



UNIVERSITY OF  
LEICESTER

# A PROPOSED DATA STANDARD FOR BREATH SAMPLE DATA AND METADATA

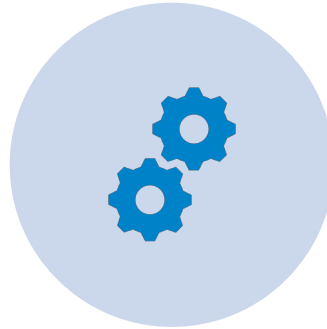
BO ZHAO

---

# BACKGROUND



Metabolomics data  
standards



Reproducibility



Quality assurance



# POPULAR METABOLOMICS DATA STANDARDS



## ISA Tools

ISA-Tab  
ISA JSON



## Metabolomics Workbench

mwTab



## HUPO Proteomics Standards Initiative

mzTab  
mzTab-M



# A GOOD DATA STANDARD



## Readability

- Human readable
- Machine processable

## Usability

- Ease of use

## Extendibility

- Configurable design



# EXTENDING ISA



## ISA data model

Diverse data support  
Rich metadata



## Enhanced relationship

Chain of processes



## Quality control

Validated results



# ENHANCED RELATIONSHIP



## Ontology based

- Ontology annotations in ISA data model

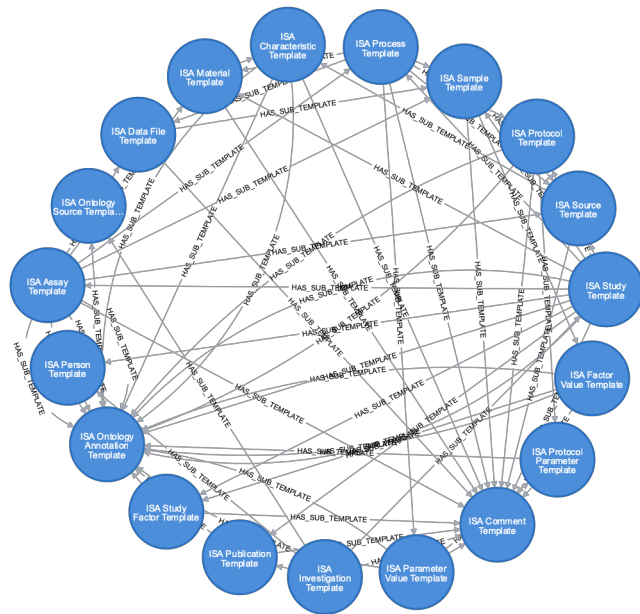
## Pre-defined

- Model-based built-in relationship definitions

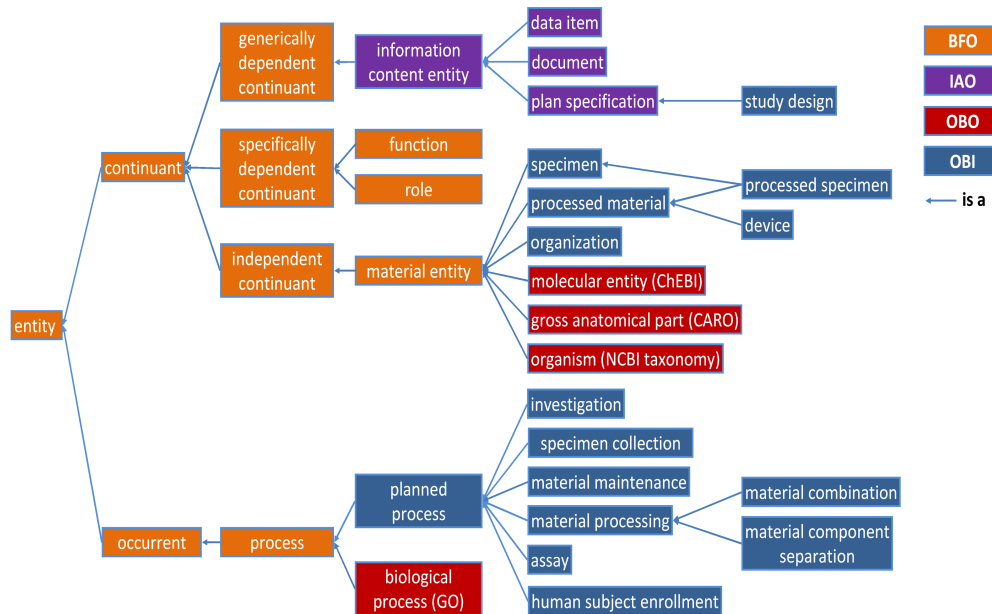
## User defined

- Study-specific configurations

# ENHANCED RELATIONSHIP

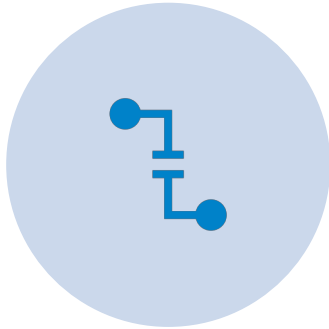


Low-level ISA components relationships



Partial high-level structure of OBI classes (*Bandrowski et al. 2016*)

# ENHANCED QUALITY CONTROL



Extensively applied



Customisable approaches  
(in development)



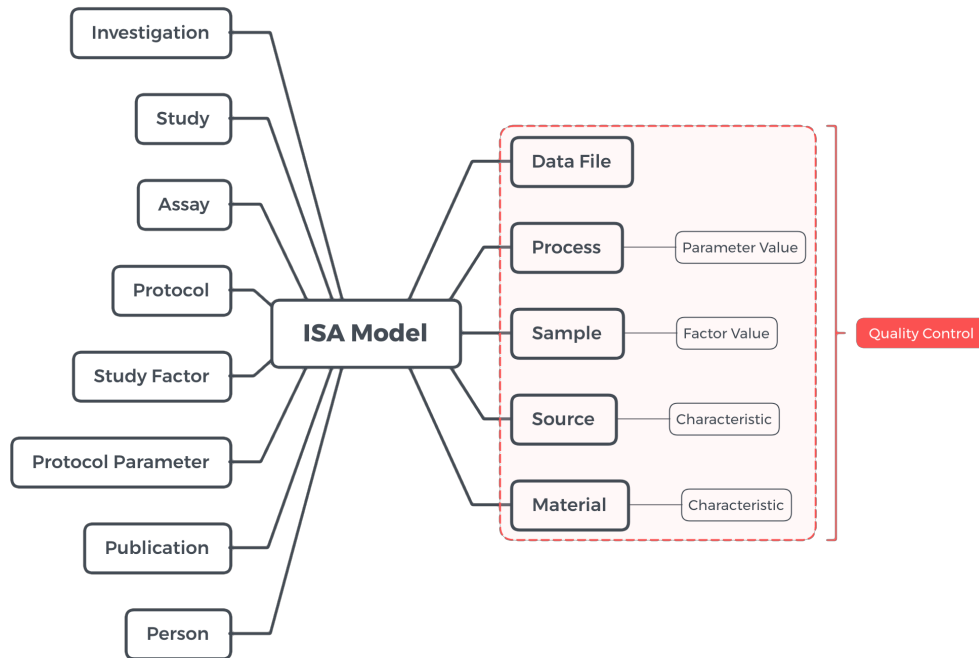
Built-in support for basic  
statistical validations



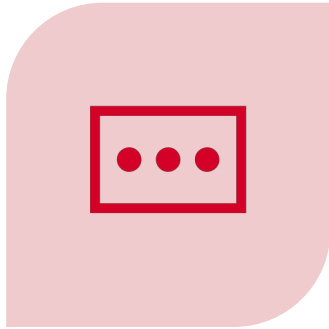


# ENHANCED QUALITY CONTROL

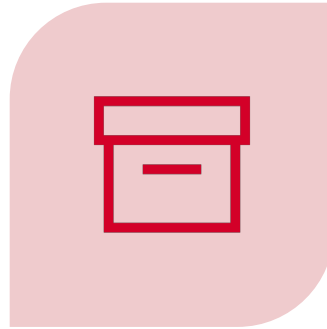
Component level verification  
Configurable in study design  
Quality control result file



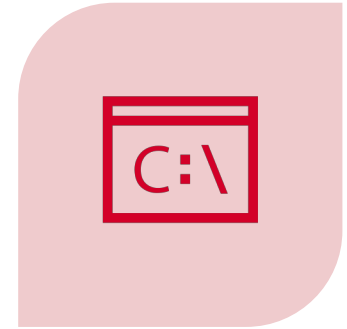
# DATA MANAGEMENT TOOL



Local installation or on a  
server



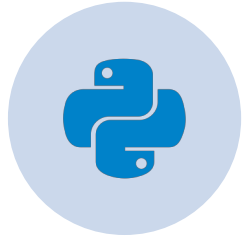
Compatible with original ISA  
archives



Command line and user-  
friendly interfaces



# TOOL FEATURES



Written in python



Easy to install



Easy to embed



Guided access



# USER INTERFACE – TEMPLATE BROWSER



Type	Description	Used in Records	Child Templates	Actions
>	ISA Factor Value Template	0	4	<a href="#">BROWSE RECORDS</a> <a href="#">CREATE RECORD</a> <a href="#">EXTEND TEMPLATE</a>
∨	ISA Ontology Annotation Template	1	1	<a href="#">BROWSE RECORDS</a> <a href="#">CREATE RECORD</a> <a href="#">EXTEND TEMPLATE</a>

Child templates		
Merge field	Child template	Multiple instances
term_source	ISA Ontology Source Template	<span style="color: red;">✘ No</span>



# USER INTERFACE – GUIDED RECORD CREATION

1. Template
2. Metadata
3. Data
4. Link
5. Review

>	✔	Select template	ISA Ontology Annotation Template
>	✔	Enter metadata	
>	✔	Enter data	✔ Required Fields Completed
>	✔	Link records	term_source 1
>	✔	Finish	

# USER INTERFACE – GUIDED RECORD CREATION

1. **Template**
2. Metadata
3. Data
4. Link
5. Review

▼ ✓ Select template ISA Ontology Annotation Template

**Search for template**

Minimum 3 characters required.

**ISA compatible level**

✓ **Exact**  ✓ **Compatible**  **Any**

**You have selected:**

ISA Ontology Annotation Template

**NEXT**

# USER INTERFACE – GUIDED RECORD CREATION

1. Template
2. Metadata
3. Data
4. Link
5. Review

Enter metadata

**Identifier**  
id\_test  
Required field. Must be unique.

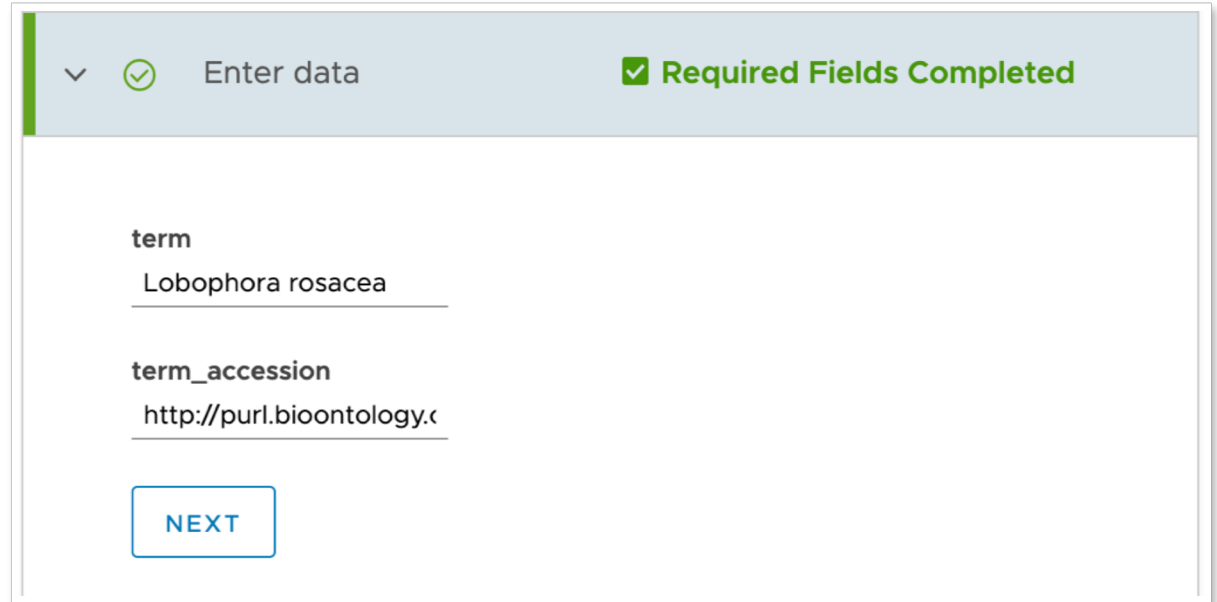
**Display**  
Test Record

**Description**  
This is a test record.

NEXT

# USER INTERFACE – GUIDED RECORD CREATION

1. Template
2. Metadata
3. **Data**
4. Link
5. Review



▼ ✓ Enter data ✓ Required Fields Completed

**term**  
Lobophora rosacea

**term\_accession**  
http://purl.bioontology.c

NEXT



# USER INTERFACE – GUIDED RECORD CREATION

1. Template
2. Metadata
3. Data
4. **Link**
5. Review

The screenshot displays a web interface for creating a guided record. At the top, there is a header bar with a dropdown arrow, a checkmark, the text "Link records", and a green pill-shaped button labeled "term\_source" with a small blue circle containing the number "1". Below the header, the text "term\_source 1" is displayed. There are two buttons: "LINK SELECTED" and "LINK NONE". A table with three columns: "Record", "Description", and "Value" is shown. The table contains two rows of data. At the bottom right of the table area, there is a pagination control showing "items per page 5" and "1 - 2 of 2 items". A "NEXT" button is located at the bottom left of the interface.

Record	Description	Value
<input type="radio"/> OntologySource Node 0	This is an example ontology source node. ID: [0]	{ "description": "OS description 0.", "file": "OS file 0.", "name": "OSName0", "version": "OS version 0." }
<input type="radio"/> OntologySource Node 9	This is an example ontology source node. ID: [9]	{ "description": "OS description 9.", "file": "OS file 9.", "name": "OSName9", "version": "OS version 9." }

# USER INTERFACE – GUIDED RECORD CREATION

1. Template
2. Metadata
3. Data
4. Link
5. Review

Finish

Please review your actions before continue.

**Template**

ISA Ontology Annotation Template

**Metadata**

term	Lobophora rosacea
term_accession	<a href="http://purl.bioontology.org/ontology/NCBITAXON/1638598">http://purl.bioontology.org/ontology/NCBITAXON/1638598</a>

**Records to link**

term_source	1
-------------	---

FINISH

# USER INTERFACE - REPORT



### Study Information

---

**Identifier** EMBER

---

**Name** East Midlands Breathomics Pathology Node

---

**Description** The East Midlands Breathomics Pathology Node (EMBER) is a Medical Research Council (MRC) & Engineering and Physical Sciences Research Council (EPSRC) funded Molecular Pathology Node hosted at the University of Leicester partnering with University Hospitals of Leicester NHS Trust, Loughborough University and Industry.

---

**Publications** [PMID 30852546](#) [PMID 30755317](#)

---

**Contacts** [Chris Brightling](#) [Salman Siddiqui](#)

---

**Submission** 2015-09-01

---

**Release** 2019-09-09

---

### Sample Stats

---

Sources	Specimen Type	Samples	Locations
400	Exhaled air specimen	1000	<a href="#">Glenfield Hospital</a> <a href="#">Leicester Royal Infirmary</a>



# USER INTERFACE - REPORT



## Protocol Information

Only showing protocols related to selected samples

Name	Version	Components	Parameters
ReCIVA Kits Preparation	2	Consumable <b>TD Tube</b> Consumable <b>Mask</b>	Seal Status <b>Yes</b>
ReCIVA Breath Collection	1	Instrument <b>ReCIVA</b> Consumable <b>TD Tube</b> Consumable <b>Mask</b>	Samples <b>1</b> Minimum air flow (L/min) <b>30</b> Minimum collection volume (%) <b>50</b> Maximum collection duration (second) <b>900</b>

## Sample Detail

### EMBER-E0001-R-RP

Processes	Performer	Date	Parameters	Quality Control
ReCIVA Kits Preparation	Bo Zhao	2019-09-09	Seal Status <b>Yes</b>	Seal Check <b>Pass</b>
ReCIVA Breath Collection	Jane Doe	2019-09-09	Samples <b>1</b> Air flow (L/min) <b>33</b> Collection volume (%) <b>100</b> Collection duration (second) <b>600</b>	Sample Count <b>Pass</b> Minimum air flow <b>Pass</b> Minimum collection volume <b>Pass</b> Maximum collection duration <b>Pass</b>



# AVAILABILITY



**GitHub**

<https://github.com/rcfgroup>



**Later this year**



# OTHER WORKS

- LabPipe
- Ontology for Breath Sample Process



# ACKNOWLEDGEMENT

- The research is funded by the EMBER project and NIHR Leicester Biomedical Research Centre – Respiratory.

