skList()`, `skListp()`, `skRelated()`
- Reading: `skGet()`, `skRead()`, `skParse()`, `skContent()`
- Creating: `skCreate()`, `skCheckAndCreate()`,
- Deleting: `skDelete()`
- Uploading: `skUpload()`, `skUplodFile()`
- **Bulk uploading**: `skFilesToUpload()`, `skUploadFiles()`

▸ To upload details

Packages and user credentials

```
> library(httr)
> library(jsonlite)
> library(seekr)
```

User credentials

```
> options(.sk$main)
> skOptions()
Hidden: sk.pwd
$sk.myid
[1] 808

$sk.url
[1] "https://fairdomhub.org"

$sk.usr
[1] "ablejec"
```

Searching

Finding ids and titles

```
> skFindId("people","Andrej Blejec")
             id            type           title 
          "808"        "people" "Andrej Blejec" 
> skFindId("projects", "pISA-tree")
         id        type       title 
       "57"  "projects" "pISA-tree" 
> skFindId("programmes", "NIBSys")
          id         type        title 
        "17" "programmes"     "NIBSys" 
> skFindTitle("programmes", 17)
          id         type        title 
        "17" "programmes"     "NIBSys" 
```

▸ To upload details

Listing

```
> # Short list of projects
> r <- skList("projects")
> r <- skContent(r)
> head(sort(sapply(r, function(x) x$attributes$title)),10)
 [1] "_p_RNAinVAL"
 [2] "_p_stRT"
 [3] "Acetyl-coa production in Pseudomonas putida"
 [4] "ADAPT - Accelerated Development of multiple-stress tolerAnt
 [5] "AGENSI"
 [6] "Agro-ecological modelling"
 [7] "An overview of systematic reviews of treatment of polycysti
 [8] "Aneurinibacillus thermoaerophilus CCM 8960 metabolic pathwa
 [9] "AquaHealth (ERA-BlueBio)"
[10] "Are microRNAs key mediators of cartilage destruction in oste
```

▸ To upload details

Reading

```
> id <- skFindId("projects","_p_stRT")["id"]
> (r <- skRead("projects",id))
Status: 200
Object: https://fairdomhub.org/projects/161
Path  : /projects/161
Title : _p_stRT
> d <- skContent(r)$attributes$description
```

Metadata

```
> cat(d)
project:        _p_stRT
Short Name:        stRT
Title:        Solanum tuberosum Reference Transcriptomes
Description:        Cultivar-specific transcriptome and pan-transc
pISA projects path:        ../pISA-Projects
Local pISA-tree organisation:        National Institute of Biolog
```

▸ To upload details

Creating and deleting

- `skCreate()`

  Create a SEEK item.
- `skCheckAndCreate()`

  Check if ISA layer exists and create it if not.
- `skDelete()`

  Delete item.

Uploading

- `skFilesToUpload()`
  Prepare a list of files to upload.
- `skUploadFiles()`
  Bulk upload of files in the file list.

  - `skCheckAndCreate()`
    Check if ISA layer exists and create it if not.
  - `skUplodFile()`
    Create an item of type 'documents' or 'data_files' and upload file content as a blob.

Step 1: `skFilesToUpload()`: prepare the list of files

- Create a list of files in a folder and subfolders.
  Typically: files in an investigation with studies and assays.
- Function `skFilesToUpload()` filters the list of files by:
  - Excluding ISA layers with the key
    `Upload to FAIRDOMHub:  No`
    in metadata.
  - Excluding files with patterns in the file seekignore.txt.

```
> seekignore[1:7]
[1] "# File patterns to be ignored by SEEK API upload"
[2] "#"
[3] "DoNotUpload/"
[4] "other/"
[5] "*.bat"
[6] "_template.txt"
[7] "common.ini"
```

Step 1: Example

```
> filesToUpload <- skFilesToUpload(.iroot)
```

```
Ignoring pISA layer: _S_NoUpload/_A_TestUpload
Ignoring pISA layer: _S_NoUpload
Ignored pISA layers: 2 / 7
Ignored README.MD files: 22 / 25
Ignoring 0 files [ DoNotUpload/  ]
Ignoring 7 files [ other/  ]
Ignoring 21 files [ *\.bat$  ]
Ignoring 3 files [ _template\.txt$  ]
Ignoring 3 files [ common\.ini$  ]
Ignoring 0 files [ *\.zip$  ]
Ignoring 0 files [ *\.tmp$  ]
...
Ignoring 1 files [ _S_Show/reports/README.MD  ]
Files to upload: 14 / 110
```

Step 1: Files to upload

```
> head(filesToUpload,20)
 [1] "_INVESTIGATION_METADATA.TXT"
 [2] "phenodata_20191015.txt"
 [3] "README.MD"
 [4] "seekignore.txt"
 [5] "_S_Show/_STUDY_METADATA.TXT"
 [6] "_S_Show/_A_Private-R/_Assay_METADATA.TXT"
 [7] "_S_Show/_A_Work-R/_Assay_METADATA.TXT"
 [8] "_S_Show/_A_Work-R/UploadSigma.txt"
 [9] "_S_Show/_A_Private-R/input/some-data.txt"
[10] "_S_Show/_A_Private-R/reports/README.MD"
[11] "_S_Show/_A_Work-R/input/bla.txt"
[12] "_S_Show/_A_Work-R/input/filename with blanks.txt"
[13] "_S_Show/_A_Work-R/input/some-data.txt"
[14] "_S_Show/_A_Work-R/reports/README.MD"
[15] "_S_Show/_A_Work-R/reports/SIGMA_vs_A_plot.pdf"
```

Step 2: skUploadFiles

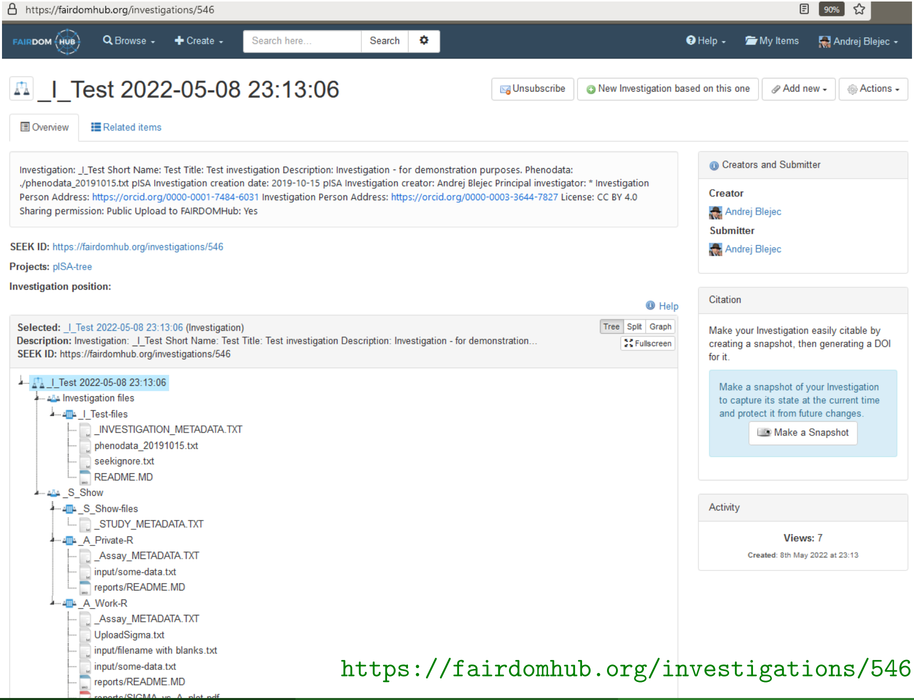
```
> skUploadFiles(
+     filesToUpload,
+     reportFile="UploadReport.txt")
```

Step 2: skUploadFiles - progress report

```
|..>______________________________________________| 1/14
_INVESTIGATION_METADATA.TXT
Status code: 200
... Uploading: _INVESTIGATION_METADATA.TXT
|......>__________________________________________| 2/14
phenodata_20191015.txt
... Uploading: phenodata_20191015.txt
|.........>_______________________________________| 3/14
```

```
|....................>____________________________| 6/14
_S_Show/_A_Private-R/_Assay_METADATA.TXT
Create: assays '_A_Private-R' (no_access)
Status: 200
Object: https://fairdomhub.org/assays/1814
Path  : /assays/1814
Title : _A_Private-R
... Uploading: _S_Show/_A_Private-R/_Assay_METADATA.TXT
```

```
|.............................................>___| 13/14
_S_Show/_A_Work-R/reports/README.MD
... Uploading: _S_Show/_A_Work-R/reports/README.MD
|.................................................>| 14/14
_S_Show/_A_Work-R/reports/SIGMA_vs_A_plot.pdf
... Uploading: _S_Show/_A_Work-R/reports/SIGMA_vs_A_plot.pdf
```



_I_Test 2022-05-08 23:13:06

Investigation: _I_Test Short Name: Test Title: Test investigation Description: Investigation - for demonstration purposes. Phenodata: ./phenodata_20191015.txt pISA Investigation creation date: 2019-10-15 pISA Investigation creator: Andrej Blejec Principal investigator: * Investigation Person Address: https://orcid.org/0000-0001-7484-6031 Investigation Person Address: https://orcid.org/0000-0003-3644-7827 License: CC BY 4.0 Sharing permission: Public Upload to FAIRDOMHub: Yes

SEEK ID: https://fairdomhub.org/investigations/546

Projects: pISA-tree

Investigation position:

Selected: _I_Test 2022-05-08 23:13:06 (Investigation)
Description: Investigation: _I_Test Short Name: Test Title: Test investigation Description: Investigation - for demonstration...
SEEK ID: https://fairdomhub.org/investigations/546

- _I_Test 2022-05-08 23:13:06
  - Investigation files
    - _I_Test-files
      - _INVESTIGATION_METADATA.TXT
      - phenodata_20191015.txt
      - seekignore.txt
      - README.MD
  - _S_Show
    - _S_Show-files
      - _STUDY_METADATA.TXT
    - _A_Private-R
      - _Assay_METADATA.TXT
      - input/some-data.txt
      - reports/README.MD
    - _A_Work-R
      - _Assay_METADATA.TXT
      - UploadSigma.txt
      - input/filename with blanks.txt
      - input/some-data.txt
      - reports/README.MD

Creators and Submitter

Creator
Andrej Blejec
Submitter
Andrej Blejec

Citation

Make your Investigation easily citable by creating a snapshot, then generating a DOI for it.

Make a snapshot of your Investigation to capture its state at the current time and protect it from future changes.

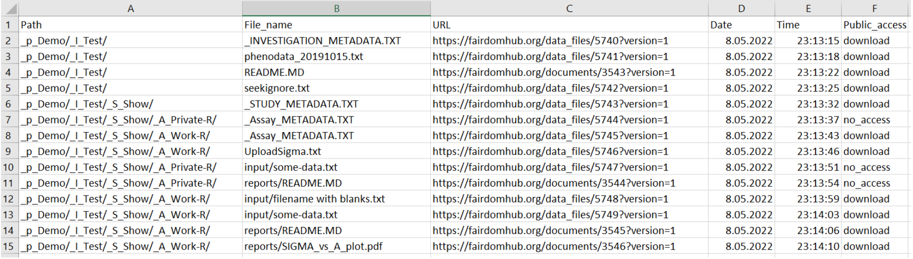
Activity

Views: 7

Created: 8th May 2022 at 23:13

`https://fairdomhub.org/investigations/546`

Step 2: Upload report



	A	B	C	D	E	F
1	Path	File_name	URL	Date	Time	Public_access
2	_p_Demo/_I_Test/	_INVESTIGATION_METADATA.TXT	https://fairdomhub.org/data_files/5740?version=1	8.05.2022	23:13:15	download
3	_p_Demo/_I_Test/	phenodata_20191015.txt	https://fairdomhub.org/data_files/5741?version=1	8.05.2022	23:13:18	download
4	_p_Demo/_I_Test/	README.MD	https://fairdomhub.org/documents/3543?version=1	8.05.2022	23:13:22	download
5	_p_Demo/_I_Test/	seekignore.txt	https://fairdomhub.org/data_files/5742?version=1	8.05.2022	23:13:25	download
6	_p_Demo/_I_Test/_S_Show/	_STUDY_METADATA.TXT	https://fairdomhub.org/data_files/5743?version=1	8.05.2022	23:13:32	download
7	_p_Demo/_I_Test/_S_Show/_A_Private-R/	_Assay_METADATA.TXT	https://fairdomhub.org/data_files/5744?version=1	8.05.2022	23:13:37	no_access
8	_p_Demo/_I_Test/_S_Show/_A_Work-R/	_Assay_METADATA.TXT	https://fairdomhub.org/data_files/5745?version=1	8.05.2022	23:13:43	download
9	_p_Demo/_I_Test/_S_Show/_A_Work-R/	UploadSigma.txt	https://fairdomhub.org/data_files/5746?version=1	8.05.2022	23:13:46	download
10	_p_Demo/_I_Test/_S_Show/_A_Private-R/	input/some-data.txt	https://fairdomhub.org/data_files/5747?version=1	8.05.2022	23:13:51	no_access
11	_p_Demo/_I_Test/_S_Show/_A_Private-R/	reports/README.MD	https://fairdomhub.org/documents/3544?version=1	8.05.2022	23:13:54	no_access
12	_p_Demo/_I_Test/_S_Show/_A_Work-R/	input/filename with blanks.txt	https://fairdomhub.org/data_files/5748?version=1	8.05.2022	23:13:59	download
13	_p_Demo/_I_Test/_S_Show/_A_Work-R/	input/some-data.txt	https://fairdomhub.org/data_files/5749?version=1	8.05.2022	23:14:03	download
14	_p_Demo/_I_Test/_S_Show/_A_Work-R/	reports/README.MD	https://fairdomhub.org/documents/3545?version=1	8.05.2022	23:14:06	download
15	_p_Demo/_I_Test/_S_Show/_A_Work-R/	reports/SIGMA_vs_A_plot.pdf	https://fairdomhub.org/documents/3546?version=1	8.05.2022	23:14:10	download

`https://fairdomhub.org/data_files/5740?version=1`

`https://fairdomhub.org/documents/3546?version=1`

Step 2: Log file

```
2022-05-08 22:45:07 Uploading file:  _p_Demo/_I_Test/_S_Show/_A_Work-R/  |  reports
2022-05-08 22:45:07 skCreate https://fairdomhub.org/documents
2022-05-08 22:45:09 200 OK 2.11 https://fairdomhub.org/documents/3542?version=1
2022-05-08 22:45:09 Upload file: _S_Show/_A_Work-R/reports/SIGMA_vs_A_plot.pdf
2022-05-08 22:45:09 skUpload https://fairdomhub.org/documents/3542/content_blobs/15
2022-05-08 22:45:11 200 OK 1.23 ../../../_S_Show/_A_Work-R/reports/SIGMA_vs_A_plot.
```

Issues (and features)

- Windows and R issue with long file names.
- Possible timeouts for huge files (rare, speed ∼5 MB/s).
- ISA structure permits files only in Assay layer
  pISA-tree has files also in other ISA layers
  seekr has a workaround for upload to FAIRDOMHub.
- Update item /`skUpdate()`/: not permitted, credentials issue?.
- How to upload blob as a new version?

Conclusion

- R package **seekr** (based on SEEK API) can be used for bulk upload to FAIRDOMHub.
- We have uploaded two large projects with several hundred files, another two in preparation.
- Software availability:
  - Package **seekr** `https://github.com/NIB-SI/seekr`
  - Package **pisar** `https://github.com/NIB-SI/pisar`
  - pISA-tree data management `https://github.com/NIB-SI/pISA-tree`

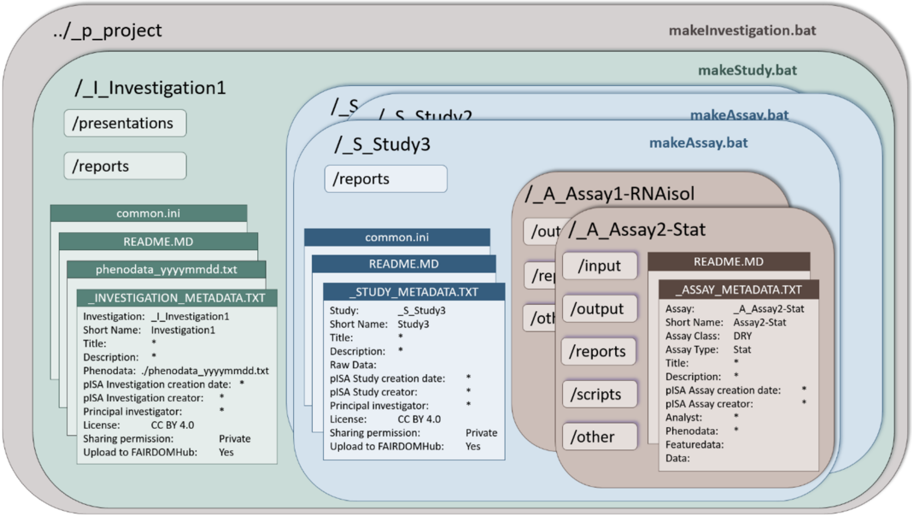
NIB Dept. of Biotechnology and Systems Biology

Bioinformatics group

Kristina Gruden
Špela Baebler
Andrej Blejec
Anna Coll
Marko Petek
Živa Ramšak
Maja Zagorščak

Thank you for your attention

pISA-tree structure



../_p_project

makeInvestigation.bat

/_I_Investigation1

makeStudy.bat

/presentations

/reports

common.ini

README.MD

phenodata_yyyymmdd.txt

_INVESTIGATION_METADATA.TXT

Investigation: _I_Investigation1
Short Name: Investigation1
Title: *
Description: *
Phenodata: ./phenodata_yyyymmdd.txt
pISA Investigation creation date: *
pISA Investigation creator: *
Principal investigator: *
License: CC BY 4.0
Sharing permission: Private
Upload to FAIRDOMHub: Yes

/_S

/_S_Study2

makeAssay.bat

/_S_Study3

makeAssay.bat

/reports

common.ini

README.MD

_STUDY_METADATA.TXT

Study: _S_Study3
Short Name: Study3
Title: *
Description: *
Raw Data:
pISA Study creation date: *
pISA Study creator: *
Principal investigator: *
License: CC BY 4.0
Sharing permission: Private
Upload to FAIRDOMHub: Yes

/_A_Assay1-RNAisol

/out

/re

/oth

/_A_Assay2-Stat

/input

/output

/reports

/scripts

/other

README.MD

_ASSAY_METADATA.TXT

Assay: _A_Assay2-Stat
Short Name: Assay2-Stat
Assay Class: DRY
Assay Type: Stat
Title: *
Description: *
pISA Assay creation date: *
pISA Assay creator: *
Analyst: *
Phenodata: *
Featuredata:
Data: