ASPIS

ASPIS – A shield used in ancient Greece; revolutionary defensive tool enabling advancing legions to join forces for better protection

Enabling innovation: from data science research to regulatory application

Philippe Rocca-Serra | University of Oxford

email: philippe.rocca-serra@oerc.ox.ac.uk
linkedin: /philipperoccaserra/
github: @proccaserra
orcid: 0000-0001-9853-5668

JRC meeting

- 31.01.2024

The ASPIS Cluster

PRECISION TOX

Leveraging evolutionary diversity to reveal the molecular basis of toxicity

ONTOX

Synthesising toxicology knowledge to support next generation risk assessment

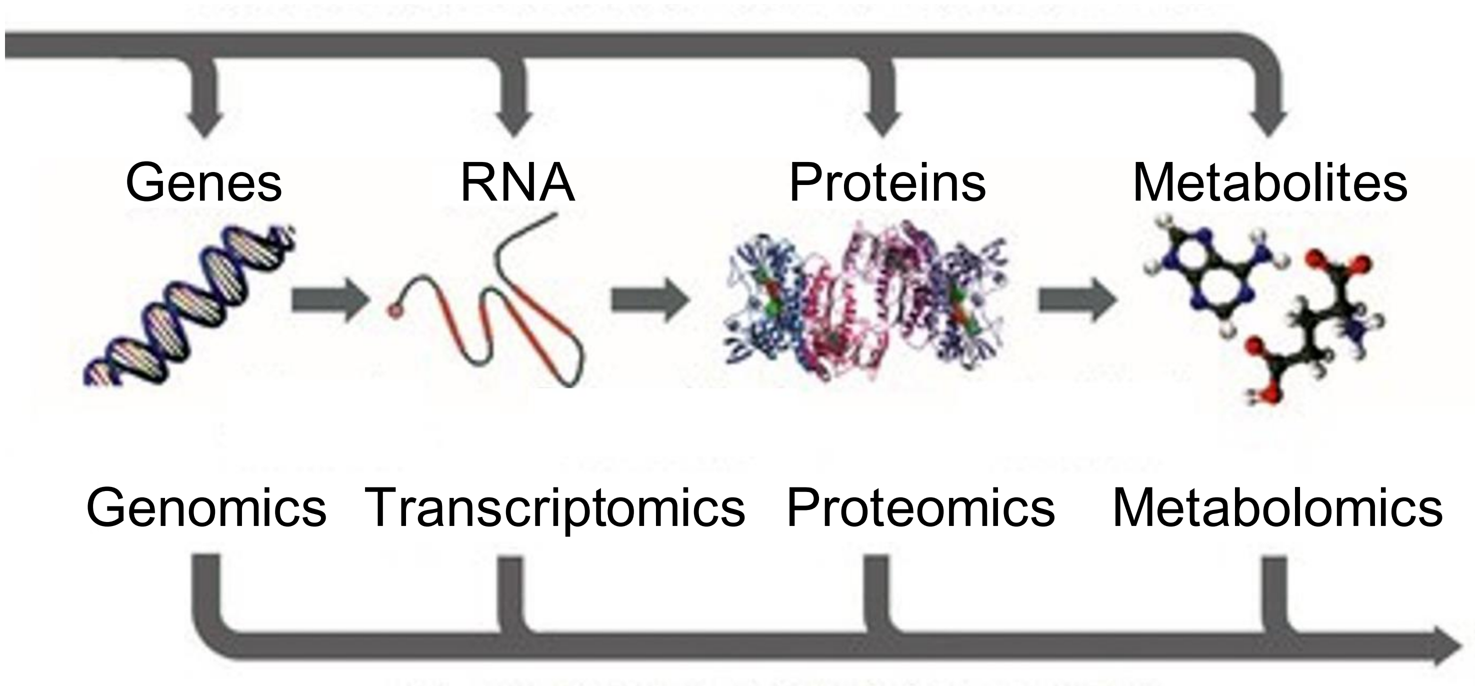
RISK HUNT3R

Human-centric chemical safety assessment utilising systems toxicology

Biomolecular Research Data

Upstream input from exposure to chemicals

Chemicals



Genes

RNA

Proteins

Metabolites

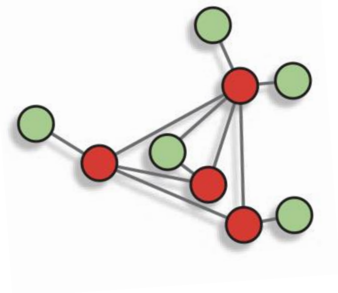
Genomics

Transcriptomics

Proteomics

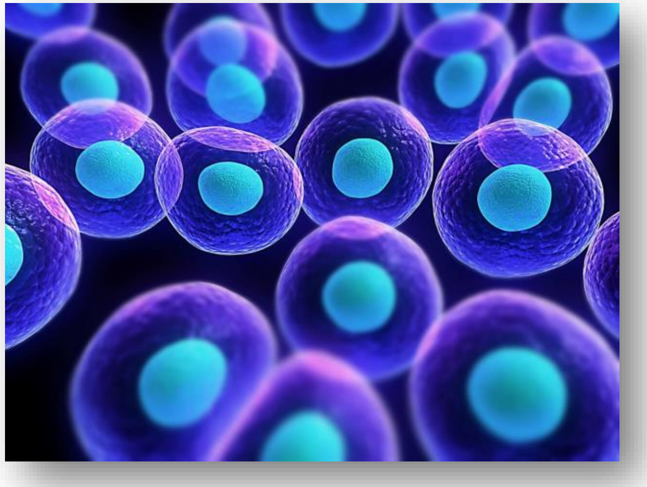
Metabolomics

Biological Adversity



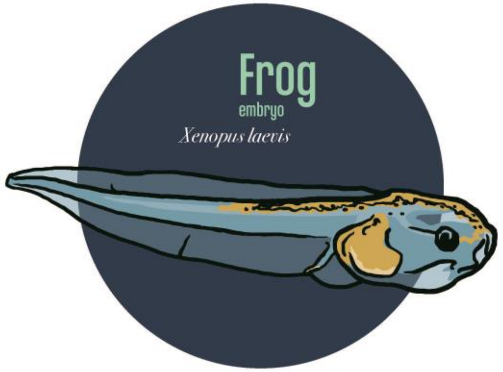
Downstream output from the biomolecular responses

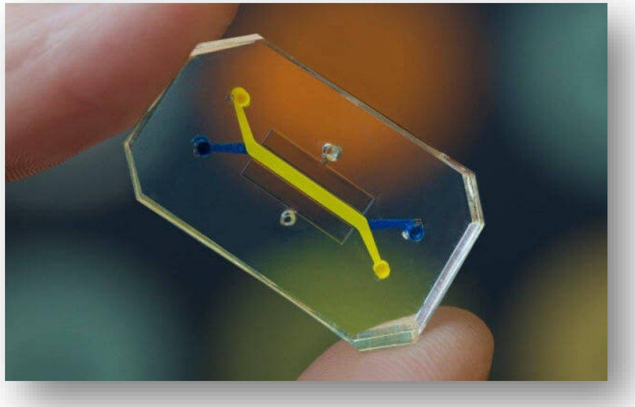
Experimental Data from Variety of Testing Platforms

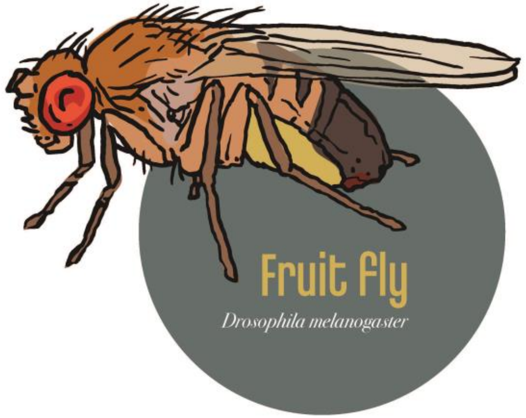


Existing Data

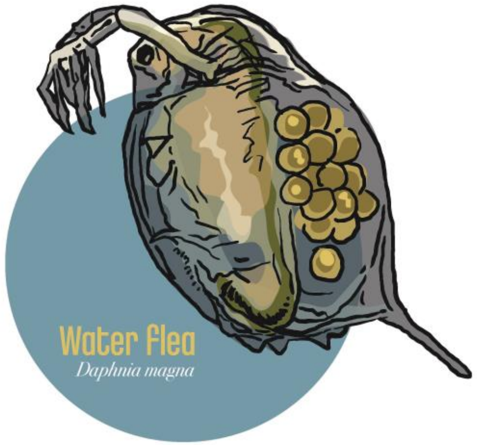
Zebrafish
embryo
Danio rerio



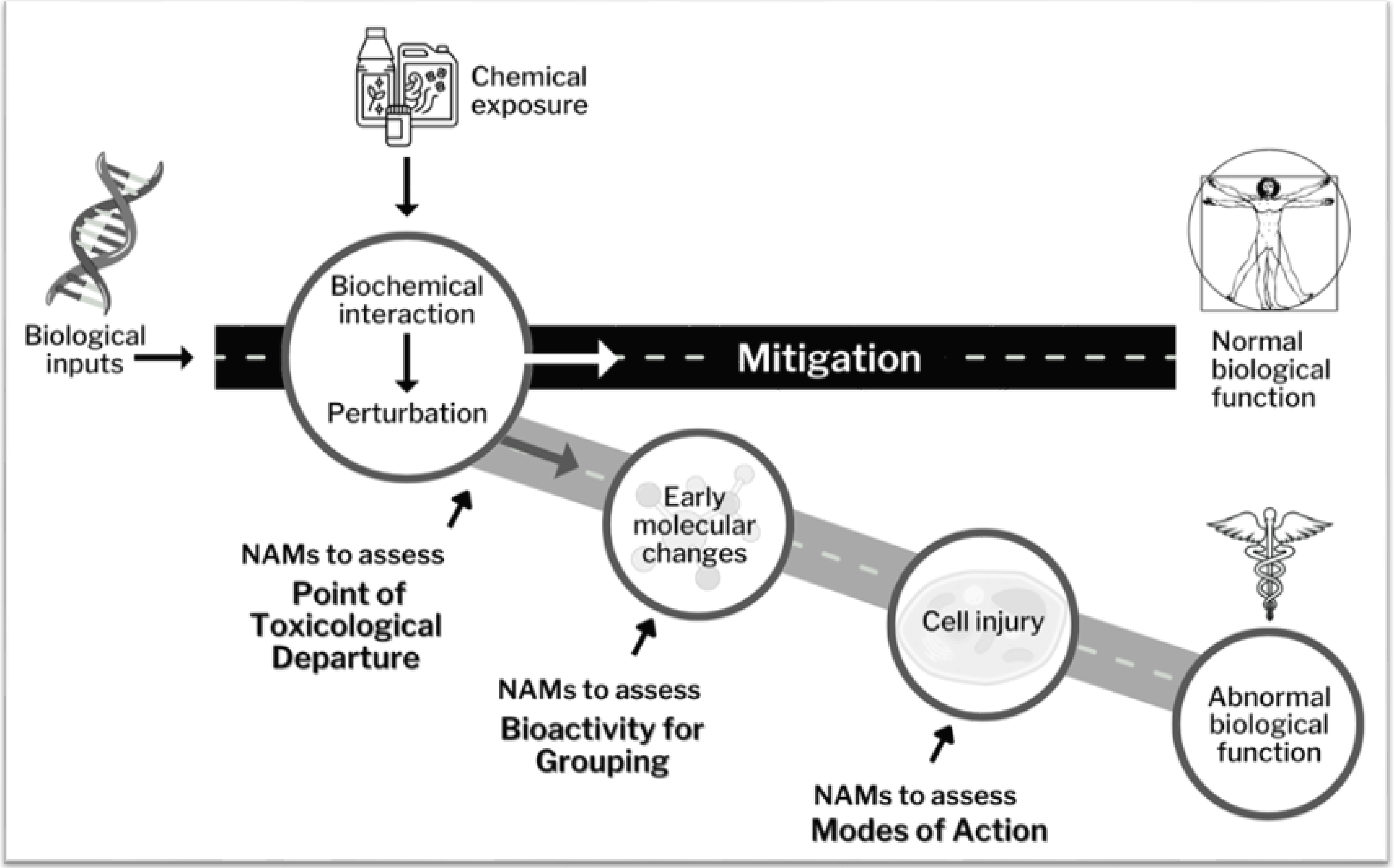




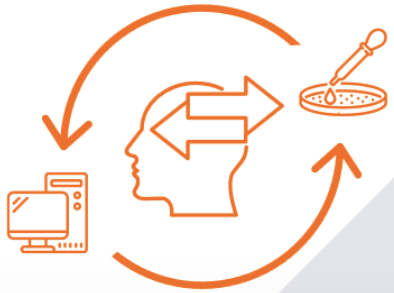
Round
worm
Caenorhabditis elegans



The New Paradigm: Activation of Toxicity Pathways



Each NAM will consist of an ontology-driven and artificial intelligence-based computational system linked with a battery of *in vitro* assays and ¨*in silico* tools for hazard prediction combined with customized exposure assessment.



Why is F.A.I.R. important for innovation and regulatory applications?

- FAIR means self-describing data
- FAIR means explicit semantics (commitment)
- FAIR means availability of data to software agents ("machine actionability")
- By ensuring sufficient metadata always accompany data, FAIR enables trust
- FAIR practices protect investments by enhancing data reuse potential

Planning for FAIR from the onset

- Survey of the art - catalogues of standards (FAIRsharing) - Promoting reuse
  - Domain-specific data repositories & associated requirements
  - Data deposition syntax
  - Controlled vocabularies / terminologies
- Semantic models and identifiers patterns
  - Preparing for Extraction Transform Load (ETL) to RDF & KG from data sources
- Licensing and access rights around in machine actionable form
- Data release and cataloguing for optimum discoverability and reuse
- Consolidation in an extensive and detailed Data Management Plan

Testing process during pilot phases

- As with all projects, a central issue is that of phasing:
  - Delivering the right tool at the right moment
  - Developing lean processes as mitigating plans
    - PrecisionTox Spreadsheet Templates with naming conventions
    - Collect signal for refine software specifications
- Taking advantage of projects pilots
  - Coordination user requirements for data collection
  - Defining annotation requirements with subject matter experts
  - Developing software solution prototypes
  - Following software engineering best practice standards
    - Continuous integration, Code review, test driven development, automation
    - Code documentation, user documentation…

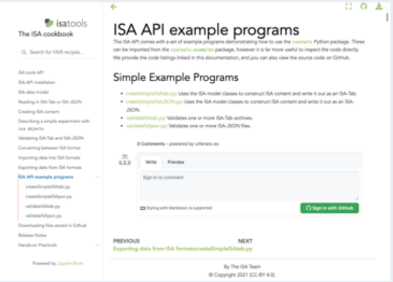
Key concept: prospective FAIR approach

- Implementation of **Metadata Manager Tool -** PrecisionTox example
- Globus based file transfer process between data producing centers
- Data publication / dissemination procedure
- Embedding FAIR practice in computational workflows
- Ready for knowledge graph representations

Developing software to make data FAIR by design

METHODOLOGY

A rich Python API supports the ISA metadata formats, allowing processing of ISA formatted documents either as ISA-Tab or ISA-JSON. Support to perform conversions to formats required by public repositories, such SRA or MAGE-TAB.



isatools
The ISA cookbook
Search for FAIR recipes...

- ISA tools API
- ISA-API installation
- ISA data model
- Reading in ISA-Tab or ISA-JSON
- Creating ISA content
- Describing a simple experiment with isa objects
- Validating ISA-Tab and ISA-JSON
- Converting between ISA format
- Importing data into ISA formats
- Exporting data from ISA formats
- ISA API example programs
  - createSimpleISAtab.py
  - createSimpleISAjson.py
  - validateISAtab.py
  - validateISAjson.py
- Downloading files stored in Github
- Release Notes
- Hands-on Practicals

Powered by Jupyter Book

ISA API example programs

The ISA API comes with a set of example programs demonstrating how to use the isatools Python package. These can be imported from the isatools.examples package, however it is far more useful to inspect the code directly. We provide the code listings linked in this documentation, and you can also view the source code on GitHub.

Simple Example Programs

- createSimpleISAtab.py: Uses the ISA model classes to construct ISA content and write it out as an ISA-Tab.
- createSimpleISAJSON.py: Uses the ISA model classes to construct ISA content and write it out as an ISA-JSON.
- validateISAtab.py: Validates one or more ISA-Tab archives.
- validateISAjson.py: Validates one or more ISA-JSON files.

0 Comments - *powered by utteranc.es*

Write Preview

Sign in to comment

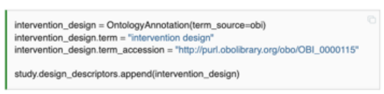
Styling with Markdown is supported

Sign in with GitHub

PREVIOUS
Exporting data from ISA formats

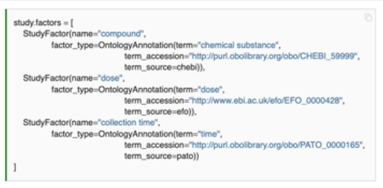
NEXT
createSimpleISAtab.py

By The ISA Team
© Copyright 2021 (CC-BY 4.0).



```
intervention_design = OntologyAnnotation(term_source=obi)
intervention_design.term = "intervention design"
intervention_design.term_accession = "http://purl.obolibrary.org/obo/OBI_0000115"

study.design_descriptors.append(intervention_design)
```



```
study.factors = [
   StudyFactor(name="compound",
            factor_type=OntologyAnnotation(term="chemical substance",
                                    term_accession="http://purl.obolibrary.org/obo/CHEBI_59999",
                                    term_source=chebi)),
   StudyFactor(name="dose",
            factor_type=OntologyAnnotation(term="dose",
                                    term_accession="http://www.ebi.ac.uk/efo/EFO_0000428",
                                    term_source=efo)),
   StudyFactor(name="collection time",
            factor_type=OntologyAnnotation(term="time",
                                    term_accession="http://purl.obolibrary.org/obo/PATO_0000165",
                                    term_source=pato))
]
```

Describe experimental plan:
identify independent variables and their levels

Generate ISA objects and serialize to TAB or JSON format

WIP

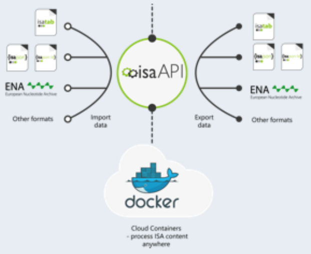
Generate robotic platform compatible list of samples
with Precision Tox Identifier scheme

WIP

Generate sample run order using randomization function during key lab steps:
material extraction and
data acquisition

Support the reporting of Quality Control (QC) samples / runs and record processing batch information

Test data deposition routes with EMBL-EBI and INSDC sequence read archives



RESULTS

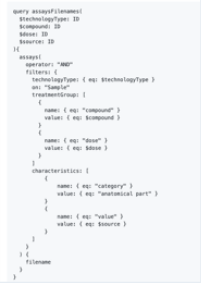
Data: capturing key information while performing behind the scene semantic markup thanks to ontology aware mappings

Code: development of new components such as:

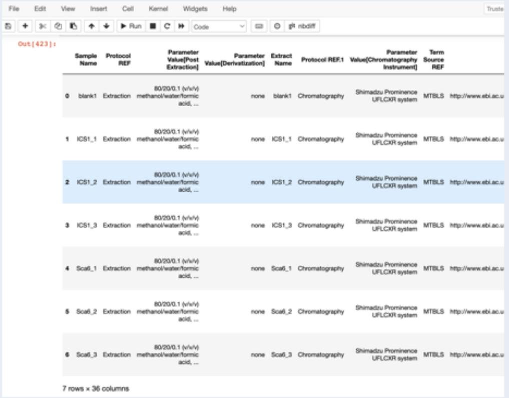
- ISAcreate mode
- ISA GraphQL
- ISA LD

ISA GraphQL interface allows efficient information extraction on individual ISA documents

https://github.com/ISA-tools/isa-api/tree/develop/isatools/graphQL



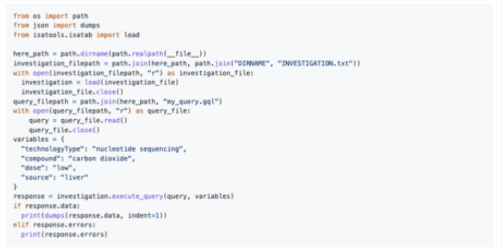
```
query assaysFilenames(
  $technologyType: ID
  $compound: ID
  $dose: ID
  $source: ID
){
  assays(
    operator: "AND"
    filters: {
      technologyType: { eq: $technologyType }
      on: "Sample"
      treatmentGroup: [
        {
          name: { eq: "compound" }
          value: { eq: $compound }
        }
        {
          name: { eq: "dose" }
          value: { eq: $dose }
        }
      ]
      characteristics: [
          {
              name: { eq: "category" }
              value: { eq: "anatomical part" }
          }
          {
              name: { eq: "value" }
              value: { eq: $source }
          }
      ]
    }
  ) {
    filename
  }
}
```



Out[423]:

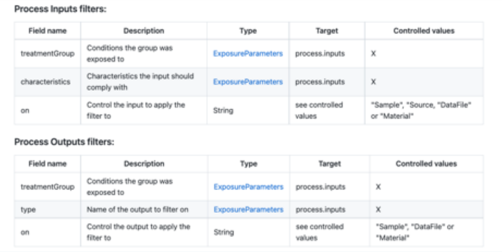
	Sample Name	Protocol REF	Parameter Value[Post Extraction]	Parameter Value[Derivatization]	Extract Name	Protocol REF.1	Parameter Value[Chromatography Instrument]	Term Source REF	
0	blank1	Extraction	80/20/0.1 (v/v/v) methanol/water/formic acid, ...	none	blank1	Chromatography	Shimadzu Prominence UFLCXR system	MTBLS	http://www.ebi.ac.u
1	ICS1_1	Extraction	80/20/0.1 (v/v/v) methanol/water/formic acid, ...	none	ICS1_1	Chromatography	Shimadzu Prominence UFLCXR system	MTBLS	http://www.ebi.ac.u
2	ICS1_2	Extraction	80/20/0.1 (v/v/v) methanol/water/formic acid, ...	none	ICS1_2	Chromatography	Shimadzu Prominence UFLCXR system	MTBLS	http://www.ebi.ac.u
3	ICS1_3	Extraction	80/20/0.1 (v/v/v) methanol/water/formic acid, ...	none	ICS1_3	Chromatography	Shimadzu Prominence UFLCXR system	MTBLS	http://www.ebi.ac.u
4	Sca6_1	Extraction	80/20/0.1 (v/v/v) methanol/water/formic acid, ...	none	Sca6_1	Chromatography	Shimadzu Prominence UFLCXR system	MTBLS	http://www.ebi.ac.u
5	Sca6_2	Extraction	80/20/0.1 (v/v/v) methanol/water/formic acid, ...	none	Sca6_2	Chromatography	Shimadzu Prominence UFLCXR system	MTBLS	http://www.ebi.ac.u
6	Sca6_3	Extraction	80/20/0.1 (v/v/v) methanol/water/formic acid, ...	none	Sca6_3	Chromatography	Shimadzu Prominence UFLCXR system	MTBLS	http://www.ebi.ac.u

7 rows × 36 columns



```
from os import path
from json import dumps
from isatools.isatab import load

here_path = path.dirname(path.realpath(__file__))
investigation_filepath = path.join(here_path, path.join("DIRNAME", "INVESTIGATION.txt"))
with open(investigation_filepath, "r") as investigation_file:
  investigation = load(investigation_file)
  investigation_file.close()
query_filepath = path.join(here_path, "my_query.gql")
with open(query_filepath, "r") as query_file:
    query = query_file.read()
    query_file.close()
variables = {
  "technologyType": "nucleotide sequencing",
  "compound": "carbon dioxide",
  "dose": "low",
  "source": "liver"
}
response = investigation.execute_query(query, variables)
if response.data:
  print(dumps(response.data, indent=1))
elif response.errors:
  print(response.errors)
```



Process Inputs filters:

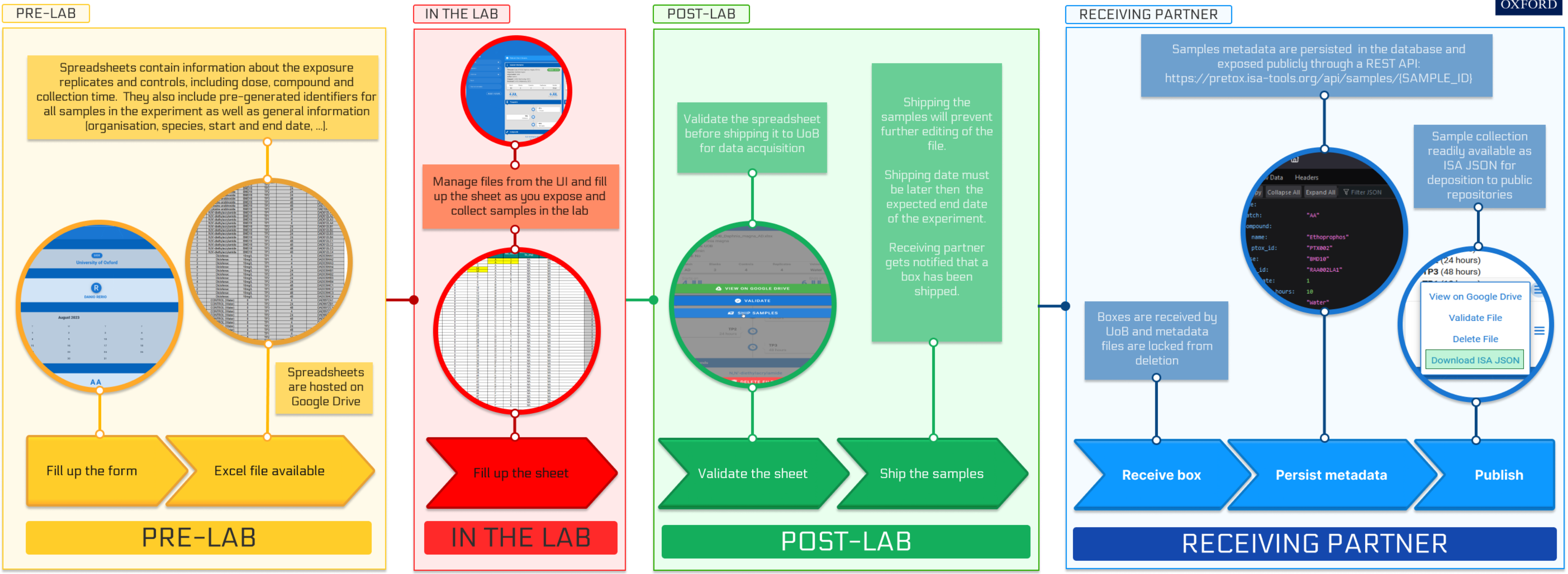
Field name	Description	Type	Target	Controlled values
treatmentGroup	Conditions the group was exposed to	ExposureParameters	process.inputs	X
characteristics	Characteristics the input should comply with	ExposureParameters	process.inputs	X
on	Control the input to apply the filter to	String	see controlled values	"Sample", "Source, "DataFile" or "Material"

Process Outputs filters:

Field name	Description	Type	Target	Controlled values
treatmentGroup	Conditions the group was exposed to	ExposureParameters	process.inputs	X
type	Name of the output to filter on	ExposureParameters	process.inputs	X
on	Control the output to apply the filter to	String	see controlled values	"Sample", "DataFile" or "Material"

Developing software to make data FAIR at design

Precision Toxicology Metadata Manager: sample exposure and collection



PRE-LAB

Spreadsheets contain information about the exposure replicates and controls, including dose, compound and collection time. They also include pre-generated identifiers for all samples in the experiment as well as general information (organisation, species, start and end date, …).

Spreadsheets are hosted on Google Drive

Fill up the form

Excel file available

PRE-LAB

IN THE LAB

Manage files from the UI and fill up the sheet as you expose and collect samples in the lab

Fill up the sheet

IN THE LAB

POST-LAB

Validate the spreadsheet before shipping it to UoB for data acquisition

Shipping the samples will prevent further editing of the file.

Shipping date must be later then the expected end date of the experiment.

Receiving partner gets notified that a box has been shipped.

Validate the sheet

Ship the samples

POST-LAB

RECEIVING PARTNER

Samples metadata are persisted in the database and exposed publicly through a REST API: https://pretox.isa-tools.org/api/samples/{SAMPLE_ID}

Sample collection readily available as ISA JSON for deposition to public repositories

Boxes are received by UoB and metadata files are locked from deletion

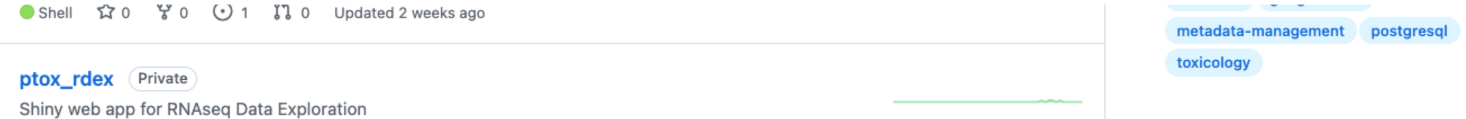
Receive box

Persist metadata

Publish

RECEIVING PARTNER

Batista Dominique (0000-0002-2109-489X), Data Readiness Group



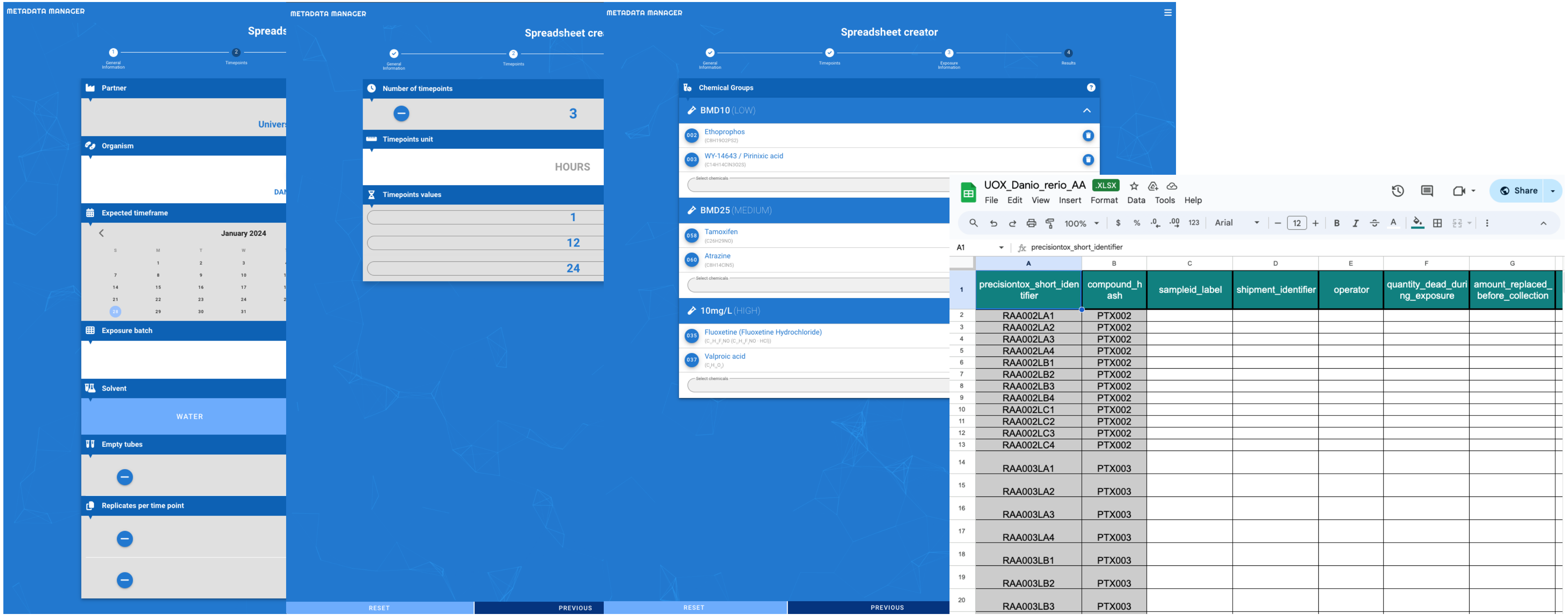
Shell ☆ 0 0 1 0 Updated 2 weeks ago

ptox_rdex Private

Shiny web app for RNAseq Data Exploration

metadata-management postgresql toxicology

Developing software to make data FAIR by design



Partner: Univer

Organism: DAI

Expected timeframe: January 2024

S	M	T	W
	1	2	3
7	8	9	10
14	15	16	17
21	22	23	24
28	29	30	31

Exposure batch

Solvent: WATER

Empty tubes

Replicates per time point

Number of timepoints: 3

Timepoints unit: HOURS

Timepoints values: 1, 12, 24

Chemical Groups

BMD10 (LOW)
- 002 Ethoprophos (C8H19O2PS2)
- 003 WY-14643 / Pirinixic acid (C14H14ClN3O2S)
- Select chemicals

BMD25 (MEDIUM)
- 058 Tamoxifen (C26H29NO)
- 060 Atrazine (C8H14ClN5)
- Select chemicals

10mg/L (HIGH)
- 035 Fluoxetine (Fluoxetine Hydrochloride) ($C_{17}H_{18}F_3NO$ ($C_{17}H_{18}F_3NO \cdot HCl$))
- 037 Valproic acid ($C_8H_{16}O_2$)
- Select chemicals

UOX_Danio_rerio_AA .XLSX

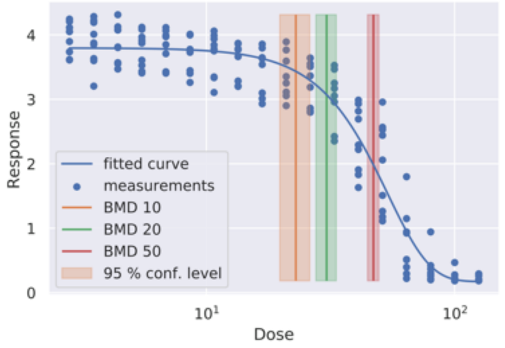
A1 precisiontox_short_identifier

	A	B	C	D	E	F	G
1	precisiontox_short_identifier	compound_hash	sampleid_label	shipment_identifier	operator	quantity_dead_during_exposure	amount_replaced_before_collection
2	RAA002LA1	PTX002					
3	RAA002LA2	PTX002					
4	RAA002LA3	PTX002					
5	RAA002LA4	PTX002					
6	RAA002LB1	PTX002					
7	RAA002LB2	PTX002					
8	RAA002LB3	PTX002					
9	RAA002LB4	PTX002					
10	RAA002LC1	PTX002					
11	RAA002LC2	PTX002					
12	RAA002LC3	PTX002					
13	RAA002LC4	PTX002					
14	RAA003LA1	PTX003					
15	RAA003LA2	PTX003					
16	RAA003LA3	PTX003					
17	RAA003LA4	PTX003					
18	RAA003LB1	PTX003					
19	RAA003LB2	PTX003					
20	RAA003LB3	PTX003					

```
{"sample":{"batch":"AA","compound":{"name":"Ethoprophos","ptox_id":"PTX002"},"dose":"BMD10","ptox_id":"RAA002LA1","replicate":1,"timepoint_hours":10,"vehicle":"Water"}}
```

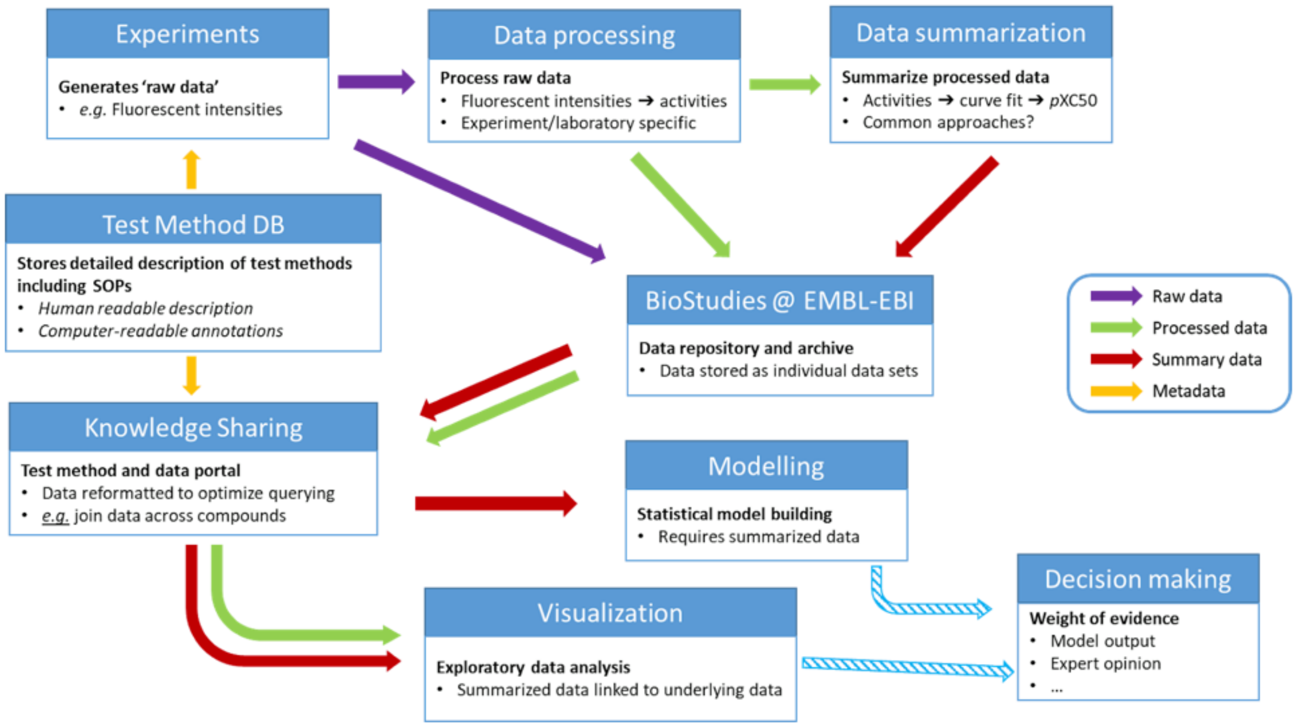
Integrated Data and Knowledge Management

RISK HUNT3R



Harmonised Data, Protocol and Metadata Management Infrastructure supported by EdelweissData

Data and Knowledge Management on EUToxRisk program



Experiments

Generates 'raw data'

- *e.g.* Fluorescent intensities

Data processing

Process raw data

- Fluorescent intensities → activities
- Experiment/laboratory specific

Data summarization

Summarize processed data

- Activities → curve fit → *pXC50*
- Common approaches?

Test Method DB

Stores detailed description of test methods including SOPs

- *Human readable description*
- *Computer-readable annotations*

BioStudies @ EMBL-EBI

Data repository and archive

- Data stored as individual data sets

Knowledge Sharing

Test method and data portal

- Data reformatted to optimize querying
- *e.g.* join data across compounds

Modelling

Statistical model building

- Requires summarized data

Visualization

Exploratory data analysis

- Summarized data linked to underlying data

Decision making

Weight of evidence

- Model output
- Expert opinion
- ...

Raw data
Processed data
Summary data
Metadata

Credits:
Barry Hardy (Edelweiss Connect)
Email: Barry.Hardy@edelweissconnect.com
LinkedIn: https://www.linkedin.com/in/barryhardy/

Data Management and Analysis

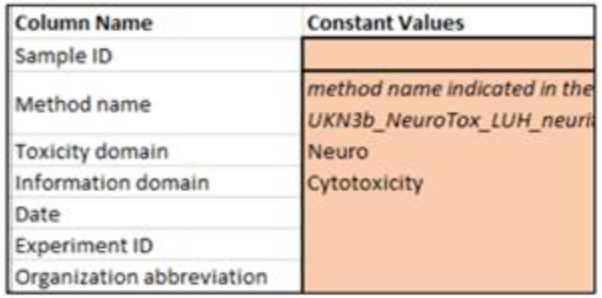
Behind the scenes

RISK HUNT3R

Data structure
Data import
Data use

Data structure

- Excel template file (EU-ToxRisk)
- Required fields for all the datasets
- Custom fields if needed



Column Name	Constant Values
Sample ID	
Method name	*method name indicated in the* *UKN3b_NeuroTox_LUH_neuri*
Toxicity domain	Neuro
Information domain	Cytotoxicity
Date	
Experiment ID	
Organization abbreviation	

Data import

- Each dataset is validated
- Published on Biostudies
- Automatically imported to EdelweissData
- Data becomes accessible through web requests (URL)

EdelweissData™

Convenient publishing of scientific data with proper versioning, rich metadata support and a powerful API

Data use

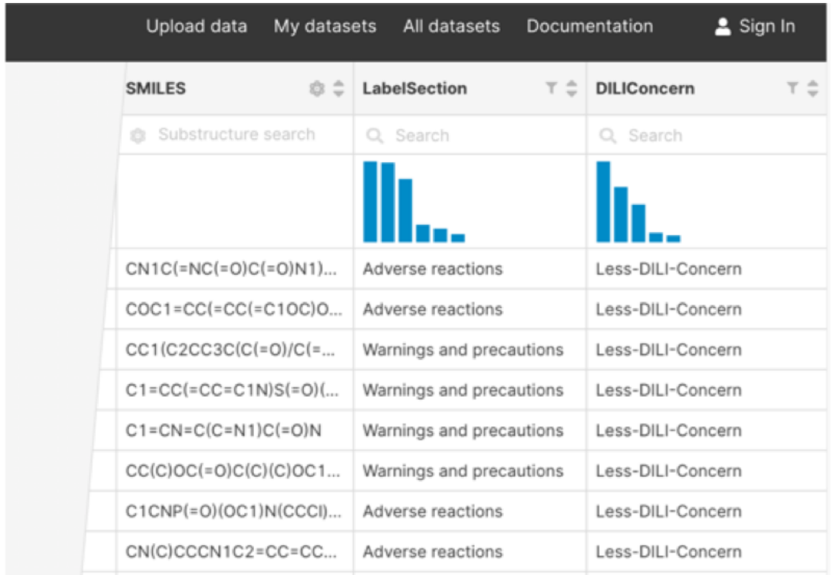
- Access data & metadata directly from the database
- Target specific version of dataset (e.g. latest)
- Consume data directly into your data analysis tool (R, Python, Excel, Jupyter, Colab, Observable)

jupyter CO

Observable

Credits:
Barry Hardy (Edelweiss Connect)
Email: Barry.Hardy@edelweissconnect.com
LinkedIn: https://www.linkedin.com/in/barryhardy/

Data Management and Analysis Workflows based on Biostudies and EdelweissData

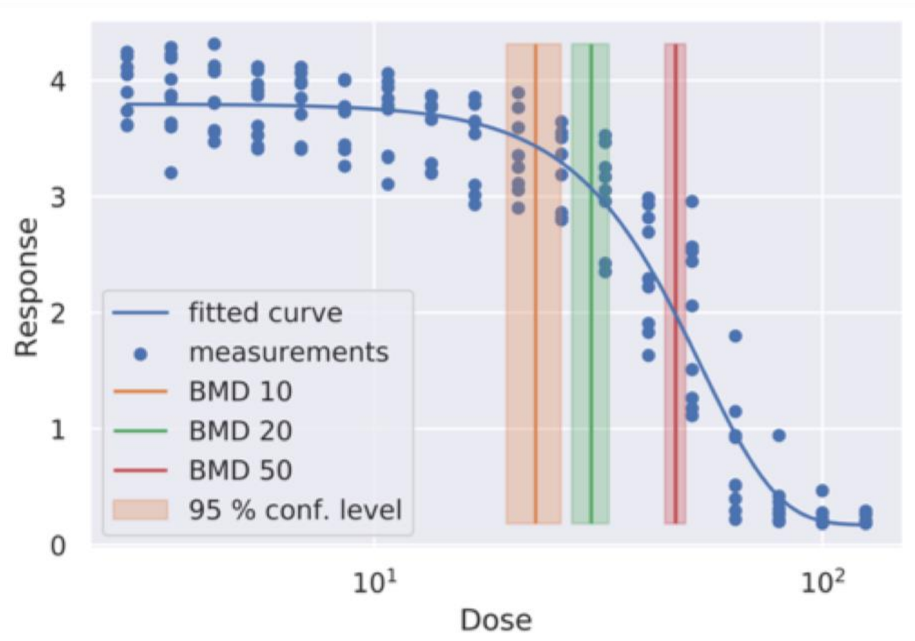


SMILES	LabelSection	DILIConcern
CN1C(=NC(=O)C(=O)N1)...	Adverse reactions	Less-DILI-Concern
COC1=CC(=CC(=C1OC)O...	Adverse reactions	Less-DILI-Concern
CC1(C2CC3C(C(=O)/C(=...	Warnings and precautions	Less-DILI-Concern
C1=CC(=CC=C1N)S(=O)(...	Warnings and precautions	Less-DILI-Concern
C1=CN=C(C=N1)C(=O)N	Warnings and precautions	Less-DILI-Concern
CC(C)OC(=O)C(C)(C)OC1...	Warnings and precautions	Less-DILI-Concern
C1CNP(=O)(OC1)N(CCCl)...	Adverse reactions	Less-DILI-Concern
CN(C)CCCN1C2=CC=CC...	Adverse reactions	Less-DILI-Concern

Supporting data life cycle from generation through to consumption.

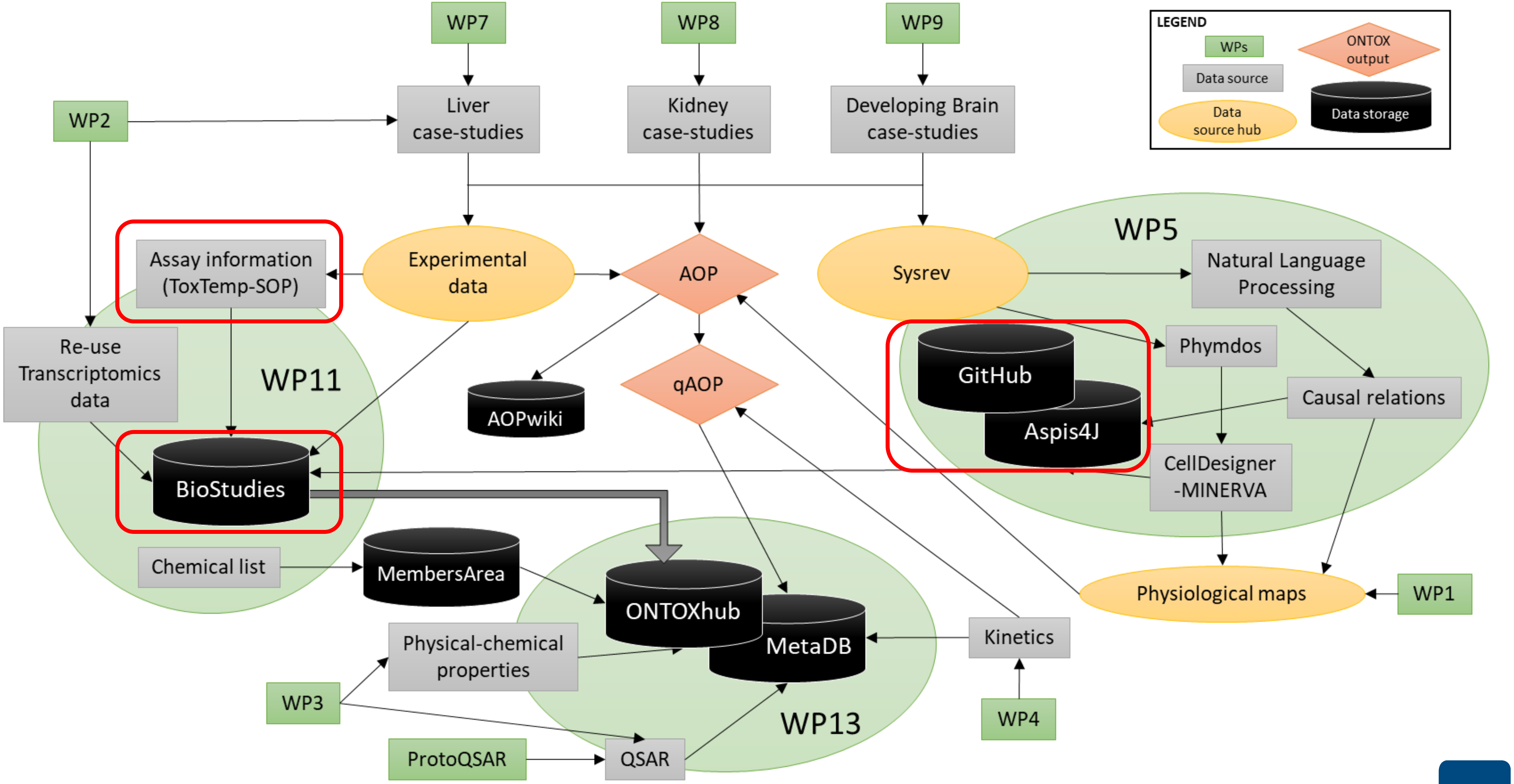
https://edelweissdata.com/

Version 1.0 of EdelweissData (harmonised data management product of Edelweiss Connect) completed by 2020 and in commercial use.

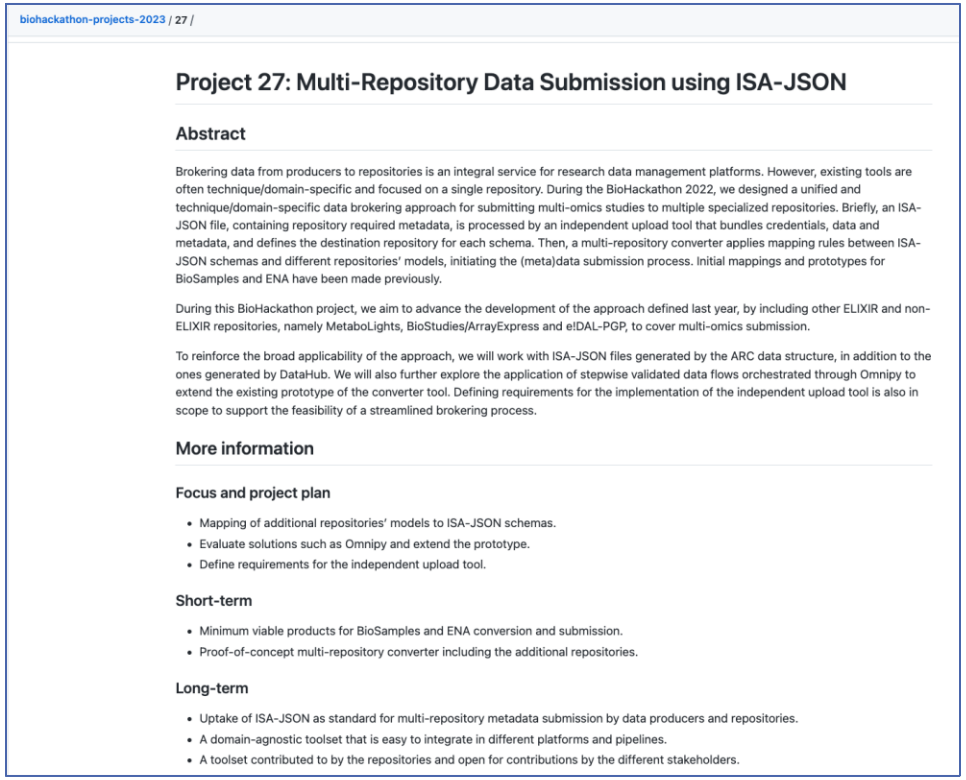


Credits:
Barry Hardy (Edelweiss Connect)
Email: Barry.Hardy@edelweissconnect.com
LinkedIn: https://www.linkedin.com/in/barryhardy/

ONTOX data flowchart



PrecisionTox and ELIXIR collaboration



Project 27: Multi-Repository Data Submission using ISA-JSON

Abstract

Brokering data from producers to repositories is an integral service for research data management platforms. However, existing tools are often technique/domain-specific and focused on a single repository. During the BioHackathon 2022, we designed a unified and technique/domain-specific data brokering approach for submitting multi-omics studies to multiple specialized repositories. Briefly, an ISA-JSON file, containing repository required metadata, is processed by an independent upload tool that bundles credentials, data and metadata, and defines the destination repository for each schema. Then, a multi-repository converter applies mapping rules between ISA-JSON schemas and different repositories' models, initiating the (meta)data submission process. Initial mappings and prototypes for BioSamples and ENA have been made previously.

During this BioHackathon project, we aim to advance the development of the approach defined last year, by including other ELIXIR and non-ELIXIR repositories, namely MetaboLights, BioStudies/ArrayExpress and e!DAL-PGP, to cover multi-omics submission.

To reinforce the broad applicability of the approach, we will work with ISA-JSON files generated by the ARC data structure, in addition to the ones generated by DataHub. We will also further explore the application of stepwise validated data flows orchestrated through Omnipy to extend the existing prototype of the converter tool. Defining requirements for the implementation of the independent upload tool is also in scope to support the feasibility of a streamlined brokering process.

More information

Focus and project plan

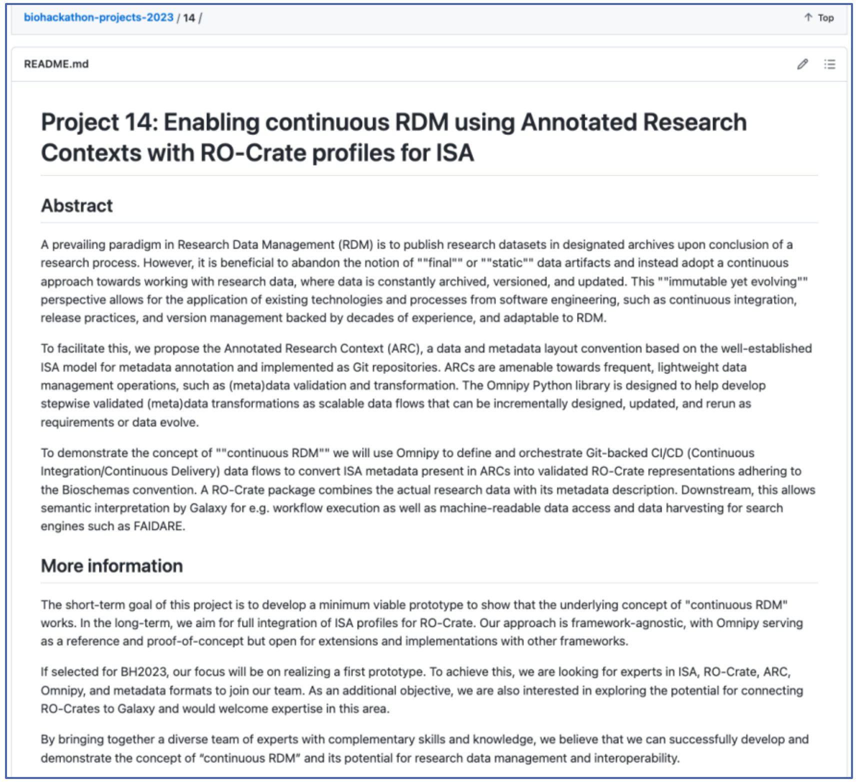
- Mapping of additional repositories' models to ISA-JSON schemas.
- Evaluate solutions such as Omnipy and extend the prototype.
- Define requirements for the independent upload tool.

Short-term

- Minimum viable products for BioSamples and ENA conversion and submission.
- Proof-of-concept multi-repository converter including the additional repositories.

Long-term

- Uptake of ISA-JSON as standard for multi-repository metadata submission by data producers and repositories.
- A domain-agnostic toolset that is easy to integrate in different platforms and pipelines.
- A toolset contributed to by the repositories and open for contributions by the different stakeholders.



Project 14: Enabling continuous RDM using Annotated Research Contexts with RO-Crate profiles for ISA

Abstract

A prevailing paradigm in Research Data Management (RDM) is to publish research datasets in designated archives upon conclusion of a research process. However, it is beneficial to abandon the notion of ""final"" or ""static"" data artifacts and instead adopt a continuous approach towards working with research data, where data is constantly archived, versioned, and updated. This ""immutable yet evolving"" perspective allows for the application of existing technologies and processes from software engineering, such as continuous integration, release practices, and version management backed by decades of experience, and adaptable to RDM.

To facilitate this, we propose the Annotated Research Context (ARC), a data and metadata layout convention based on the well-established ISA model for metadata annotation and implemented as Git repositories. ARCs are amenable towards frequent, lightweight data management operations, such as (meta)data validation and transformation. The Omnipy Python library is designed to help develop stepwise validated (meta)data transformations as scalable data flows that can be incrementally designed, updated, and rerun as requirements or data evolve.

To demonstrate the concept of ""continuous RDM"" we will use Omnipy to define and orchestrate Git-backed CI/CD (Continuous Integration/Continuous Delivery) data flows to convert ISA metadata present in ARCs into validated RO-Crate representations adhering to the Bioschemas convention. A RO-Crate package combines the actual research data with its metadata description. Downstream, this allows semantic interpretation by Galaxy for e.g. workflow execution as well as machine-readable data access and data harvesting for search engines such as FAIDARE.

More information

The short-term goal of this project is to develop a minimum viable prototype to show that the underlying concept of "continuous RDM" works. In the long-term, we aim for full integration of ISA profiles for RO-Crate. Our approach is framework-agnostic, with Omnipy serving as a reference and proof-of-concept but open for extensions and implementations with other frameworks.

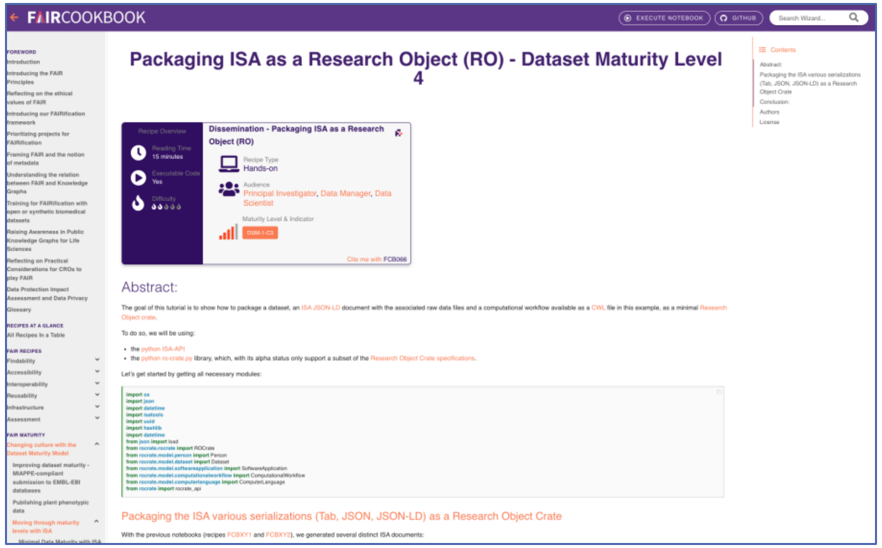
If selected for BH2023, our focus will be on realizing a first prototype. To achieve this, we are looking for experts in ISA, RO-Crate, ARC, Omnipy, and metadata formats to join our team. As an additional objective, we are also interested in exploring the potential for connecting RO-Crates to Galaxy and would welcome expertise in this area.

By bringing together a diverse team of experts with complementary skills and knowledge, we believe that we can successfully develop and demonstrate the concept of “continuous RDM” and its potential for research data management and interoperability.

Convergence & Alignment

Packaging Metadata, Data and Computational Worflow as a Research Object

Promoting adoption and convergence



FOREWORD
- Introduction
- Introducing the FAIR Principles
- Reflecting on the ethical values of FAIR
- Introducing our FAIRification framework
- Prioritizing projects for FAIRification
- Framing FAIR and the notion of metadata
- Understanding the relation between FAIR and Knowledge Graphs
- Training for FAIRification with open or synthetic biomedical datasets
- Raising Awareness in Public Knowledge Graphs for Life Sciences
- Reflecting on Practical Considerations for CROs to play FAIR
- Data Protection Impact Assessment and Data Privacy
- Glossary

RECIPES AT A GLANCE
- All Recipes in a Table

FAIR RECIPES
- Findability
- Accessibility
- Interoperability
- Reusability
- Infrastructure
- Assessment

FAIR MATURITY
- Changing culture with the Dataset Maturity Model
  - Improving dataset maturity - MIAPPE-compliant submission to EMBL-EBI databases
  - Publishing plant phenotypic data
  - Moving through maturity levels with ISA
    - Minimal Data Maturity with ISA

Packaging ISA as a Research Object (RO) - Dataset Maturity Level 4

Recipe Overview
Reading Time: 15 minutes
Executable Code: Yes
Difficulty

Dissemination - Packaging ISA as a Research Object (RO)
Recipe Type: Hands-on
Audience: Principal Investigator, Data Manager, Data Scientist
Maturity Level & Indicator: DSM-1-C3
Cite me with FCB066

Abstract:

The goal of this tutorial is to show how to package a dataset, an ISA JSON-LD document with the associated raw data files and a computational workflow available as a CWL file in this example, as a minimal Research Object crate.

To do so, we will be using:

- the python ISA-API
- the python ro-crate.py library, which, with its alpha status only support a subset of the Research Object Crate specifications.

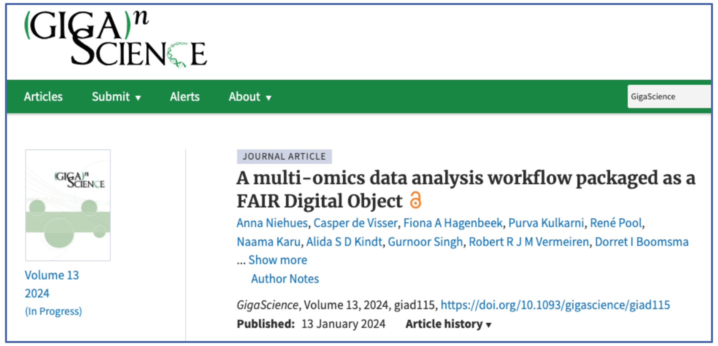
Let's get started by getting all necessary modules:

```
import os
import json
import datetime
import isatools
import uuid
import hashlib
import datetime
from json import load
from rocrate.rocrate import ROCrate
from rocrate.model.person import Person
from rocrate.model.dataset import Dataset
from rocrate.model.softwareapplication import SoftwareApplication
from rocrate.model.computationalworkflow import ComputationalWorkflow
from rocrate.model.computerlanguage import ComputerLanguage
from rocrate import rocrate_api
```

Packaging the ISA various serializations (Tab, JSON, JSON-LD) as a Research Object Crate

With the previous notebooks (recipes FCBXY1 and FCBXY2), we generated several distinct ISA documents:

Contents
- Abstract:
- Packaging the ISA various serializations (Tab, JSON, JSON-LD) as a Research Object Crate
- Conclusion:
- Authors
- License



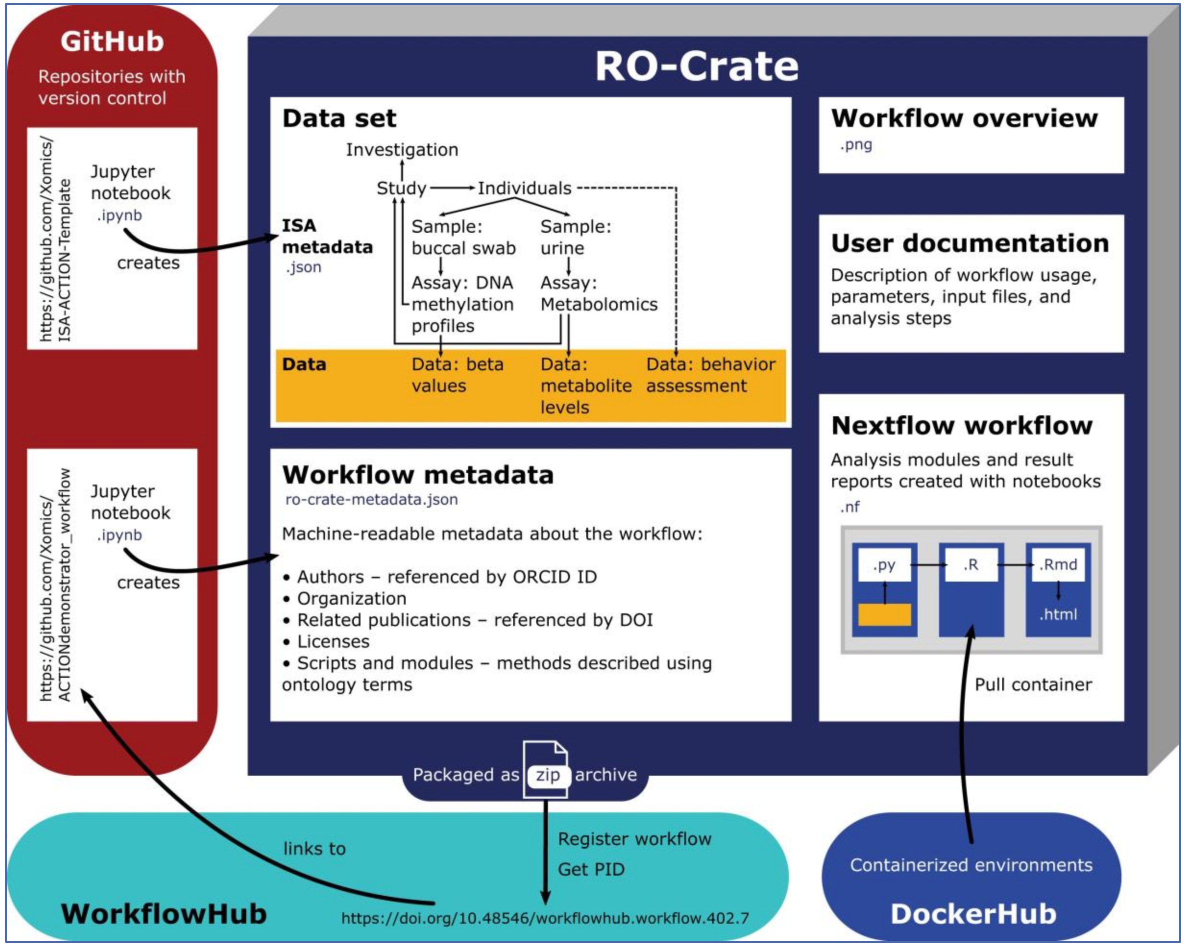
JOURNAL ARTICLE

A multi-omics data analysis workflow packaged as a FAIR Digital Object

Anna Niehues, Casper de Visser, Fiona A Hagenbeek, Purva Kulkarni, René Pool, Naama Karu, Alida S D Kindt, Gurnoor Singh, Robert R J M Vermeiren, Dorret I Boomsma
... Show more
Author Notes

Volume 13
2024
(In Progress)

GigaScience, Volume 13, 2024, giad115, https://doi.org/10.1093/gigascience/giad115
Published: 13 January 2024 **Article history** ▾



GitHub
Repositories with version control

https://github.com/Xomics/ISA-ACTION-Template
Jupyter notebook .ipynb
creates

https://github.com/Xomics/ACTIONdemonstrator_workflow
Jupyter notebook .ipynb
creates

RO-Crate

Data set
Investigation
Study → Individuals
ISA metadata .json
Sample: buccal swab
Assay: DNA methylation profiles
Sample: urine
Assay: Metabolomics
Data
Data: beta values
Data: metabolite levels
Data: behavior assessment

Workflow metadata
ro-crate-metadata.json
Machine-readable metadata about the workflow:
- Authors – referenced by ORCID ID
- Organization
- Related publications – referenced by DOI
- Licenses
- Scripts and modules – methods described using ontology terms

Workflow overview
.png

User documentation
Description of workflow usage, parameters, input files, and analysis steps

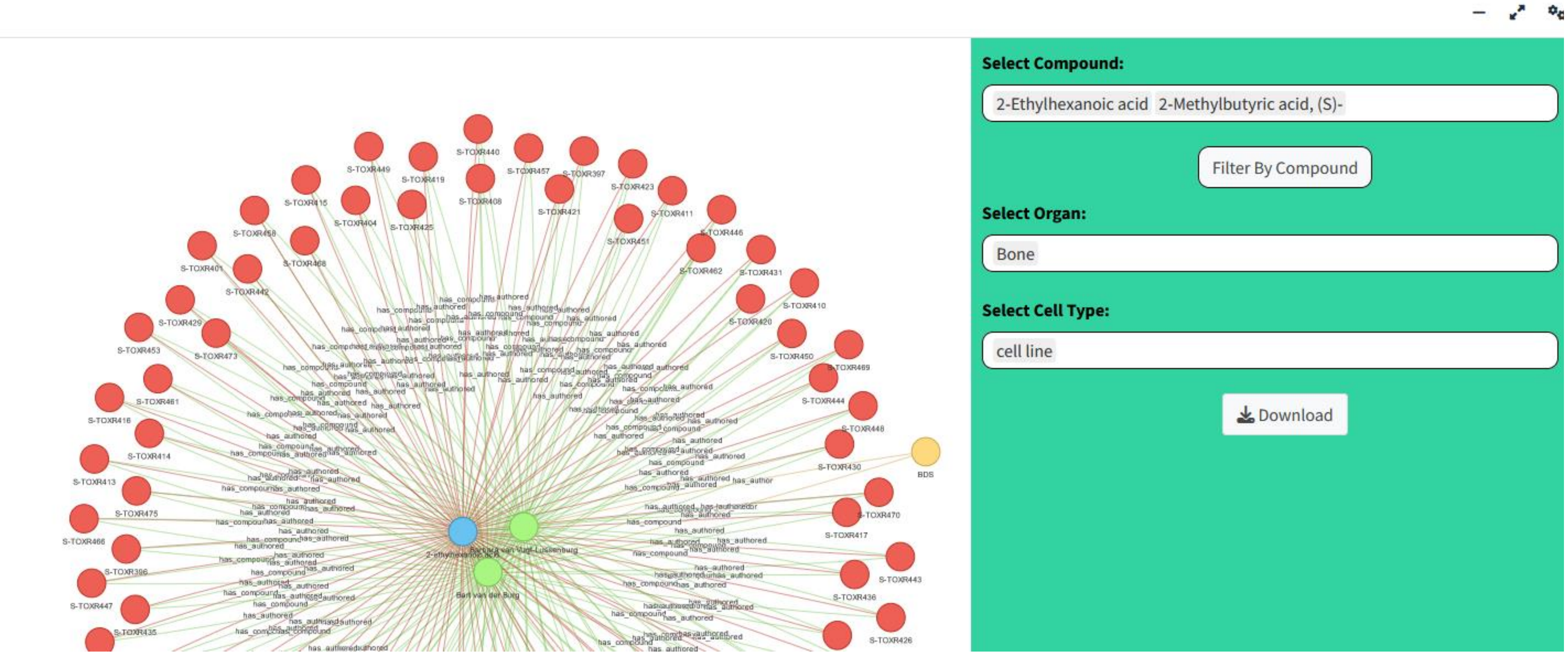
Nextflow workflow
Analysis modules and result reports created with notebooks
.nf
.py → .R → .Rmd
.html
Pull container

Packaged as zip archive

links to
Register workflow
Get PID
WorkflowHub
https://doi.org/10.48546/workflowhub.workflow.402.7

Containerized environments
DockerHub

Knowledge Graph (EU-ToxRisk data being extended to ASPIS)



Select Compound:

2-Ethylhexanoic acid 2-Methylbutyric acid, (S)-

Filter By Compound

Select Organ:

Bone

Select Cell Type:

cell line

Download

RISK HUNT3R

Credits:
Barry Hardy (Edelweiss Connect)
Email: Barry.Hardy@edelweissconnect.com
LinkedIn: https://www.linkedin.com/in/barryhardy/

Acknowledgments

Pr. John Colbourne (UoB)

Dr Ralf Weber, Dr Martin Jones, Marianne Barnard (UoB)

D Batista (PhD candidate) (UOXF)

Dr Juan Carlos Gonzales-Sanchez, Prof Rob Russel (HEI)

Dr Nadine Cistiakova (MGI)

Dr Anna Vogt (CRG)

All our colleagues from PrecisionTox

Dr. Barry Hardy (CEO Edelweiss Connect)

All our colleagues from RISK-HUNT3R

Dr. Danyel Jennen

All our colleagues from Ontox