2025 EColi Workshop: Curating with the GRDI Contextual Data Specification - Worked Examples

Created August 2025

Background

“One Health” is a collaborative approach that recognizes that the health of people is closely connected to the health of animals and our shared environment. Genomic surveillance using a One Health approach is a powerful tool for understanding and tracking how pathogens affecting human health evolve and spread. By structuring contextual data using community standards such as minimum information checklists and ontologies, this information can be more easily understood and used by both humans and computers, and can be more easily reused for different types of analyses. The GRDI contextual data specification was designed to enable consistent, high-quality capture of sample and isolate information across agencies and projects. Its implementation has been iterative, with multiple rounds of testing and refinement to ensure that the specification remains practical and fit-for-purpose. A key lesson from these efforts is that curators achieve better results when training goes beyond written protocols and includes demonstrations and hands-on practice.

This document provides a set of worked examples that illustrate how the specification can be applied to real-world scenarios. These examples highlight common data harmonisation challenges—such as missing terminology, provenance tracking, or inconsistent vocabulary—and show how the DataHarmonizer tool and the GRDI specification can be used to resolve them. They are intended as a training resource for curators, helping to build capacity across departments and agencies while maintaining data autonomy and quality. By walking through structured examples, curators can see how to align their own data with the specification, apply best practices in curation, and recognize checkpoints that support overall quality control. These worked examples complement the training workshops and guidance materials, providing a practical foundation for consistent and reliable contextual data curation.

Worked Examples

Scenario 1: Food (meat)

A sample of frozen ground beef was obtained from a retail environment on Sept 15 2019. The sample was collected by CFIA (sample ID ABC123) as part of its routine Ontario retail surveillance (sample plan “search-and-destroy2019”) and later shared with PHAC (sample ID B1234-a1). The sample was found to contain Escherichia coli via microbiological isolation and PCR testing (isolate ID badbug444) conducted at PHAC and will be sequenced as part of the GRDI by PHAC.

Sample collection and processing
specimen_collector_sample_ID: ABC123
sample_collected_by: Canadian Food Inspection Agency (CFIA) [GENEPIO:0100552]
alternative_sample_ID: B1234-a1
sample_collector_contact_email: meatlab@cfia.ca
sample_plan_name: search-and-destroy2019
purpose_of_sampling: Surveillance [GENEPIO:0100004]
sample_collection_date: 2019-09-15
sample_collection_date_precision: day [UO:0000033]
geo_loc name (country): Canada [GAZ:00002560]
geo_loc name (state/province/territory): Ontario [GAZ:00002563]
food_product: Beef (ground or minced) [FOODON:00001282]
food_product_properties: Food (frozen) [FOODON:03302148]
animal_source_of_food: Cow [NCBITaxon:9913]
environmental_site: Retail environment [ENVO:01001448]

Strain and isolate information
organism: Escherichia coli [NCBITaxon:561]
isolate_ID: badbug444
isolated_by: Public Health Agency of Canada (PHAC) [GENEPIO:0100551]
taxonomic_identification_process: PCR assay [OBI:00002740]

Scenario 2: Food (produce)

A sample of water from rinsed lettuce was obtained in BC in 2020. The lettuce had been shipped from the US.

The lettuce was involved in an outbreak and the sample was assigned an identifier by PHAC (sample ID P5677-jkwncjwnc).

The sample was found to contain Escherichia coli serotype O157:H7 via traditional serotyping (isolate ID coli8352835) and was sequenced as part of the GRDI by PHAC (IRIDA ID GRDI_LL_12345, IRIDA Project ID 666, contact Philomena.cunk@phac.ca) using an Illumina HiSeq 4000.

MICs for a Sensititre 123 (broth dilution) drug panel are available (SIR phenotypes interpreted using CLSI M100 standard).

Sample collection and processing
specimen_collector_sample_ID: P5677-jkwncjwnc
sample_collected_by: Public Health Agency of Canada (PHAC) [GENEPIO:0100551]
sample_collector_contact_email: Missing [GENEPIO:0001618]
purpose_of_sampling: Cluster/Outbreak investigation [GENEPIO:0100001]
sample_collection_date: 2020
geo_loc name (country): Canada [GAZ:00002560]
geo_loc name (state/province/territory): British Columbia [GAZ:00002562]
food_product_origin geo_loc name (country): United States of America [GAZ:00002459]
environmental_material: Water [CHEBI:15377]
collection_method: Rinsing for specimen collection [GENEPIO_0002116]
food_product: Lettuce head (whole or parts) [FOODON:03000239]

Strain and isolate information
organism: Escherichia coli [NCBITaxon:561]
serovar: O157:H7
serotyping method: Traditional serotyping
isolate_ID: coli8352835
isolated_by: Public Health Agency of Canada (PHAC) [GENEPIO:0100551]
taxonomic_identification_process: Whole genome sequencing assay [OBI:0002117]
Sequencing Information
sequenced_by: Public Health Agency of Canada (PHAC) [GENEPIO:0100551]
sequenced_by_contact_email: Philomena.cunk@phac.ca
purpose_of_sequencing: Research [GENEPIO:0100003]
sequencing_instrument: Illumina HiSeq 4000 [GENEPIO:0100119]
IRIDA_isolate_ID: GRDI_LL_12345
IRIDA_project_ID: 666
AMR Testing Information
AMR_testing_by: Public Health Agency of Canada (PHAC) [GENEPIO:0100551]
AMR_testing_by_contact_name: Missing [GENEPIO:0001618]
AMR_testing_by_contact_email: Missing [GENEPIO:0001618]
AMR_laboratory_typing_platform: Sensititre [ARO:3004402]
AMR_laboratory_typing_method: Broth dilution [ARO:3004397]
gentamicin_measurement: 64
gentamicin_measurement_unit: ug/mL
gentamicin_measurement_sign: greater than or equal to (>=) [GENEPIO:0001005]
gentamicin_susceptible_breakpoint: 4

gentamicin_resistance_breakpoint: 32
gentamicin_resistance_phenotype: Resistant antimicrobial phenotype [ARO:3004301]
gentamicin_testing_standard: Clinical Laboratory and Standards Institute (CLSI) [ARO:3004366]
gentamicin_testing_standard_version: M100

Scenario 3: Environmental (Abx pre-treatment)

The AAFC performed a field experiment examining AMR spread on pig farms in the prairies via aerosolization. Feed was pre-treated with antibiotic X (400ppm) and distributed to animals, and the air was sampled (sample ID: QWERTY888) by leaving an agar dish (MacConkey) open for 3hrs on March 3 2021. E. coli isolates are currently being selected for sequencing as part of the GRDI.

Sample collection and processing
specimen_collector_sample_ID: QWERTY888
sample_collected_by: Agriculture and Agri-Food Canada (AAFC) [GENEPIO:0100553]
sample_collector_contact_email: Missing [GENEPIO:0001618]
purpose_of_sampling: Field experiment [GENEPIO:0100550]
sample_collection_date: 2021-03-21
geo_loc name (country): Canada [GAZ:00002560]
geo_loc name (state/province/territory): Prairie region (Canada) [wikidata:Q1364746]
original_sample_description: air - MacConkey II plate open 3 hours (pig farm)
environmental_material: Air [ENVO:00002005]
environmental_site: Farm [ENVO:00000078]
animal_or_plant_population: Pig [NCBITaxon:9823]
collection_device: Culture plate [GENEPIO:0004318]

Stain and isolate information
organism: Escherichia coli [NCBITaxon:561]
isolated_by: Agriculture and Agri-Food Canada (AAFC) [GENEPIO:0100553]

Risk assessment
experimental_intervention: Addition of substances to food/water [GENEPIO:0100536];Antimicrobial pre-treatment [GENEPIO:0100537]
experimental_intervention_details: feed supplemented with 440 ppm antibiotic X

Scenario 4: Host (certified practices)

The AAFC performed a survey study examining pig husbandry practices on AMR spread in abattoirs in Quebec. A fecal sample (single source; sample ID AAFC-PP-456701b) was obtained from a pig in an abattoir on August 25 2025. The pig had been raised on a farm that was certified for implementing the AGRO-GENTLEMENPIGS husbandry practices.

E. coli isolates are currently being sequenced as part of the GRDI. The study confirmed 47/1000 E.coli isolates tested positive for the same AMR determinant at the harvest stage of production (abattoir) when husbandry practices were used, compared to 89/1000 positive isolates on uncertified farms.

Sample collection and processing
specimen_collector_sample_ID: AAFC-PP-456701b
sample_collected_by: Agriculture and Agri-Food Canada (AAFC) [GENEPIO:0100553]
sample_collector_contact_email: Missing [GENEPIO:0001618]
purpose_of_sampling: Survey study [GENEPIO:0100582]
sample_collection_date: 2025-08-25
geo_loc name (country): Canada [GAZ:00002560]
geo_loc name (state/province/territory): Quebec [GAZ:00002569]
original_sample_description: pig feces, abattoir, AGRO-GENTLEMENPIGS
body_product: Feces [UBERON:0001988]
environmental_site: Abattoir [ENVO:01000925]
specimen_processing: Isolated from single source [OBI:0002079]
Host information
host_(scientific_name): Sus scrofa domesticus [NCBITaxon:9825]
host_(common_name): Pig [NCBITaxon:9823]

Stain and isolate information
organism: Escherichia coli [NCBITaxon:561]
isolated_by: Agriculture and Agri-Food Canada (AAFC) [GENEPIO:0100553]

Risk assessment
experimental_intervention: Certified animal husbandry practices [GENEPIO:0100538]
experimental_intervention_details: testing effects of animal husbandry practices - AGRO-GENTLEMENPIGS
prevalence_metrics: 47/1000 E.coli isolates tested positive for the same AMR determinant
prevalence_metrics_details: 89/1000 positive isolates on uncertified farms
stage_of_production: Abattoir [ENVO:01000925]

Scenario 5: Environmental (manure)

The AAFC performed a research project examining the prevalence and diversity of AMR determinants in manure on conventional dairy farms in Saskatchewan. A sample of manure (sample ID AAFC-MF-8648045) was obtained from an animal pen on a dairy farm in early 2025. The farm is part of a dairy production stream, and the lab also has feed history and Vitek microbial profile data available. E. coli isolates are currently being selected for sequencing as part of the GRDI.

Sample collection and processing
specimen_collector_sample_ID: AAFC-PP-456701b
sample_collected_by: Agriculture and Agri-Food Canada (AAFC) [GENEPIO:0100553]
sample_collector_contact_email: Missing [GENEPIO:0001618]
purpose_of_sampling: Survey study [GENEPIO:0100582]
presampling activity: Conventional farming practices [GENEPIO:0100895]
sample_collection_date: 2025
geo_loc name (country): Canada [GAZ:00002560]
geo_loc name (state/province/territory): Saskatchewan [GAZ:00002564]
original_sample_description: manure from a conventional dairy farm, animal pen
food_product_production_stream: Milk production stream [FOODON:03000459]
Sample collection and processing (cont'd)
environmental_site: Dairy farm [ENVO:03501416];Animal pen [ENVO:03501387]
environmental_material: Animal manure [AGRO:00000079]
available_data_types: Feed history [GENEPIO:0100704];Microbiological identification (ViTek) [GENEPIO:0100721]

Stain and isolate information
organism: Escherichia coli [NCBITaxon:561]
isolated_by: Agriculture and Agri-Food Canada (AAFC) [GENEPIO:0100553]

Scenario 6: Environmental (surface water)

The ECCC performed a research project examining the prevalence and diversity of AMR determinants in surface water. Five water samples were collected from different locations in Lake Ontario and pooled (sample ID ECCC-LO-99666) on a sunny day (July 23 2024), at a depth of 1m. It had rained prior to sample collection (5cm). E. coli isolates are currently being selected for sequencing as part of the GRDI. While most of the sample was used for microbiological testing, some of the water sample (residual sample) is still available and is currently stored at -80C.

Sample collection and processing
specimen_collector_sample_ID: ECCC-LO-99666

sample_collected_by: Environment and Climate Change Canada (ECCC) [GENEPIO:0100555]
sample_collector_contact_email: Missing [GENEPIO:0001618]
purpose_of_sampling: Research [GENEPIO:0100003]
sample_collection_date: 2024-07-23
geo_loc name (country): Canada [GAZ:00002560]
geo_loc name (state/province/territory): Ontario [GAZ:00002563]
geo_loc_name (site): Lake Ontario
original_sample_description: surface water, pooled
specimen_processing: Samples pooled [OBI:0600016]
specimen_processing_details: 5 samples from different sites in the lake pooled
environmental_site: Lake [ENVO:00000020]
environmental_material: Surface water [ENVO:00002042]
residual_sample_status: Residual sample remaining (some sample left) [GENEPIO:0101087]
Environmental conditions and measurements
sampling_weather_conditions: Sunny/Clear [ENVO:03501421]
water_depth: 1
water_depth_units: meter (m) [UO:0000008]
presampling_weather_conditions: Rain [ENVO:01001564]
precipitation_measurement_value: 5
precipitation_measurement_unit: centimeter (cm) [UO:0000015]

Stain and isolate information
organism: Escherichia coli [NCBITaxon:561]
isolated_by: Environment and Climate Change Canada (ECCC) [GENEPIO:0100555]