



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

PHA4GECon 2025: Scoping

Defining your audience, data need and setting boundaries

This template is designed to guide defining the purpose and rationale for a new standard by systematically evaluating the requirements and stakeholder need, ensuring the new standard addresses genuine gaps and has a clear well-defined scope.

Scope

1.1: Problem Statement

A novel pathogen, *E. griffithsyi*, has emerged across **multiple regions**, with indications of both **human** and **environmental** transmission. There is currently no widely used data standard to capture contextual information (clinical, epidemiological, and environmental) required to support early genomic surveillance and outbreak investigation.

Existing laboratory systems and surveillance databases capture fragmented contextual data that **lacks integration** across clinical and environmental domains. This has resulted in **inconsistent data sharing** and **delayed response** of this emerging pandemic.

1.2 Requirement for a new standard

Existing standards

- MxS (Genomic Standards Consortium): general but lacks additional information required to facilitate both clinical and environmental sample collection.
- CanCOGeN COVID-19 specification: clinical focus only.
- PHES-ODM: targeted for SARS-CoV-2 pathogen and PCR data.

Identified gaps

- No schema combining human clinical, environmental, and sequencing metadata.
- Limited fields for infection source tracing (e.g., surface, wastewater, vector).

1.3 Defining the boundaries:

In-scope

- Data elements from specimen collection through sequencing.
- Human and environmental samples linked to outbreak or surveillance activities.
- Core contextual data: sampling event, host/exposure details, sequencing metadata, QC.
- Interoperable exports to INSDC BioSample, ENA, and GISAID formats.

Out-of-scope

- Individual clinical health records or patient identifiers.
- Case management, diagnostic treatment outcomes, or contact tracing datasets.
- Analysis results

Stakeholders and Audience

2.1 Define the key stakeholders

Stakeholder Group	Role / Contribution	Data Produced or Used
Public Health Labs	Surveillance, sequencing, variant detection, QC reporting	Sample metadata, sequencing data, QC metrics
Clinical Researchers	Investigate host–pathogen interactions, phylogenetic and epidemiological analyses	Clinical context, sampling strategy, analytical metadata
Environmental Monitoring Teams	Collect and process wastewater and environmental samples	Sampling site metadata, environmental measurements

Preliminary Standards Review

3.1 Existing Standards

MIxS, PHA4GE SARS-CoV-2, CanCOGeN, PHES-ODM

3.2 Identify common data elements

Across standards:

- *sample ID, collection date, geo location, organism/taxon, purpose of sampling, sequencing instrument, QC method.*

3.3 Missing data elements

- Environmental context **and** host exposure linkages within a single record.
- Sample receipt and chain-of-custody metadata
- Versioned metadata for rapid updates during outbreak events.

