



Designed by Charlie Barclay, Emma Griffiths and Rhiannon Cameron as part of PHA4GECon 2025 pre-conference workshops

## PHA4GECon 2025: Developing Data Standard Workshop Agenda

### Overview

This hands-on workshop introduces the end-to-end process of data standard development, from defining scope and needs, to selecting ontology terms, to implementing a working schema. Participants will use an example scenario and curated templates to practice designing structured, interoperable data specifications.

**Audience:** Researchers, data managers, and public health professionals involved in developing or implementing standards for infectious disease response.

**Format:** Short lectures, guided exercises, group discussions using scenarios grounded in real-world experience and reusable templates to facilitate your data standards journey.

### Agenda

Session	Duration	Objective	Focus	Activity
1 Identifying Need: Why Standards Matter	08:30 –10:30 2 hrs	Why standards exist; what makes a good one.	Scope, audience, ethics, FAIR.	Scoping & needs assessment (boundaries, audiences, interviews).
2 Standardising Your Terms	10:45 –12:30 1 hr 45 min	Identify and select ontology-based terms.	Ontologies vs standards, LinkML view, OLS demo.	Term curation for environmental/anatomical fields.
3 Implementation & Testing, Testing, Testing	13:30 –15:30 2 hrs	Apply and test curated data in tools.	Tooling (DataHarmonizer), validation, feedback loops.	Curation & feedback on schema and validation issues.

# Plan

## Session 1 – Identifying the need: Why standards matter (2 hours)

**Objective:** Understand why standards exist, the conditions that shape them, and what makes a good one

**Lecture focus:**

- Motivations: data needs, ethical and practical drivers.
- What makes a “good” standard vs. purely technical compliance.
- Case examples:
  - Ontology alignment and FAIR data (from *Examples of Ontology Use in Public Health* ).
  - The PHA4GE Wastewater specification structure and use cases.

**Activity:** *Scoping & Needs Assessment*

- **Scoping exercise** – identify boundaries, audiences, and similar initiatives.
- **Needs assessment** – interview stakeholder personas.

## Session 2 – Standardising your terms (1 hour 45 minutes)

**Objective:** Learn to identify and select ontology-based terms for a specification.

**Lecture focus:**

- What ontologies are; standards vs ontologies vs tools.
- Examples from CanCOGeN and PHA4GE schemas (LinkML view).
- Demonstrate how term sourcing works (EBI OLS demo).

**Activity:** *Ontology Term Curation*

- Each group sources terms for environmental/anatomical fields in the scenario
- Time permitting, add validation to a DataHarmonizer schema

## Session 3 – Implementation and testing, testing, testing (2 hours)

**Objective:** Apply curated data elements into a data-handling tool and design for feedback.

**Lecture focus:**

- Implementing schema in tooling (DataHarmonizer).
- Validation and feedback loops.
- Promoting adoption, community testing and maintenance (link to *Data Specification Development Process* ).

**Activity:** *Curation & Feedback*

- Populate example data in DataHarmonizer using provided PHA4GE schema.
- Identify validation issues and discuss improvement.