A Galaxy tool definition powered by the ISA-API

The Galaxy tool is built around a capability recently added to the ISA API. The ISA Create Mode, by embedding key knowledge about design of experiments and practicing study design descriptions at the top of the reporting requirements, allows the rapid creation of regularized data management templates.

ISA-Tab format as Galaxy data type

These can be tuned to implement annotation requirements and naming conventions, provided a set of core rules underwriting the templating engine are followed. We demonstrate the advantages of the approach by developing a component aimed at reporting experiments using metabolomics techniques, and that supports both NMR and Mass spectrometry approaches, and their various data acquisition modalities. It is being tested at some of the UK National Phenome Centres.

Semantic markup of ISA templates is achieved by populating many of the select options from an external file that stores all controlled vocabulary and terminology values and feeding those values to the templating engine. Study design information is semantically marked up with ontology terms coming from STANDO (https://stando-ontology.org), while other terms are drawn from the Metabolomics Standards Initiative Ontology (https://github.com/ISA-tools/MSIO).

A Galaxy workflow for study deposition/pre-registration to MetaboLights

Future developments:

The flexibility and generivity of the approach is highlighted by extending the initial metabolomics use case to support the reporting of genomics data according to similar principles. Hence, a complementary tool offers similar capabilities for prospective metadata templating for functional genomics applications using DNA microarray or NGS techniques. This further validates the versatility of the ISA specifications, supporting multiomics experiments, which allow molecular phenotyping to be tracked in a consistent fashion.

Expand the tools and workflows ISA aware and capable of producing FAIR datasets:

ISAcreate offers a central component to the data custody chain, by providing means to envisage study pre-registration in the context of Galaxy, the analysis workflow engine, thus placing scientists closer to be able to do reproducible data science.