

Some Considerations on Samples

Rolf Krahl 

ICAT F2F Meeting, 03 May 2023, Berlin



- Hypothesis: the best sample description is its history record.
- Most samples in our facilities are specifically created for the experiment.
- Sample preparation and characterization often create data and/or logbook entries.
- If all the artifacts generated along the sample preparation are made FAIR and linked together, they probably constitute the best sample description we could get.
- Conclusion: don't worry too much on how to describe samples. Rather collect and combine all the information that is already there.
- We need a sample PID for linking all that information together.
- Some (rather generic) ontology or controlled vocabulary for the classification of sample types might still be useful.

Sample PID Option: IGSN

- International Generic Sample Number (IGSN) may be a candidate for a sample PID.
- Technically, (since this year) IGSNs are normal DataCite DOIs. We'd be constrained to normal DataCite Metadata schema in the PID record.
- DataCite recommends (and may even require) to use a dedicated prefix for IGSNs.

Sample Database

- We need a facility sample database.
- Integrating with various facility workflows will be the most involved part in the implementation.
- Sample database should provide an API to allow to register samples.
- Minting of IGSNs could be done from the database and it could provide a sample landing page.
- The database needs to have at least the PID metadata, plus some generic sample description (tbd).
- Suggestion: prototype that as a separate database first. Full integration into ICAT may be considered as a later step.
- In any case, each entry in that database should link to a corresponding entry in the sample table in ICAT.
- Note: having a dedicated prefix for IGSNs means that we could use a local database name as suffix for the IGSN, e.g. we could predict the final IGSN even before minting them.