

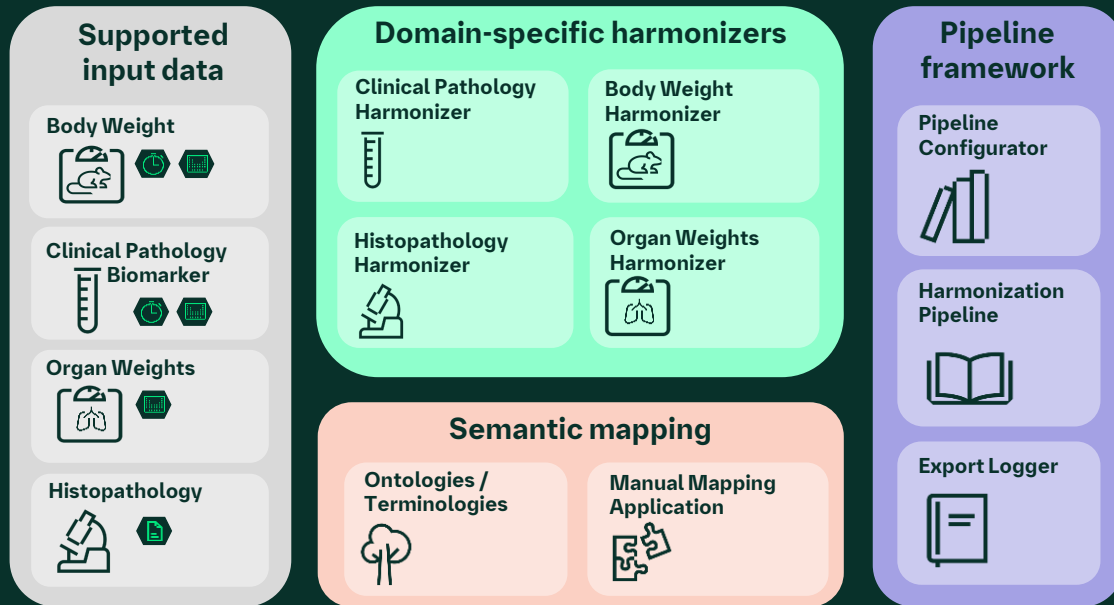
ONITH: An Ontology-based Non-clinical Integration and Term Harmonization Pipeline facilitating Data Reuse in 3R

Lisa-Maria Winter^{*1,2,3}, Johann Kraus², Birgit Stierstorfer¹, Natalie Ketter¹, Katja Hempel¹, Kathrin Stucke-Straub³, Hans A. Kestler², Mario Pepe¹

¹ Boehringer Ingelheim Pharma GmbH & Co. KG

² Ulm University, Institute of Medical Systems Biology

³ Ulm University of Applied Sciences, Institute of Computer Science



Non-Clinical Statistics Conference 2026



universität
uulm



THU
Technische
Hochschule
Ulm



**Boehringer
Ingelheim**

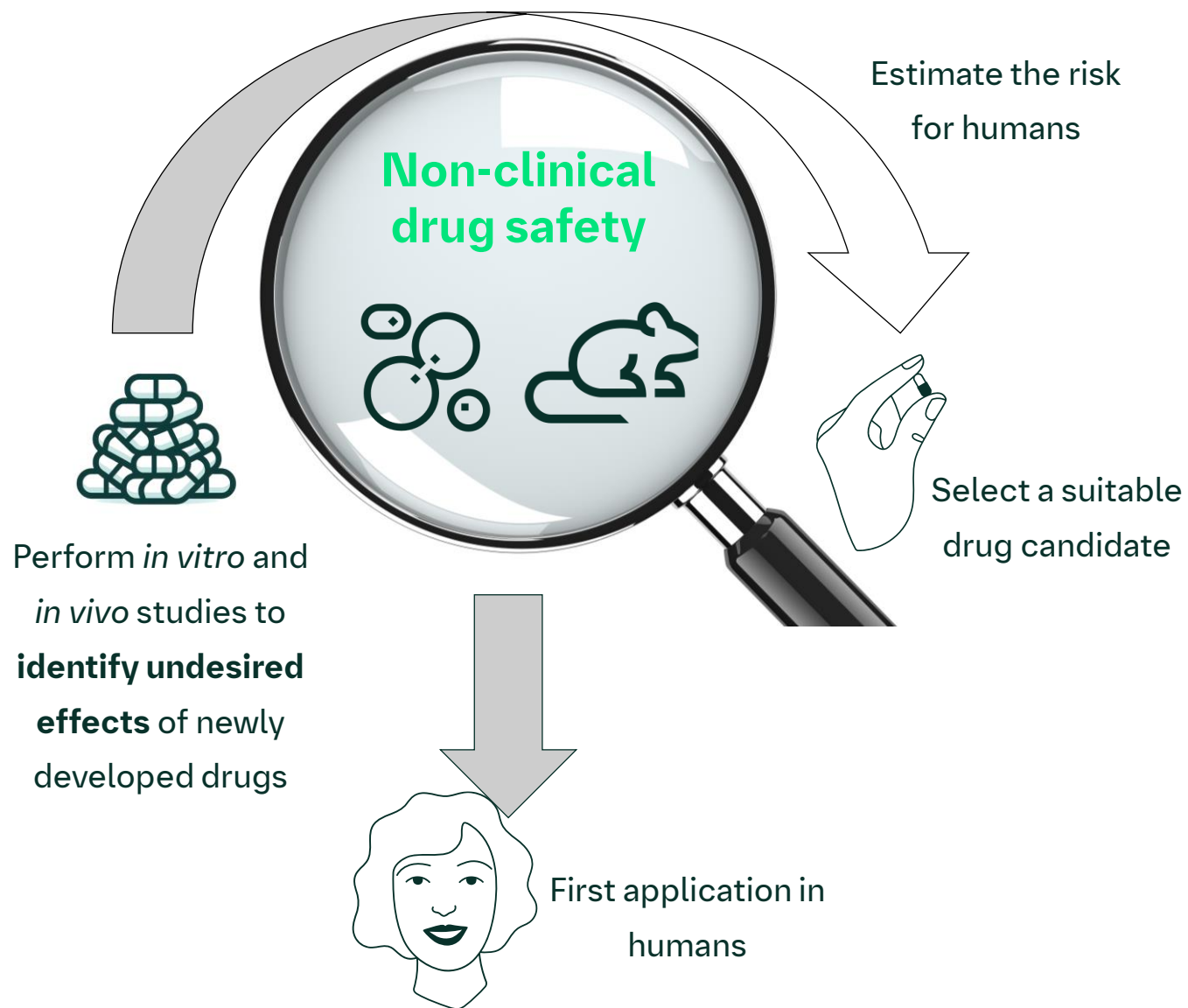
Our Mission

Animal studies continue to play a critical role in non-clinical drug safety

Data-driven approaches are one possibility to reduce animal use^[1]

We want to **facilitate data reuse** and thereby enable **computational modeling on historical non-clinical data** to contribute to 3R

[1] FDA Modernization Act 3.0. 118th Congress. H.R.7248. 2024.



Data Reuse

The Goal

Compiling a large cross-study dataset from historical non-clinical studies

The Challenge

Historical datasets are very heterogenous and often lack comparability.

The Solution

We developed the harmonization toolkit “ONITH” to help us make this data usable

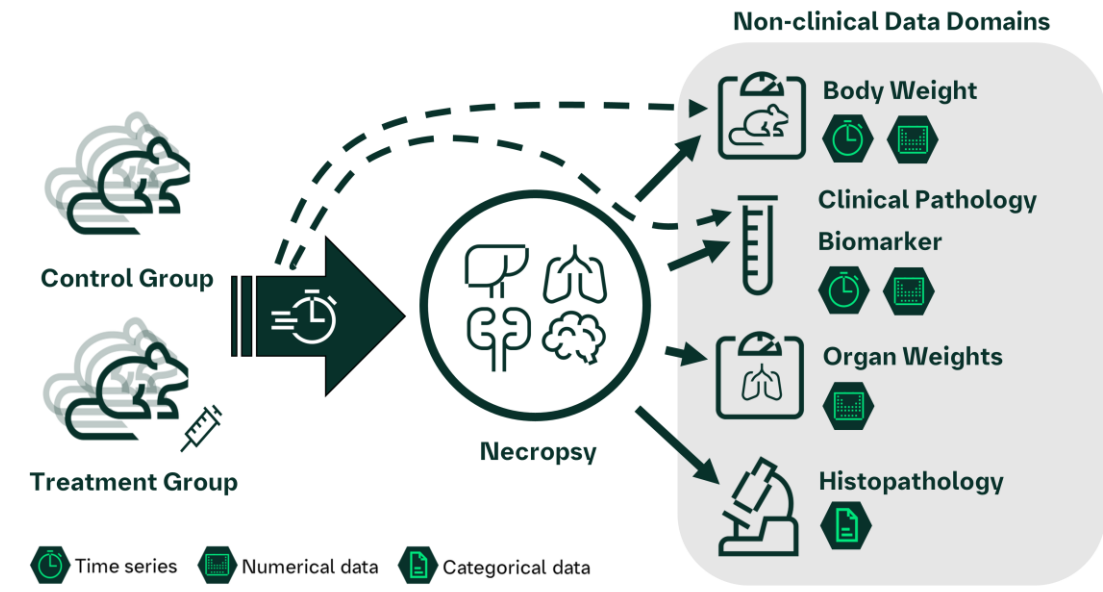


Fig 1: Scheme of a non-clinical safety study and a subset of data domains that are collected.

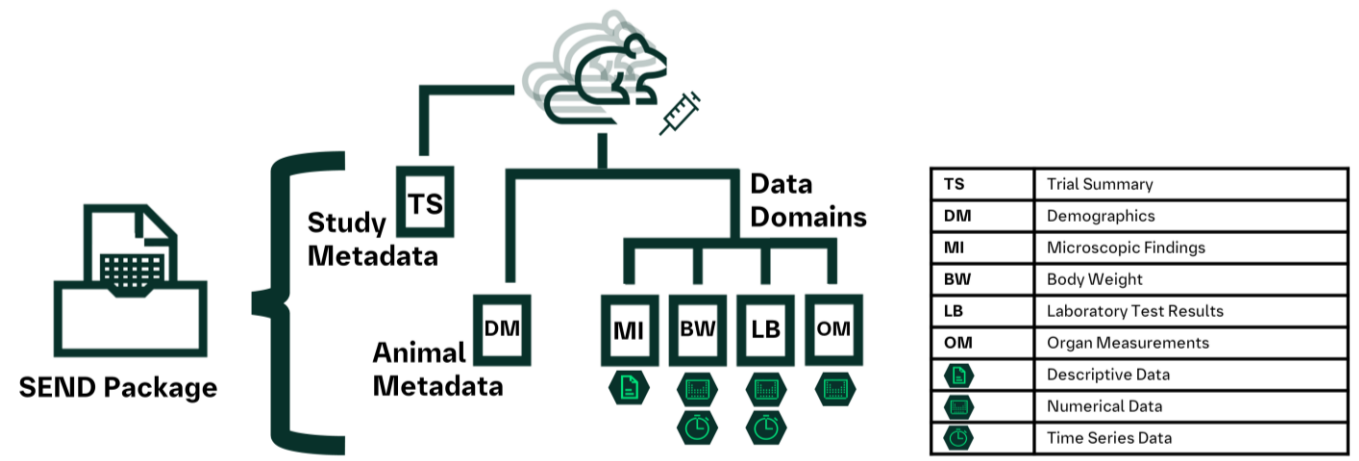
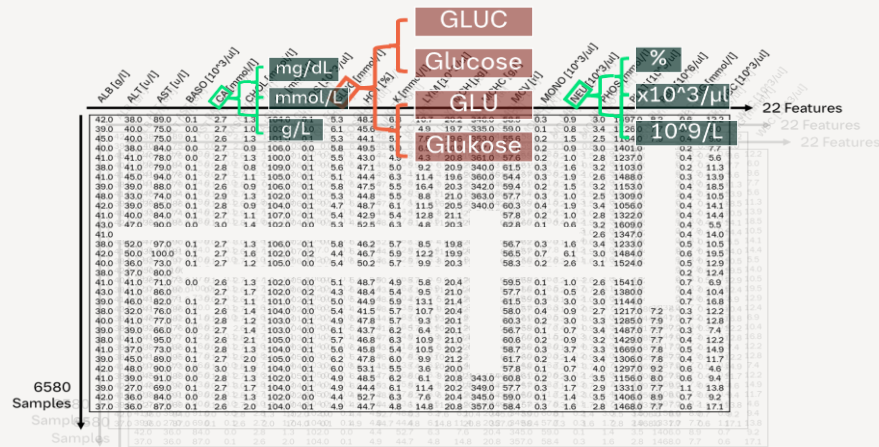


Fig 2: Architecture of a data package from a non-clinical study (simplified) containing the collected data domains in separate tables with defined format and column names, as well as animals and study metadata.

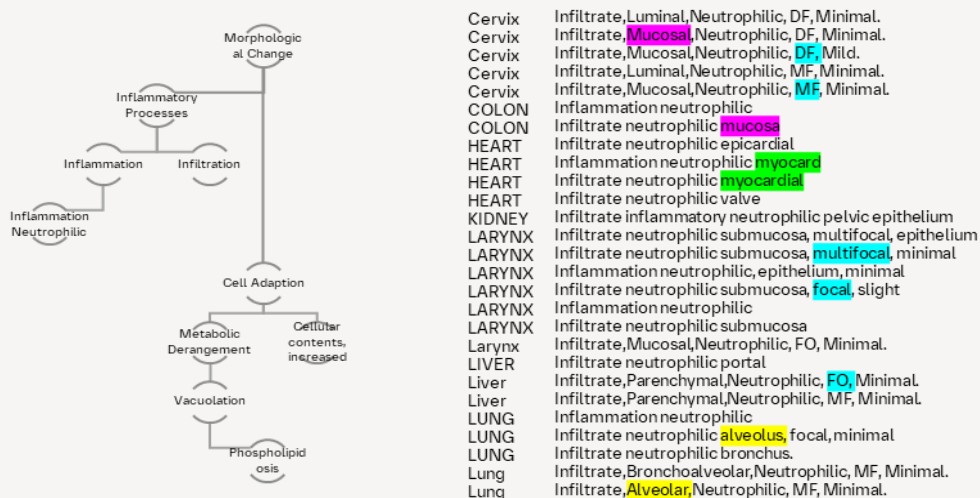
ONITH (Ontology-based Nonclinical Integration & Term Harmonization)

ONITH is published as official python package
-> run ***pip install onith*** to install it

Example 1: Clinical Pathology Biomarker



Example 2: Histopathology Data



Ontology/Terminology

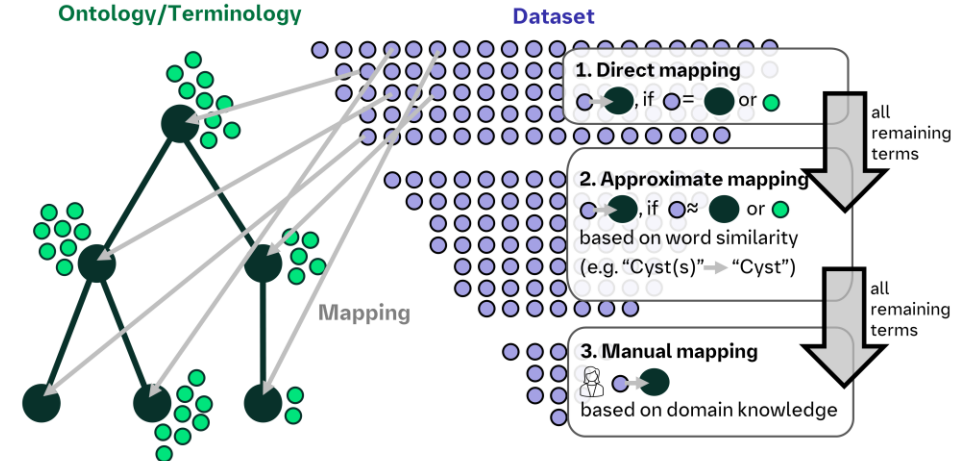
●	Main term
●	Synonym term
●	Synonym set

Dataset

●	Unmapped term
---	---------------

Mapping

↓	Assigning each term of the dataset to a main term of the ontology or terminology
---	--



Ontology-based Term Harmonization Implementation Strategy

Ontologies used in our current implementation of ONITH (selection based on the presence of synonym terms and semantic relationships were necessary):

- CDISC SEND[1] for LB terms (LB)
- INHAND[2] for Organ terms (MI, OM)
- HPATH[3] for Lesion terms (MI)

Users can import their own ontologies for the harmonization in ONITH

[1] CDISC SEND Controlled Terminology. Accessed 2025 Jun 26.

<https://evs.nci.nih.gov/ftp1/CDISC/SEND/SEND%20Terminology.html>.

[2] Keenan CM, Baker J, Bradley A, et al. International Harmonization of Nomenclature and Diagnostic Criteria (INHAND): Progress to Date and Future Plans. Toxicol Pathol. 43(5): 730–732, 2015.

[3] Piaia A. Hpath. 2016 Accessed 2025 Jun 26. <https://reni.item.fraunhofer.de/hpath/index.htm>.

Currently Under Revision at „Computer Methods and Programs in Biomedicine“

Supported Input Types:

- SEND Packages
- Report derived findings
- Group-wise statistics

Depending on the input type, specific pipeline steps can be omitted or re-ordered.

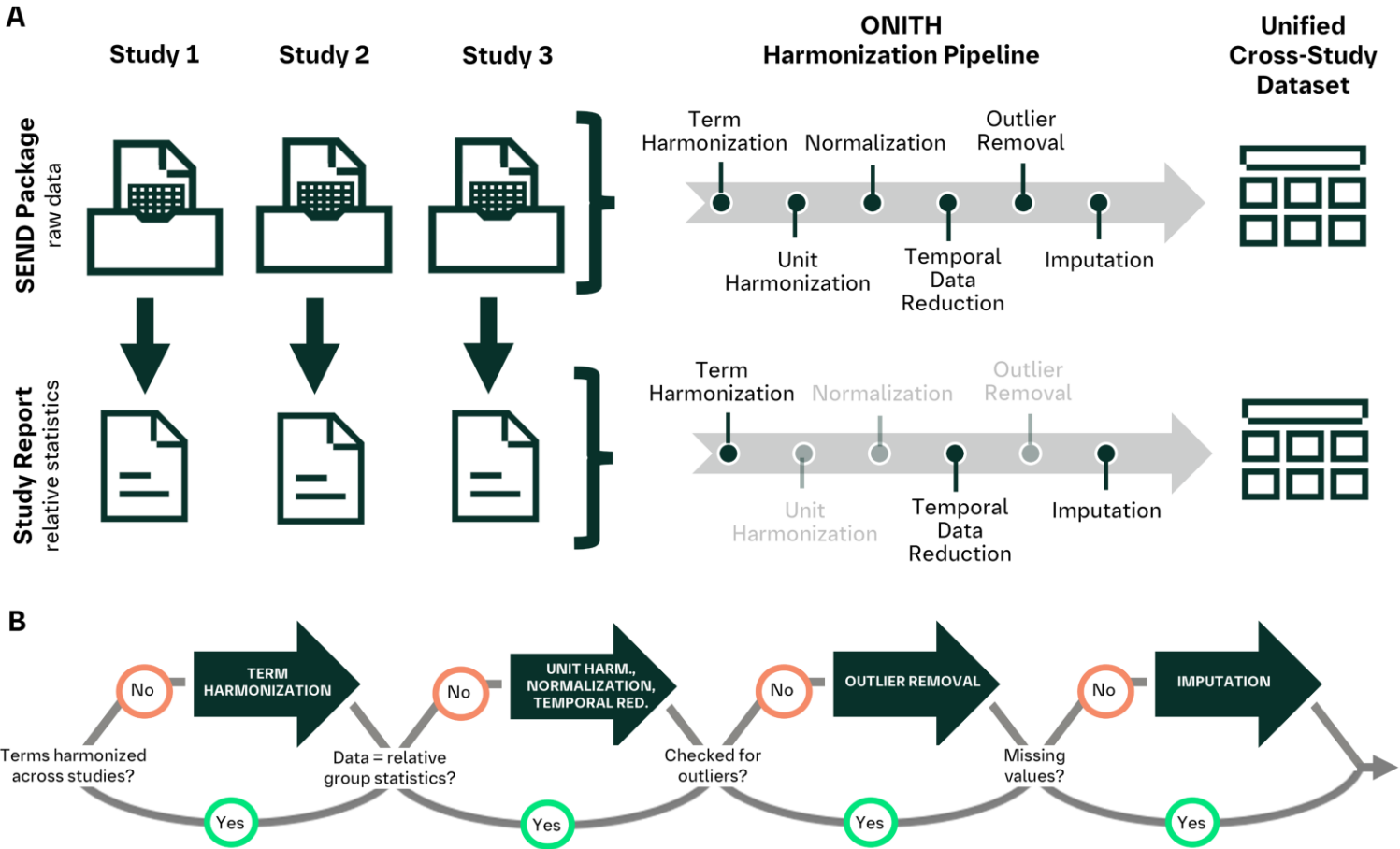
Supported Domains:

- BW (Body Weight)
- OM (Organ Measurements)
- MI (Histopathology)
- LB (Clinical Pathology Biomarker)

Each domain has its own dedicated harmonization pipeline.



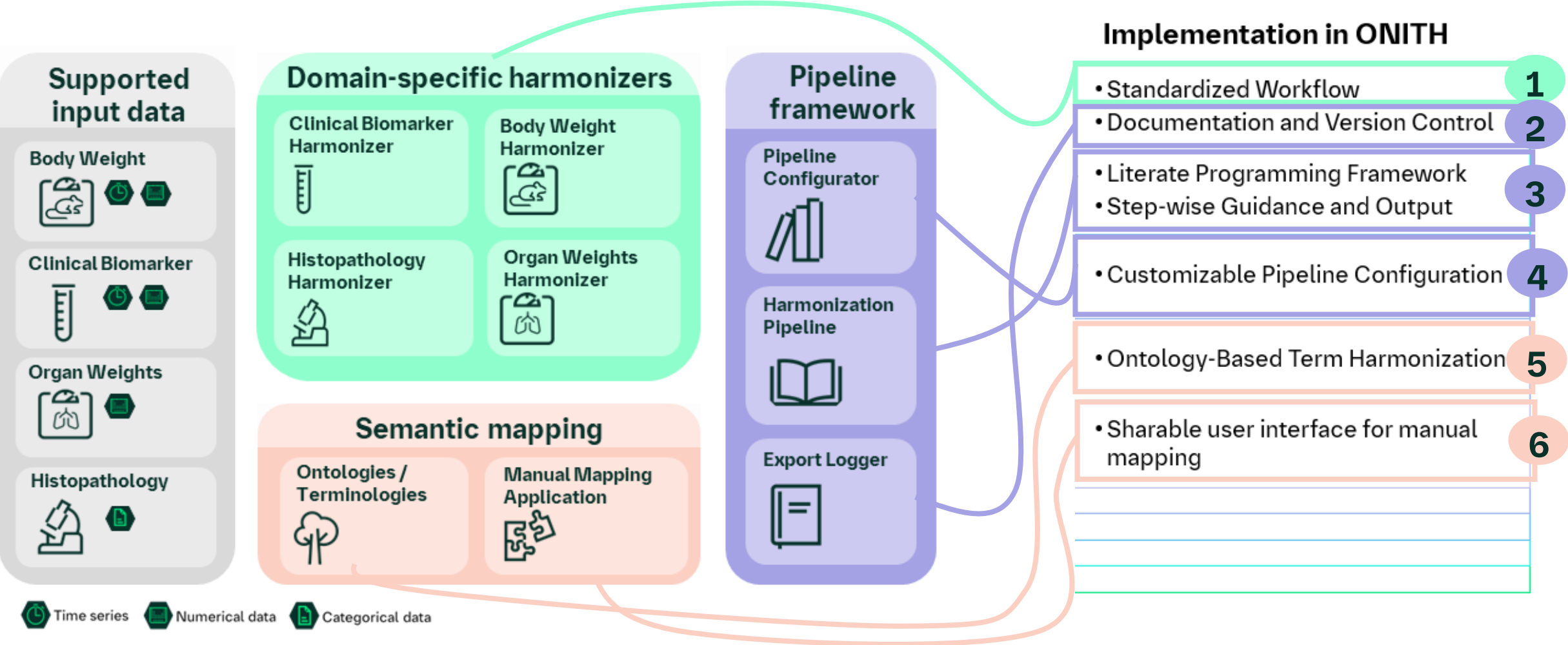
Human-in-the-loop Harmonization Pipeline Design



Key concept to enable FAIR cross-study integration and harmonization

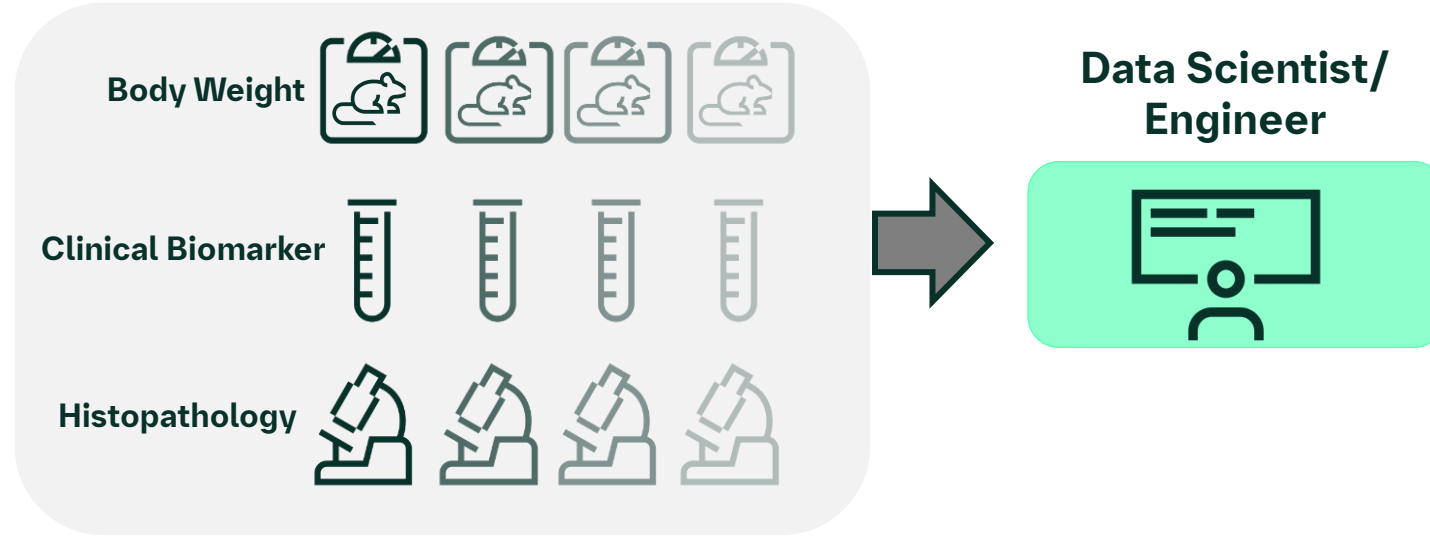
	Key Concept	Implementation in ONITH
	Reproducibility	• Standardized Workflow • Documentation and Version Control
	Interpretability	• Literate Programming Framework • Step-wise Guidance and Output
	Adaptable Input Handling	• Customizable Pipeline Configuration
	Preservation of Semantic Relationships	• Ontology-Based Term Harmonization
	Integration of Expert Knowledge	• Sharable user interface for manual mapping

Modular Package Architecture



Example Use Case

10 studies, 3 domains



The Scenario:

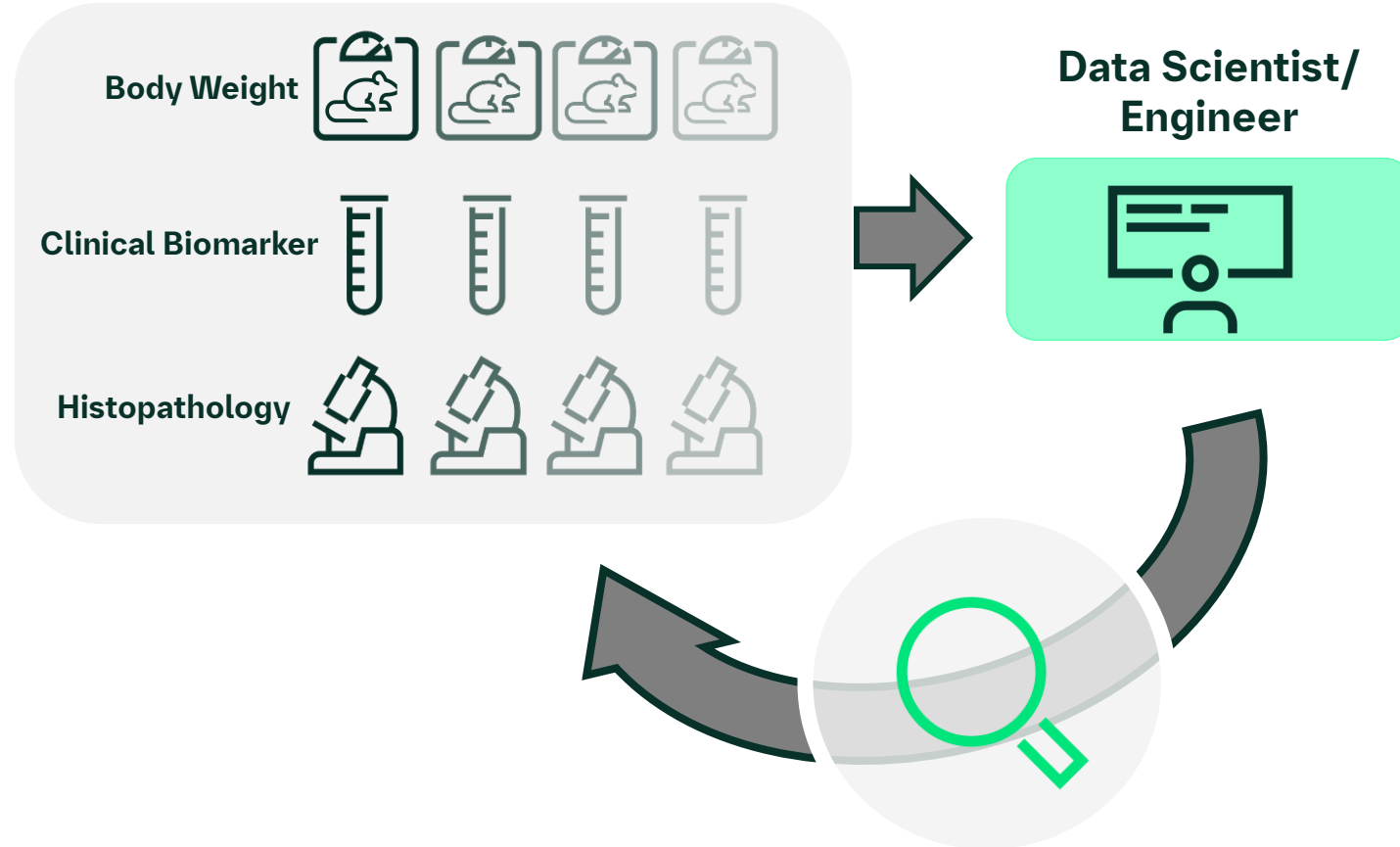
A data scientist / engineer gets a new project assignment:
Computational modeling on historical non-clinical data.

His dataset:

10 internal historical non-clinical studies
3 domains of interest for particular project

Example Use Case

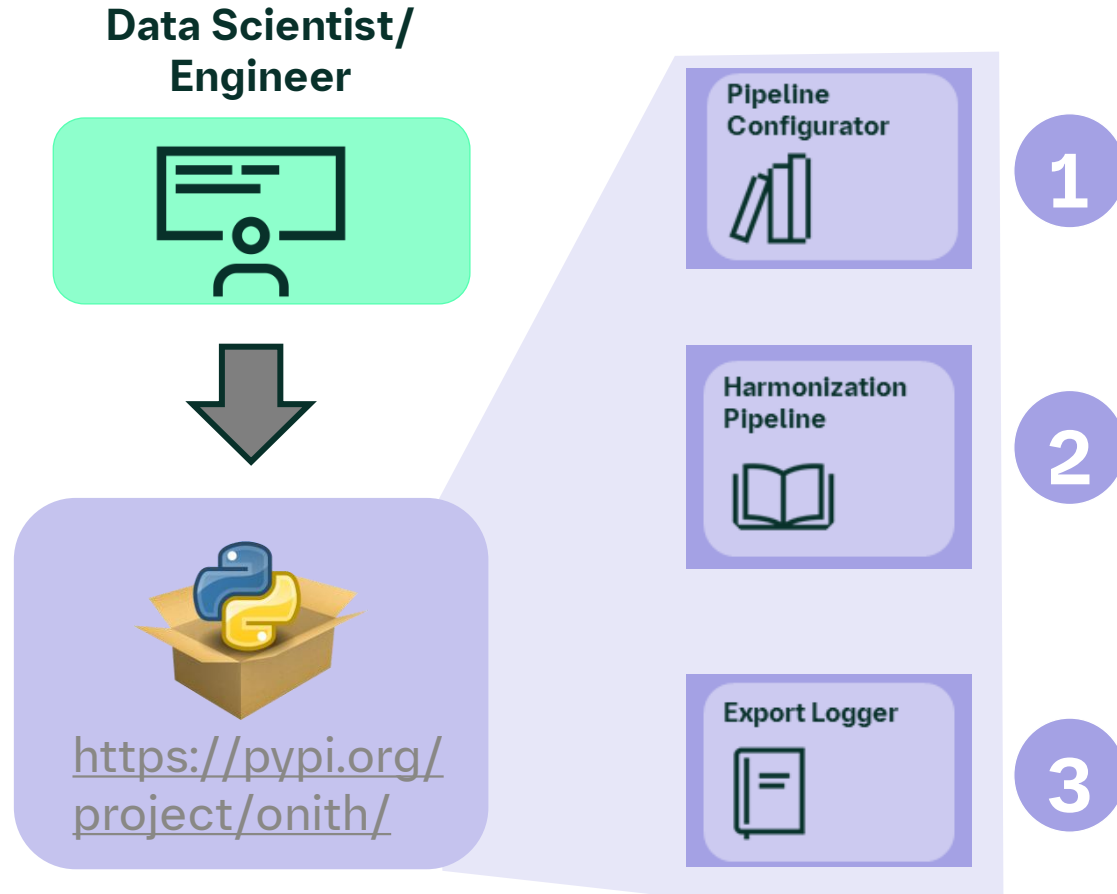
10 studies, 3 domains



The data scientist / engineer explores his dataset and finds:

- The studies come from different study sites, different devices and units were used to quantify clinical pathology biomarkers
- The study designs vary in their measurement timepoints: (single / multiple time point measurements)
- The organ panel for histopathology differs between studies
- The studies come from different years, during which the reference terminology for histopathological findings was updated several times

Example Use Case



The data scientist / engineer decides to use ONITH to facilitate his harmonization process.

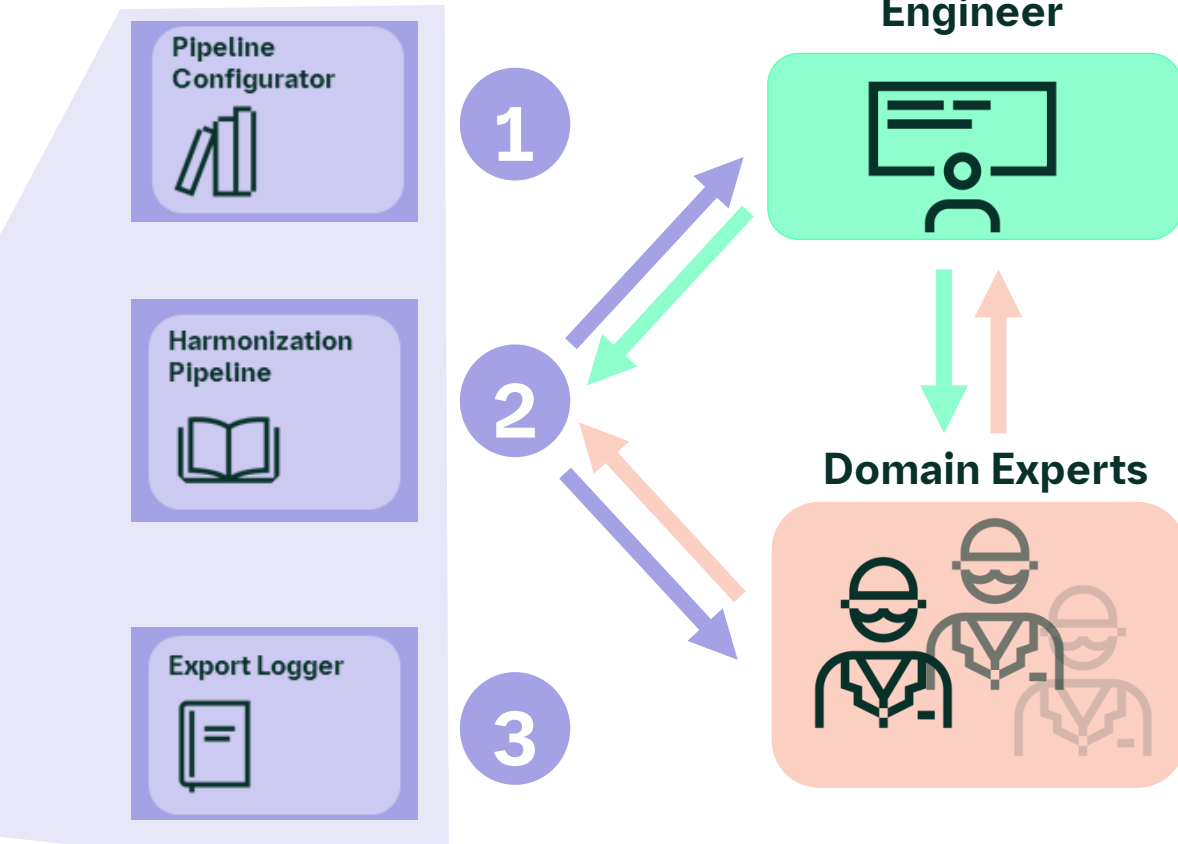
ONITH guides him through every step:

He has to configure his own customized harmonization pipeline,

then work through the all domain-specific harmonization steps within the pipeline, while integrating expertise from domain experts whenever necessary,

and let ONITH document everything he did automatically and store it together with his harmonized dataset.

Example Use Case



The data scientist / engineer decides to use **ONITH** to facilitate his harmonization process.

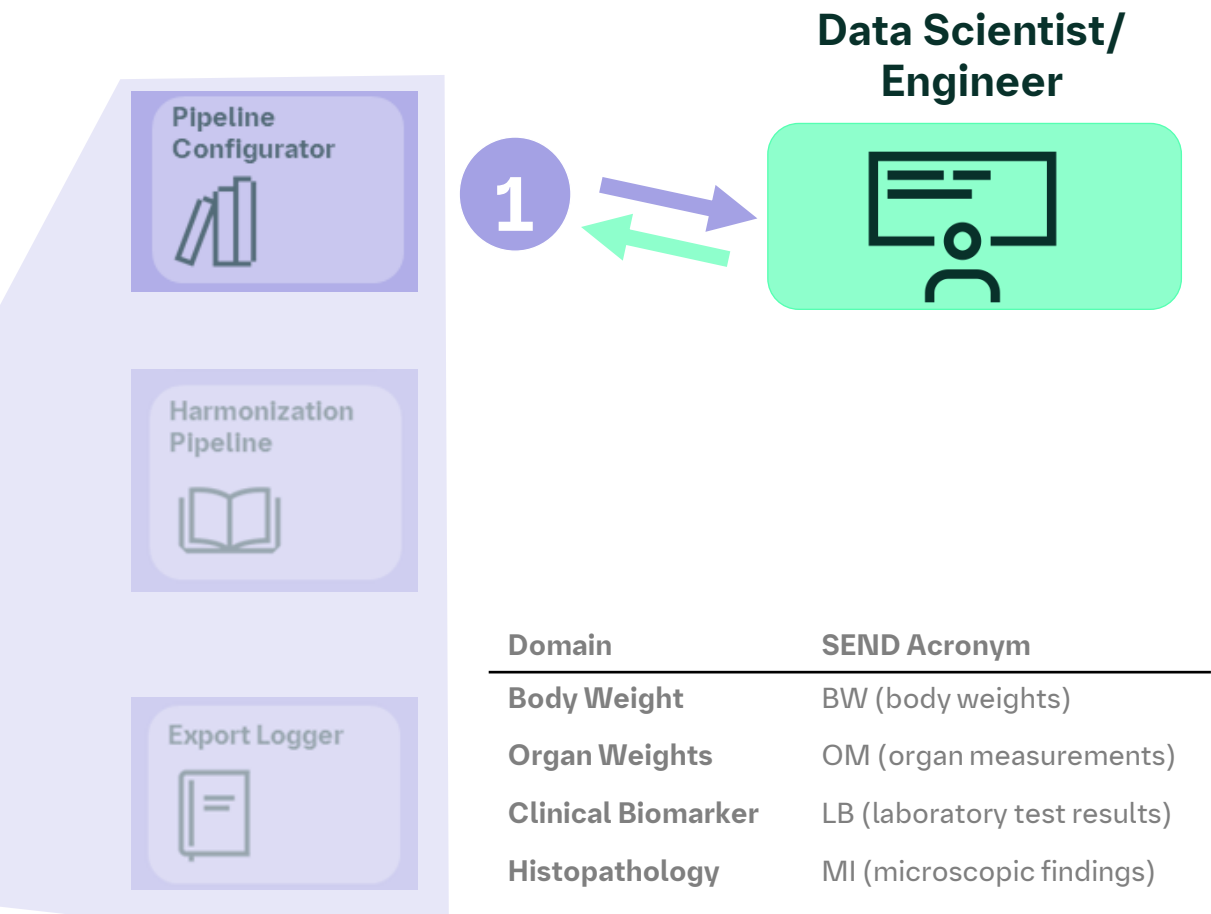
ONITH guides him through every step:

He has to configure his own customized harmonization pipeline,

then work through the all domain-specific harmonization steps within the pipeline, while integrating expertise from domain experts whenever necessary,

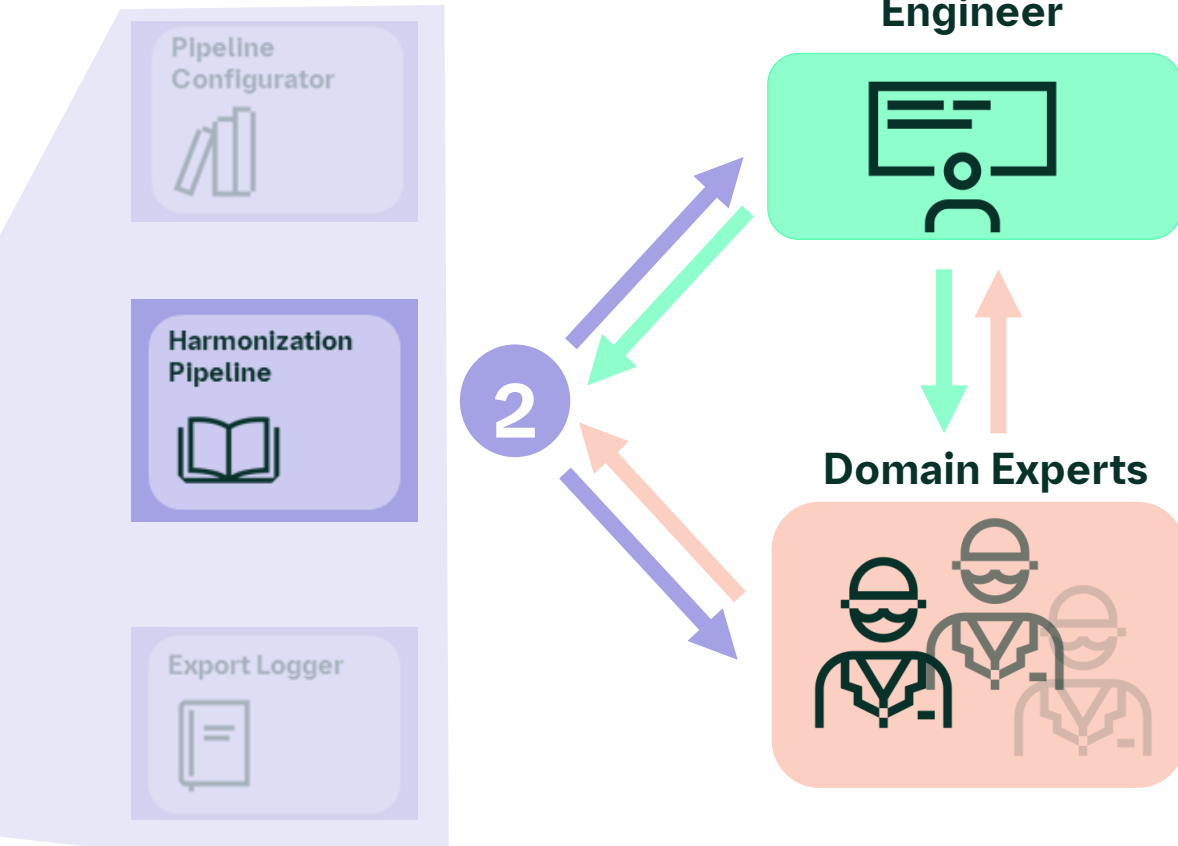
and let ONITH document everything he did automatically and store it together with his harmonized dataset.

Example Use Case



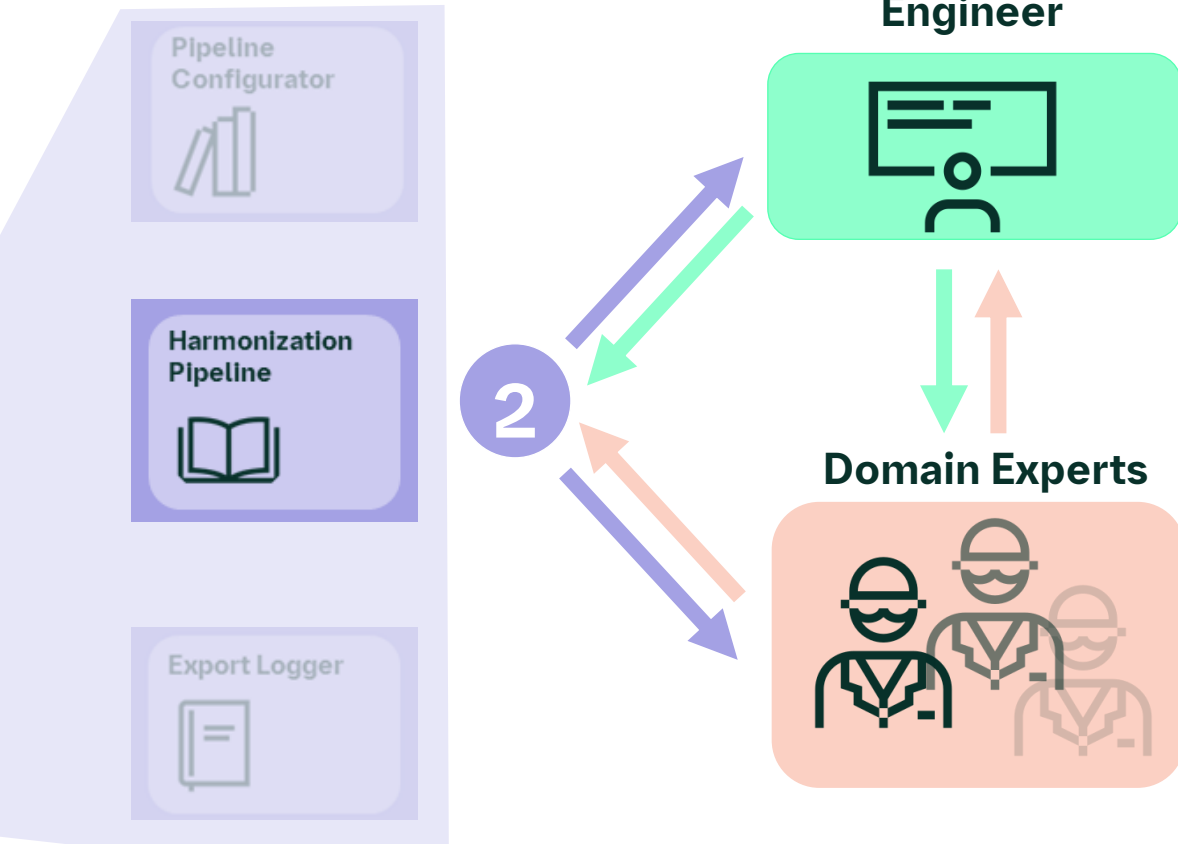
```
example_pipeline_config.ipynb •
C: > Temp > example_pipeline_config.ipynb > from onith import *
+ Code + Markdown | ▶ Run All ↺ Restart ≡ Clear All Outputs ... Python 3.12.2
```

Example Use Case



- The pipeline is implemented as Jupyter Notebook and customized to the domains included
- Domain-specific instructions are already included
- The necessary functions are already prepared in the right execution order
- Outputs are automatically generated to inform user decisions

Example Use Case



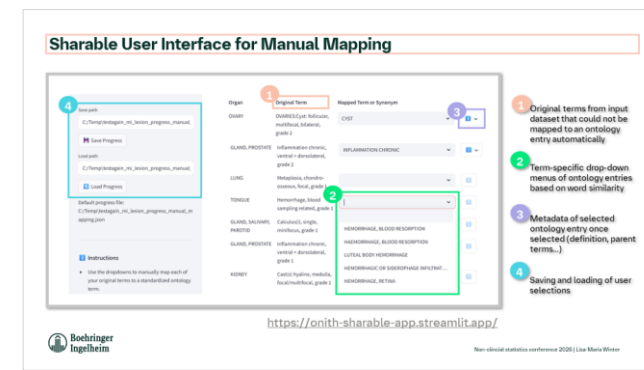
Example for expert knowledge integration:



Ontology-based Term Harmonization

- 1) Fully automatic mapping based on synonym terms
- 2) Fuzzy matching (based on word similarity) $\geq 95\%$
- 3) Manual mapping of remaining terms

-> Sharable manual mapping application:



Sharable User Interface for Manual Mapping

4 Save path
C:/Temp/testagain_mi_lesion_progress_manual.
Save Progress
Load path
C:/Temp/testagain_mi_lesion_progress_manual.
Load Progress
Default progress file:
C:/Temp/testagain_mi_lesion_progress_manual_mapping.json
Instructions
• Use the dropdowns to manually map each of your original terms to a standardized ontology term.

Organ	Original Term	Mapped Term or Synonym
OVARY	OVARIES:Cyst: follicular, multifocal, bilateral, grade 2	CYST
GLAND, PROSTATE	Inflammation chronic, ventral + dorsolateral, grade 2	INFLAMMATION CHRONIC
LUNG	Metaplasia, chondro-osseous, focal, grade 1	
TONGUE	Hemorrhage, blood sampling related, grade 1	
GLAND, SALIVARY, PAROTID	Calculus(i), single, minifocus, grade 1	
GLAND, PROSTATE	Inflammation chronic, ventral + dorsolateral, grade 1	
KIDNEY	Cast(s) hyaline, medulla, focal/multifocal, grade 1	

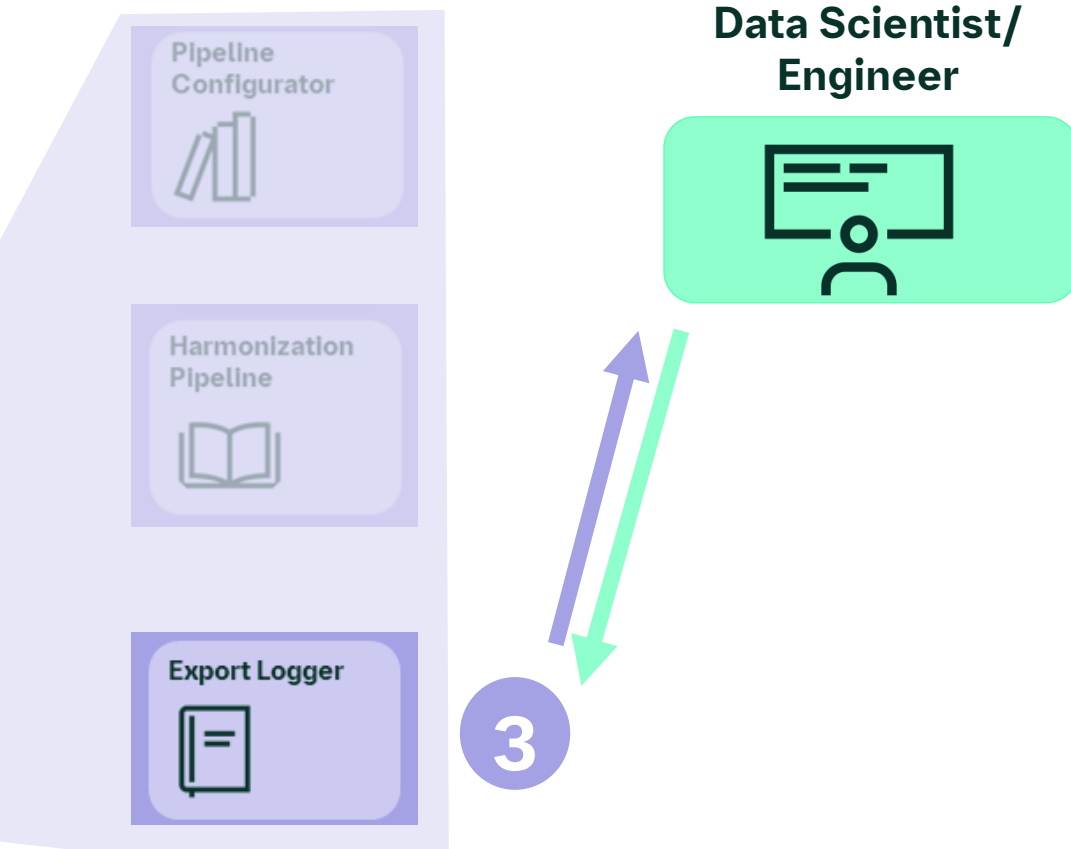
2 Hemorrhage, blood sampling related, grade 1
HEMORRHAGE, BLOOD RESORPTION
HAEMORRHAGE, BLOOD RESORPTION
LUTEAL BODY HEMORRHAGE
HEMORRHAGIC OR SIDEROPHAGE INFILTRAT...
HEMORRHAGE, RETINA

3 Metadata of selected ontology entry once selected (definition, parent terms...)

- 1 Original terms from input dataset that could not be mapped to an ontology entry automatically
- 2 Term-specific drop-down menus of ontology entries based on word similarity
- 3 Metadata of selected ontology entry once selected (definition, parent terms...)
- 4 Saving and loading of user selections

<https://onith-sharable-app.streamlit.app/>

Example Use Case



Documentation upon export of the harmonized dataset

- All user decisions (html export of the Jupyter Notebook)
- Environment status (package versions) etc.
- Dictionary Status

test_project_mi_output_20250618_110216	6/18/2025 11:02 AM
dict_snapshots	6/18/2025 11:02 AM
test_project_mi_harmonization_output.csv	6/18/2025 11:02 AM
test_project_mi_metadata.txt	6/18/2025 11:02 AM
test_project_mi_notebook_snapshot.html	6/18/2025 11:02 AM

Version Control

- Manually mapped terms are retracable within the local ontology copy

```
[Term]
id: OD:1003
name: Adrenal gland
is_a: 0102 ! Endocrine System
synonym: Endocrine System, ADRENAL GLAND
synonym: "GLAND, ADRENAL" MANUAL [2025-05-30]
synonym: "GLND, ADRENAL" MANUAL [2025-06-10]
```

Using ONITH on a Reference Dataset

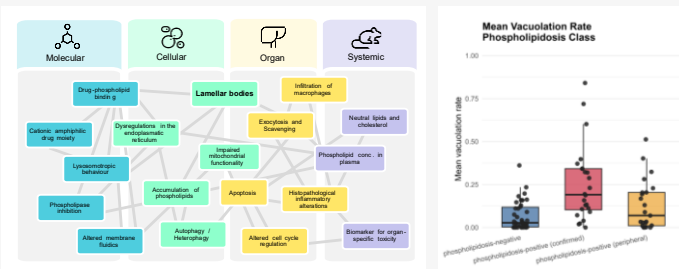
Domain-specific data reduction and harmonization outcomes for SEND datasets



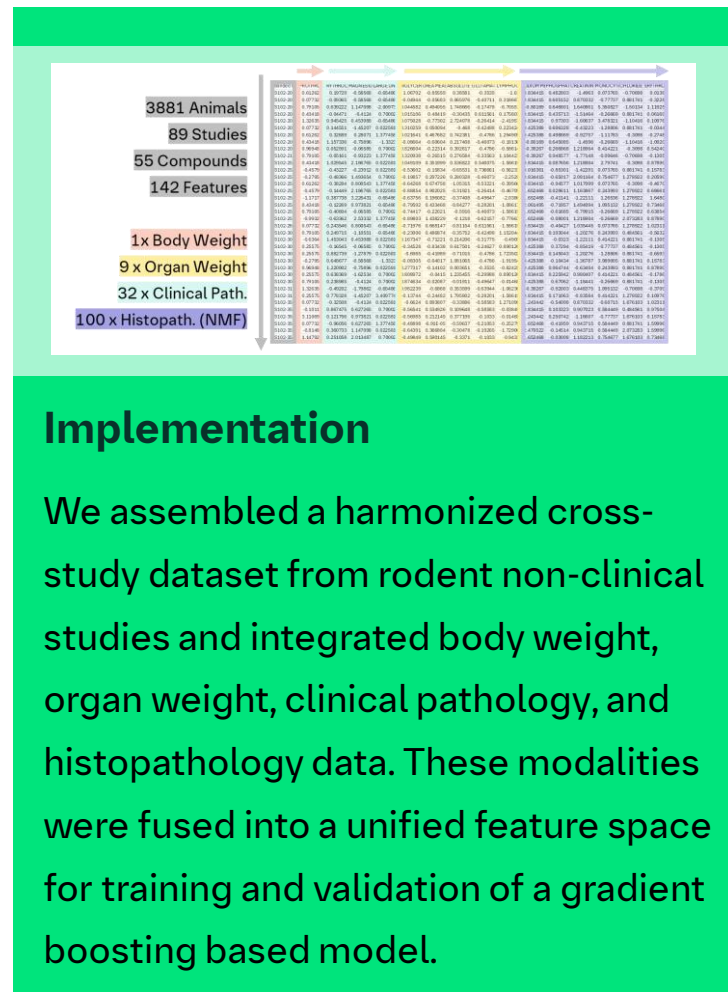
Use Case 1: Multi-modal Data Fusion Using Historical Non-Clinical Data for Predicting Drug-Induced Phospholipidosis In Vivo

Challenge

Phospholipidosis (PL) is a heterogeneous and mechanistically complex form of vacuolation. Definitive confirmation requires electron microscopy.

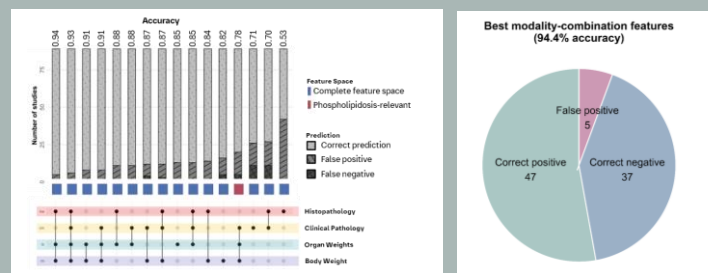


What if we can predict PL from routinely collected non-clinical parameters?



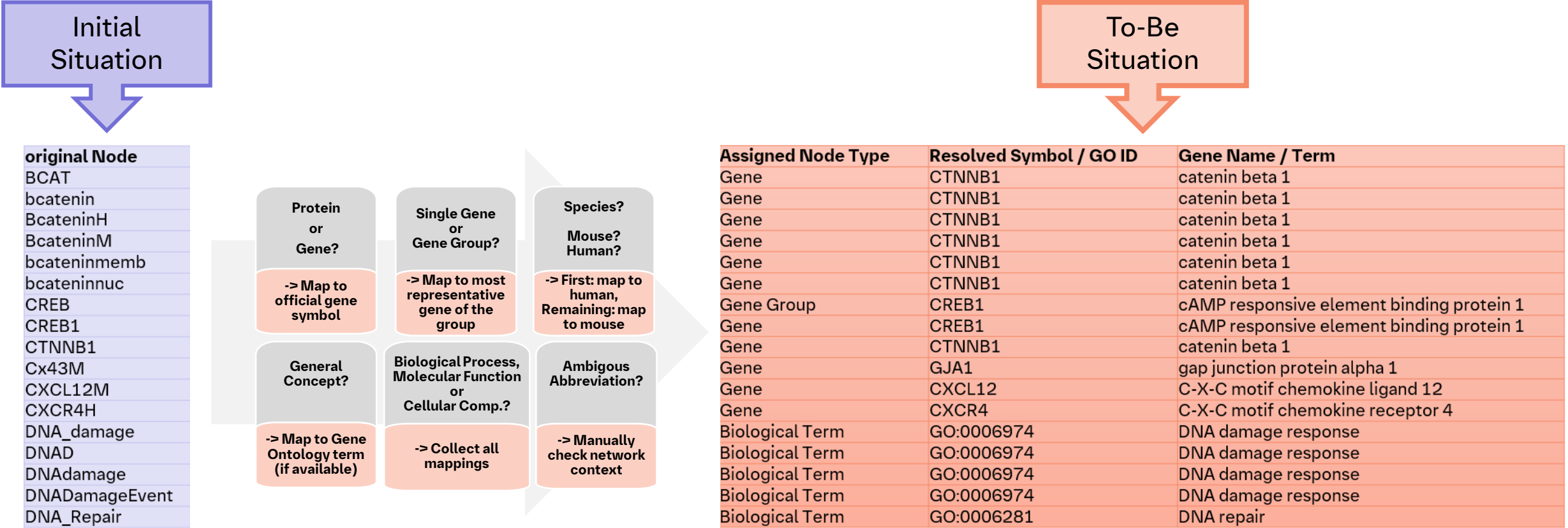
Outcome

Routinely collected non-clinical parameters contain meaningful PL-related signal. The model achieved 94% accuracy on unseen studies and 88% on unseen compounds with the full feature space retrieved from multimodal data fusion outperforming individual modalities.



Use Case 2: Repurposing ONITH for the Mapping and Harmonization of a Boolean Network Repository

Nensi Ikonomi^{1, †}, Lisa-Maria Winter, Natalie Ketter¹, Mario A.A. Pepe^{1*}





universität
uulm



**Boehringer
Ingelheim**



THU
Technische
Hochschule
Ulm

Thank you for your attention!

Questions?

Boehringer Ingelheim

- Dr. Mario Pepe
- Dr. Birgit Stierstorfer
- Dr. Natalie Ketter
- Dr. Katja Hempel

Universität Ulm

Institute for Medical Systems Biology

- Prof. Dr. Hans Kestler
- Dr. Johann Kraus

Technische Hochschule Ulm

Faculty for Computer Science

- Prof. Dr. Kathrin Stucke-Straub

Backup Slides



HPATH



Ontology-based Term Harmonization

HPATH

- 13 companies
 - 9 species
 - 6000+ reports from 40 years
 - **31000+ terms**
 - Cross-referenced to INHAND and CDISC SEND
 - Synonyms
 - Definitions
 - Hierarchical tree structure
- > **Suitable to preserve semantic relationships between lesion terms**

