

What are investigations?



The smallest amount of samples that still form a complete story

Table of Contents

- [The smallest amount of samples that still form a complete story](#)
 - [Background](#)
 - [FAIR data](#)
 - [ISA data model](#)
 - [PISA - the DataHub implementation](#)
 - [PISA definitions](#)
 - [Project](#)
 - [Investigation](#)
 - [Sample](#)
 - [Assay](#)
 - [Mapping PISA to iRODS](#)

Background

To answer this question, you need some information about our metadata model (the PISA-model) first.

FAIR data

At DataHub we believe in FAIR data. One of the aspects to accomplish FAIRness is to accompany datasets with rich, machine-readable and semantic metadata. Annotating data with proper metadata is often a cumbersome and boring process. As such, we have developed a web portal that uses intuitive forms to enter metadata about your data set.

ISA data model

The elements in this web form are inspired by the [ISA data model](#) that was originally developed by the [ISA A working group](#) coordinated by the Oxford e-Research Centre. The ISA model was designed for use in *o* *m*ics-related domains, but is flexible enough for metadata modeling in other research domains. See [Figure 1](#).

PISA - the DataHub implementation

At DataHub we use the ISA data model with an additional top-level category, i.e. P for Project, to be more aligned with the hierarchical folder structure in the iRODS system. Additionally we made design choices about cardinality of the individual levels in PISA.

PISA definitions

PISA is an acronym for **P**roject, **I**nvestigation, **S**ample, **A**ssay and forms a layered model; each layer being a level at which metadata can be entered. As such, each PISA-level has its own metadata template. See [Figure 3](#).



Currently, DataHub only provides metadata web forms for the Project- and Investigation-level. We will extend our functionality in the future to support user-friendly metadata entry on all PISA-levels. In the meantime we highly encourage each user to provide metadata on the Sample- and Assay levels as well, by using spreadsheets or other formats.

For your convenience, we provide example spreadsheets as attachments to this page.

- [Example of a Sample metadata spreadsheet](#)
- [Example of an Assay metadata spreadsheet](#)

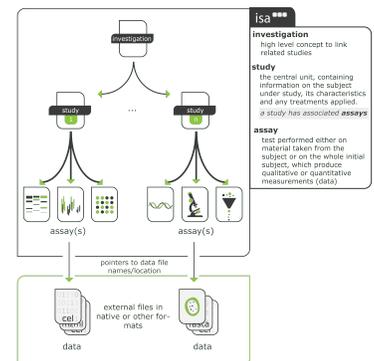


Figure 1: The ISA model



Figure 2: The tower of PISA as mascot for the PISA model

Project

The highest level of organization. It encompasses all data from a same context. Metadata on this level will typically be defined during your [project intake process](#).

Project examples

- PhD project
- Consortium project
- Clinical Study
- Student's thesis project

Investigation

The smallest amount of samples that still form a complete story. The data that shape a complete investigation are therefore highly influenced by the research question.

Investigation examples

Research Question	Investigation
What is the optimal setting for my microscope to measure slices of tissue X?	Microscope images of tissue X taken on different apertures, exposure time etc.
What is the dose- and time-effect of compound Y on HepG2 cells	Measurement results of different doses and time points from both the treated and control samples.
What is the level of CXCR4 expression in T-lymphocytes	qPCR results of expression of CXCR4- and corresponding household gene.



Conceptually, defining the scope of an investigation can be very challenging. The investigation should not be:

- **too small**, because selection of samples belonging together can become difficult during the analysis phase
- **too large**, because data in your collection can become too complicated and interrelated. Additionally, there is a risk that the duration of your sample collection phase transcends the maximum lifetime of a drop zone, being 3 months.

If you need help defining the scope of your investigation, please don't hesitate to [contact us](#).

Sample

Biological material that acts as a central unit in the experiment to which treatments or measurements are applied. Each investigation contains 1 to n samples. Each sample should be accompanied with proper metadata about biological origin, species, treatment, etc.



Feel free to use [this spreadsheet](#) as starting point for your Sample metadata.

Assay

Measurements performed on samples. Each sample in an investigation is associated with 1 to n assays. Each sample in an assay should be properly annotated with (technical) metadata about machine settings, machine type, measurement date, etc. and **most importantly**: the pointer to the resulting **data file of this sample-assay combination**.



Feel free to use [this spreadsheet](#) as starting point for your Assay metadata.

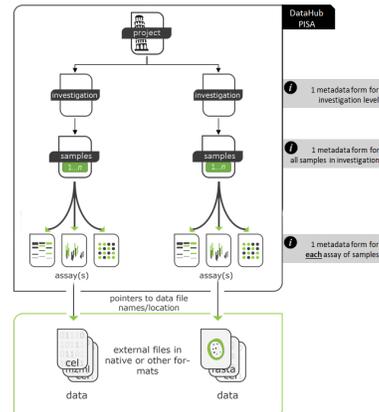


Figure 3: The DataHub PISA implementation

Mapping PISA to iRODS

Figure 4 shows the relationship between concepts in PISA and their corresponding element in the iRODS data structure.

- **Project**; metadata is registered in the iRODS database. Each project has its own path in iRODS.
 - For example: `/nlmumc/projects/P000000009`
- **Investigation**; metadata is partly registered in the iRODS database and partly in the *metadata.xml* file. Each investigation is stored as a new Collection in iRODS.
 - For example: `/nlmumc/projects/P000000009/C000000002`
- **Sample**; metadata is registered in spreadsheets, XML, RDF or any other preferred format by the researcher. Sample metadata is stored as file(s) in the Collection.
 - For example: `/nlmumc/projects/P000000009/C000000002/s_study_sample.txt`
- **Assay**; metadata is registered in spreadsheets, XML, RDF or any other preferred format by the researcher. Sample metadata is stored as file(s) in the Collection.
 - For example: `/nlmumc/projects/P000000009/C000000002/a_transcription_micro_1.txt`
- **Data files**; are stored in separate files or subfolders of the collection. It's up to the researcher how the data files are being organised, as long as there is a valid link from Assay-metadata to data file!
 - For example: `/nlmumc/projects/P000000009/C000000002/52078100929382020215419332913403.cel`
 - For example: `/nlmumc/projects/P000000009/C000000002/microscope_data/sample_id_data_file.tif`

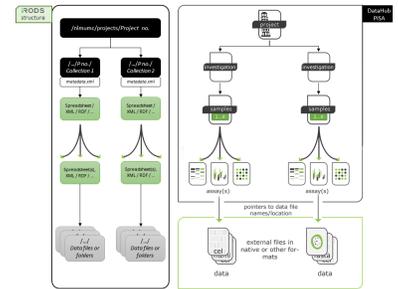


Figure 4: Relationship between the PISA model and the data organisation in iRODS