

Made for the PALs meeting in Munich, 9-10th December 2015.

Our goal in this tutorial is to arm you with the knowledge and skills essential for generating citable data. There are 4 broad stepping stones to generating citable data, which are explained [here](#). In our tutorial we will focus mostly on Steps 2 to 4.

You will use a special version of SEEK for the workshop: <https://pals.fairdomhub.org> in this SEEK you will login and generate your own citable data structure. It is a workshop specific version of the main SEEK that is publicly used so we can demonstrate brand new functionality. So please keep this in mind during the tutorial.

You will first need to register an account on this SEEK.

Any questions for discussion that come up during the workshop can be posted here, please add your name if you're not too shy:

XXX

Task 1 - The basics

Before we begin the hands on we ask you to consider the following questions, and prepare answers to feed back to the group:

1. What is meant by citable data? Are there any key guidelines to assist us?
2. What is ISA structure (hint: Investigation, Study, Assay)? How is it relevant to presenting scientific investigations?
3. What is a commons? What is the advantage over a database?
4. What is a persistent identifier? Provide examples.
5. Why is making data available in supplementary material not enough?
6. During this workshop we will not use versioning of data, we will take a snapshot of the data we want to make citable for a publication. Why do you think that is?

Task 2 - Setting up the data set we want to make citable

One of our best examples of citable (and reproducible!) data sets in SEEK is the supporting data from an extensive study into the glucose metabolism of plasmodium falciparum (<http://www.ncbi.nlm.nih.gov/pubmed/25693925>). Rather than having to download the data via supplementary folder, the user can access the entire data structure stored in SEEK directly from a link provided in the paper ([10.15490/seek.1.investigation.56](#))

In order to start with the hands-on session you can either

(i) Use your own data to set up an ISA structure in SEEK which you will snapshot and make citable. **OR**

(ii) Use a small subsection of the existing data set of the plasmodium falciparum model to complete your ISA structure in SEEK, which you will snapshot and make citable.

Please note that the data you attach will become public for the duration of the workshop and deleted after the workshop, and that the ISA structure, and DOI we create will also be erased after the workshop.

For more detailed documentation you can see <http://fairdomhub.org/presentations/175>

(i) Using your own data

1. Log in
2. Identify how you want to structure your data within the ISA framework. If you need assistance then please chat with any member of the FAIRDOM Team. There is no one correct way of representing your work within the ISA structure, so feel free to have a play around. You can use the plasmodium falciparum investigation ([10.15490/seek.1.investigation.56](https://seek.sysmo-db.org/investigations/10.15490/seek.1.investigation.56)) to help you.
3. Generate the Investigation, which is the high level concept all Studies and Assays will belong to.
4. Generate a Study onto which your first (or both) data set(s) can be linked.
5. Generate an Assay for just one data set, and link the data files to the Assay.
6. Generate a publication (if appropriate) and link the publication to the Study and Investigation.

(ii) Using the plasmodium falciparum data

1. Log in
2. Create a new Investigation using on “Glucose metabolism in plasmodium falciparum trophozoites” as a guide [10.15490/seek.1.investigation.56](https://seek.sysmo-db.org/investigations/10.15490/seek.1.investigation.56). (NB This investigation will open in the main public SEEK, not in the special version of SEEK you will build your ISA structure).
3. Create a Study using the “model construction” study in the ISA as a guide (<https://seek.sysmo-db.org/studies/118> NB This investigation will open in the main public SEEK, not in the special version of SEEK you will build your ISA structure).
4. Select one of the Assays attached to the “model construction” study, and use it as a guide to generate an Assay in SEEK.
5. Attach the appropriate data files to the Assay you have made.
6. Generate a publication and link it to the Studies, and Investigation.

Task 3 - Generating snapshots of the data and associated DOIs

The ISA structures can be modified through time, therefore we use snapshots to record an instance of the ISA structure at a given point in time. These snapshots can be accessed at a later date.

1. Create a snapshot of your current ISA structure, and request a DOI for that snapshot.
2. Now we want to modify the ISA structure:
 - a. If you are using your own data include another Assay, or Study and Assay with your second data set.
 - b. If you are using the plasmodium falciparum ISA structure, select an additional Assay to append to the “model generation” study.
3. Create a second snapshot of the ISA structure, and request a DOI.
4. Share both DOIs with your neighbour and have them check the differences between the snapshots.

Task 4 - Downloading a Research Object, taking a peek inside.

(This will be demonstrated by the FAIRDOM Team).