GRDI-AMR Metadata Curation SOP 8.9

May 2023

By Emma Griffiths, Julie Shay, Rhiannon Cameron, and Anoosha Sehar

Table of Contents

1. Background

Genomic surveillance using a One Health approach is a powerful tool for understanding and tracking how pathogens affecting human health evolve and spread. Genomic surveillance of pathogens requires high quality sequence data as well as well structured contextual data. Contextual data is the sample, laboratory, clinical, epidemiological, and methods information that enables the interpretation of sequence data. One Health initiatives often involve data streams from different sources, agencies, sectors, and information management systems, and because the data is structured in different ways it is often difficult to harmonize and integrate. 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 Canadian Antimicrobial Resistance Genomics Research and Development Initiative (AMR-GRDI) uses a genomics-based approach to understand how food production contributes to the development of antimicrobial resistance of human health concern, and explore strategies for reducing antimicrobial resistance in food production systems. The AMR-GRDI is a component of the Federal Action Plan for Antimicrobial Resistance and Use in Canada, and involves data streams from five federal departments and agencies spanning human health, agriculture, the environment, and food regulation.

To better harmonize AMR-GRDI contextual data across sectors and agencies, the Metadata Harmonization Working Group has developed an ontology-based One Health AMR data standard for foodborne pathogens, which provides standardized fields, pick lists of controlled vocabulary and prescribed formats for the harmonized capture of contextual data. The standardized fields are based on community standards such as NCBI's combined Pathogen and Environmental attribute package derived from internationally agreed upon Minimal Data for Matching (MDM) standards, as well as applicable fields from different MIxS packages (Genomic Standards Consortium). The data standard will also be harmonized with recently developed standards (e.g. One Health Enteric Package v1.0: Expanded and Standardized Metadata for Enteric Genomic Epidemiology in the U.S, and the Food MIxS package). The AMR-GRDI standard is implemented via a spreadsheet-based data collection instrument (i.e. metadata template) and accompanying Field and Term reference guides (which provide definitions and additional specific guidance) and this curation SOP.

2. What's New in the AMR-GRDI standard?

An initial template was tested by nominated curators representing different data providers, and feedback from that pilot implementation was incorporated in a revised collection template (GRDI Template 2.0), Reference Guide, and SOP to improve usability and clarity. As the template is put into practice, feedback from users is (and will be) continually incorporated. Based on lessons learned, the improved collection template contains refined field labels, as well

as a number of new fields in order to enhance isolate tracking and include more information. The fields have now been organized into separate tabs according to whether they pertain to Sample Collection and Processing, Strain and Isolate Information, Sequence and Public Repository Information, AMR Phenotypic Testing Information, and Risk Assessment Information. To ensure a minimal set of metadata across all sample types, a colour-code scheme was introduced highlighting essential epidemiological data elements for establishing place and time, and for establishing provenance and attribution of information. “Required” fields are colour-coded in yellow, recommended fields are coloured purple, and optional fields are in white. We recommend that all required fields be filled, however curators should work with data stewards to decide how much or how little of the other information can be provided. The full spectrum of fields is meant to address a wide variety of use cases and sample types, however, only fields pertaining to a curator’s use case are expected to be populated.

To reduce the burden of data entry, curators need only enter sample and/or isolate information once i.e. sample information should be entered in the Sample Collection and Processing Information sheet, isolate information should be entered in Strain and Isolate Information sheet). Subsequently, the sample and isolate identifiers need only be included in other tabs. Sample and isolate information will then be automatically populated in the merged sheet (second tab in the template) - curators need not enter any information into the merged sheet. The fully populated summary sheet can then be uploaded to IRIDA.

In July 2022, another updated version of the GRDI template (GRDI-AMR Template 3.0) was released which included additional environmental fields and a new Term Reference Guide (in addition to the Field Reference Guide). A New Term Request System and SOP was also developed.

In May 2023, a new version of the GRDI template (GRDI-AMR Template 7.6.4) was released which included a variety of new sample collection fields requested from users, new guidance for attributing and formatting agency-specific identifiers for improved traceability, as well as an expansion and in some cases, restructuring, of picklists for several existing fields. The template described below is now also available in a downloadable data creation application called the DataHarmonizer, which offers additional curation features, validation, and automated data transformation capabilities. The latest GRDI-AMR harmonization package provides curators with two different tools for capturing contextual data in the same way (the Excel template, and the DataHarmonizer app) - and users can choose whichever tool they prefer for harmonizing information. Instructions for downloading and using the DataHarmonizer app are linked in the GRDI-AMR GitHub repository README file.

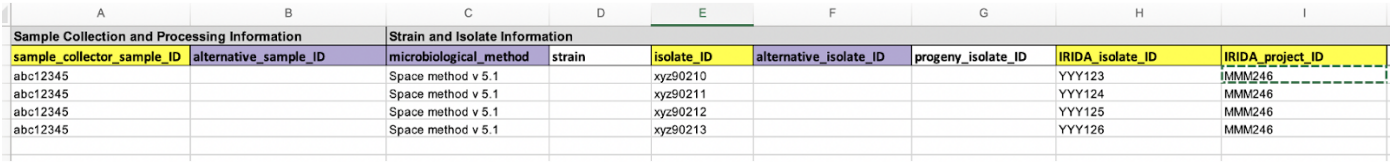
How to use the template: Worked examples

Enter sample information once in the Sample Collection & Processing tab.

e.g.

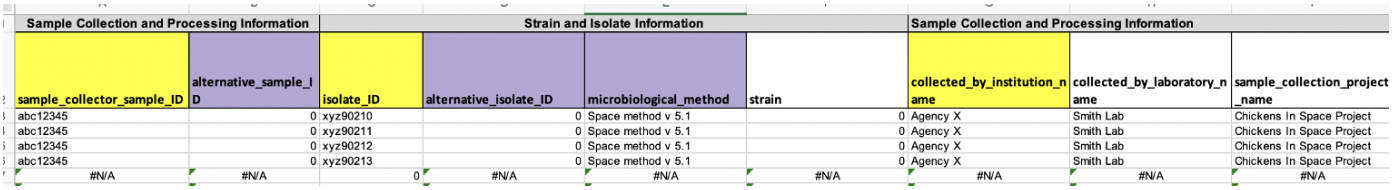
Sample Collection and Processing Information						
sample_collector_sample_ID	alternative_sample_ID	collected_by_institution_name	collected_by_laboratory_name	sample_collection_project_name	sample_plan_name	sample_plan_ID
abc12345		Agency X	Smith Lab	Chickens In Space Project	Space Chickens Sample Collection v 3.5	X9876

In the Isolate Information tab, simply reuse the sample ID for all isolates from the same sample.
e.g.



A	B	C	D	E	F	G	H	I
Sample Collection and Processing Information		Strain and Isolate Information						
sample_collector_sample_ID	alternative_sample_ID	microbiological_method	strain	isolate_ID	alternative_isolate_ID	progeny_isolate_ID	IRIDA_isolate_ID	IRIDA_project_ID
abc12345		Space method v 5.1		xyz90210			YYY123	MMM246
abc12345		Space method v 5.1		xyz90211			YYY124	MMM246
abc12345		Space method v 5.1		xyz90212			YYY125	MMM246
abc12345		Space method v 5.1		xyz90213			YYY126	MMM246

The “Merged Sheet” will automatically generate the complete set of contextual data for you. You can use the same technique if an isolate has more than one sequence, or more than one set of AMR testing results (i.e. use the isolate ID multiple times in subsequent tabs without the need for repeat entry of isolate information).
e.g.



Sample Collection and Processing Information		Strain and Isolate Information				Sample Collection and Processing Information		
sample_collector_sample_ID	alternative_sample_ID	isolate_ID	alternative_isolate_ID	microbiological_method	strain	collected_by_institution_name	collected_by_laboratory_name	sample_collection_project_name
abc12345	0	xyz90210	0	Space method v 5.1	0	Agency X	Smith Lab	Chickens In Space Project
abc12345	0	xyz90211	0	Space method v 5.1	0	Agency X	Smith Lab	Chickens In Space Project
abc12345	0	xyz90212	0	Space method v 5.1	0	Agency X	Smith Lab	Chickens In Space Project
abc12345	0	xyz90213	0	Space method v 5.1	0	Agency X	Smith Lab	Chickens In Space Project
#N/A	#N/A	0	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A

Below are instructions for curating contextual data according to fields prescribed by the data standard. Field-level definitions, guidance, and examples of use are also provided in the Reference Guide (first tab in collection template). If information cannot be supplied for any of the required fields, please provide an acceptable null value such as: Missing, Not Collected, Not Provided, Not Applicable, Restricted Access. Recommended or optional fields that do not apply or cannot be filled, should be left blank.
*Please note, the contextual data captured in these spreadsheets is for internal use (GRDI) only, unless consent for broader sharing is provided by data stewards according to jurisdictional data sharing policies. As such, please provide the most detailed information available and permissible. There is no need to provide higher level categories e.g. commodity types, which may require reclassification or may not be provided by other groups.

For troubleshooting, further training, and/or for more information, please contact Dr. Emma Griffiths at ega12@sfu.ca.

3. The Importance of FAIR Data

It might be cliche but it’s true - pathogens do not acknowledge geographical borders. Sharing of data is fundamental for tracking and controlling infectious diseases, and understanding how pathogens evolve and share genetic material. Appropriate storage and management of genomics data (both sequence and contextual) is key to ensuring its longevity for verification, reuse, and innovation. In 2015, a consortium of authors from both academia and industry who believed the current ways of processing and managing data were not “extracting the maximum

benefit from research investments" published a paper describing four foundational principles that would encourage knowledge discovery, integration, and innovation known as the FAIR Guiding Principles which stand for Findability, Accessibility, Interoperability, and Reusability.

The FAIR guiding principles are unique as they promote accessibility and longevity of our digital data to ensure data can be "easily found, accessed, understood, exchanged, and reused".
In its simplest form, **Findable** means that other users can discover your data. Data should be easy to find, for both humans AND machines. While humans can understand domain-specific vocabulary, jargon, slang and colloquialisms, machines cannot (unless you tell them how) so contextual data needs to be machine readable. **Accessibility** means that the data can be made readily available to other potential users. There should be clear protocols in place that allow data to be easily retrieved. **Interoperability** refers to the ability of the data or contextual data to be integrated with other datasets, databases, and platforms. Often datasets from different sources and laboratories must be combined for different types analyses so it is important that the data is encoded with similar formats, and is well described to ensure that the appropriate interpretations can be made for different use cases. The old saying "time is money" in the genomics era should be updated to "data is money". The wealth of information genomic sequences contain can only be extracted if sufficient and interoperable contextual data is stored and shared along with them. Because genomics data can be used in different analyses, they are profoundly **Reusable**.

Using ontologies to standardize contextual data helps make your data FAIR by implementing a shared, interoperable, common language. Ontologies are open source and meant to represent "universal truth" as much as possible (so not tied to one organization's vocabulary of use case). Ontologies encode synonyms, which enables mapping between the specific languages used by different organizations, and every term in the ontology is assigned a globally unique and persistent identifier. Good ontologies are registered in public registries (such as the OBO Foundry Library of ontologies) which means their terms, relations, identifiers, and definitions are globally accessible and available for community input. In the GRDI-AMR, the fields and terms are ontology-based. Terms in the pick lists are associated with ontology language and identifiers in order to make contextual data FAIR.

4. How to Begin Using the Standard

Before you begin to curate sample metadata, there are a few things you should familiarize yourself with first.

a) **Review your dataset.**
   Look over the fields of information that you have. Which fields represent the original data and which fields represent downstream information (progeny isolates, IDs assigned by partner labs)? **Make sure that the sample and isolate IDs you have been given represent those from the originating lab i.e. pertain to the original sample or the**

isolate obtained directly from the sample. If you're not sure, consult the data provider. Why were the samples collected, and why were the isolates selected for sequencing? Were there any experimental conditions or variables tested prior to sample collection (e.g. pre-treatments, genetic mutation)? If a food product was tested, did it (or the animal source of the food) originate in another country? The dates you are including - do they pertain to when a sample was collected or when an isolate originally obtained, or when the sample/isolate was received by your lab? Which agency/lab performed the sample collection vs organism isolation vs sequencing?

b) **Review the Collection Template.**
Review the fields - do you have all the necessary information e.g. sample collector contact email? If not, discuss missing metadata with the Harmonization team and the data provider. Look over the definitions provided in the Reference Guide tab. Review the links to standardized term look-up services also provided in the Reference Guide.

5. Populating Fields - Putting Information in the Right Place

Use the field definitions and guidance provided in the Reference Guide (see first tab of the collection template spreadsheet) to fill out the fields and terms in the collection template. The fields used to describe the sample, the isolate, the sequence, and to provide information regarding AMR testing and risk assessment, are outlined below. Remember, only the fields that apply to your use case and your sample/isolate/sequence need to be filled. Curation training sessions are also available and may help to answer specific questions and inform practice.

6. Sample Collection and Processing Information

In this section, enter information about the ***original sample***.
A sample may be from a host (direct sampling of an animal or plant, and usually involves an anatomical part/material/body product), a natural or built environment (usually involves an environmental site/material; environmental materials can be substances or objects such as water or tractors whereas environmental sites are places or things that do not usually move such as farms or parts of buildings), or from food. A sample may also be collected using a particular device or technique, which is informative for downstream analyses.

Ideally, you will have several types of information about your sample which you can enter into a number of different fields. Provide as much information as you can about the sample and its processing e.g. where and when it was collected, how it was collected, what the sample

contains, qualities or characteristics that are known about the original material sampled, and whether the sample is from a single source or was pooled from multiple sources.

Pick lists of standardized ontology terms (organization-agnostic controlled vocabulary arranged in a hierarchy, with terms specified by unique identifiers and linked by logical relationships) and their identifiers are provided for a wide variety of fields. If a desired term is missing from a field, first check that a synonym has not been used in its place. If no equivalent term is present, check that the term is not available in another field. If the desired term cannot be found, new terms can be requested by the process outlined in Section 6 below.

If you have multiple isolates, sequences or tests, you only need to enter sample information once in this section, and then can refer to the corresponding sample ID in the subsequent sections.

Curators should try to fill as many of the following fields that apply to their samples, as permitted according to organizational policies.

Field Name	Field Definition
sample_collector_sample_ID	The user-defined identifier for the sample, as provided by the laboratory that collected the sample.
alternative_sample_ID	An alternative sample_ID assigned to the sample by another organization.
sample_collected_by	The name of the agency, organization or institution with which the sample collector is affiliated.
sample_collected_by_laboratory_name	The specific laboratory affiliation of the sample collector.
sample_collection_project_name	The name of the project/initiative/program for which the sample was collected.
sample_plan_name	The name of the study design for a surveillance project.
sample_plan_ID	The identifier of the study design for a surveillance project.
sample_collector_contact_name	The name or job title of the contact responsible for follow-up regarding the sample.
sample_collector_contact_email	The email address of the contact responsible for follow-up regarding the sample.
purpose_of_sampling	The purpose of sample collection.
presampling_activity	The activities or variables introduced upstream of sample collection that may affect the sample

	collected.
presampling_activity_details	The details of the activities or variables introduced upstream of sample collection that may affect the sample collected.
specimen_processing	The processing applied to samples post-collection, prior to further testing, characterization, or isolation procedures.
geo_loc (country)	The country where the sample was collected.
geo_loc (state/province/region)	The province/territory where the sample was collected.
geo_loc_name (site)	The name of a specific geographical location e.g. Credit River (rather than river).
food_product_origin geo_loc (country)	The country of origin of a food product.
host_origin geo_loc (country)	The country of origin of the host.
geo_loc latitude	The latitude coordinates of the geographical location of sample collection.
geo_loc longitude	The longitude coordinates of the geographical location of sample collection.
sample_collection_date	The date on which the sample was collected.
sample_received_date	The date on which the sample was received by the laboratory.
original_sample_description	The original sample description provided by the sample collector.
environmental_site	An environmental site is a location in the natural or built environment.
water_depth	The depth of some water.
sediment_depth	The depth of some sediment.
air_temperature	The temperature of some air.
water_temperature	The temperature of some water.
weather_type	The state of the atmosphere at a place and time as regards heat, dryness, sunshine, wind, rain, etc.
available_data_types	The type of data that is available, that may or may not require permission to access.
animal_or_plant_population	The type of animal or plant population inhabiting an area.
environmental_material	A substance or object obtained from the natural or man-made environment .
anatomical_material	A substance obtained from an anatomical part of an

	organism e.g. tissue, blood.
body_product	A substance excreted/secreted from an organism.
anatomical_part	A substance obtained from an anatomical part of an organism.
food_product	A material consumed and digested for nutritional value or enjoyment.
food_product_properties	Any characteristic of the food product pertaining to its state, processing, a label claim, or implications for consumers.
animal_source_of_food	The animal from which the food product was derived.
production_stream	A production pathway incorporating the processes, material entities (e.g. equipment, animals, locations), and conditions that participate in the generation of a particular commodity.
sample_storage_method	A specification of the way that a specimen is or was stored.
sample_storage_medium	The material or matrix in which a sample is stored.
collection_device	The instrument or container used to collect the sample e.g. swab.
collection_method	The process used to collect the sample e.g. phlebotomy, necropsy.
food_packaging	The type of packaging used to contain a food product.
food_quality_date	A date recommended for the use of a product while at peak quality, this date is not a reflection of safety unless used on infant formula.
food_packaging_date	A food product's packaging date as marked by a food manufacturer or retailer.

Sample Collection and Processing curation tips:

1. **sample_collector_sample_ID** refers to the **original ID of the sample**.
2. **alternative_sample_ID** refers to any ID assigned to the sample by any other lab. If you received a sample from another lab and assigned it an ID, put your ID in the alternative_sample_ID field. Make sure to include the original sample_ID to establish the chain of custody, and to ensure that samples are properly tracked between labs. Alternative sample IDs should be provided in the in a prescribed format which consists of the ID followed by square brackets (no space in between the ID and bracket) containing the short form of ID provider's agency name i.e. ID[short organization code]. Agency short forms include the following:

Public Health Agency of Canada: **PHAC**
Canadian Food Inspection Agency: **CFIA**
Agriculture and Agri-Food Canada: **AAFC**
Fisheries and Oceans Canada: **DFO**
Environment and Climate Change Canada: **ECCC**
Health Canada: **HC**
An example of a properly formatted alternative_sample_identifier would be e.g. **ABCD1234[PHAC]**
Multiple identifiers can be provided and separated by a semi-colon. If an agency name does not appear in this list and has provided an identifier that should be tracked, contact Emma Griffiths (ega12@sfu.ca).

3. There are fields to specify the **organization** that collected samples (e.g. CFIA), as well as **specific labs** (Smith lab or Food Pathogen Lab). These fields represent provenance information at different levels of granularity and help establish chain of custody and also provide attribution for crediting contributions to projects. Agency names should be provided in the "**sample_collected_by**" field. "**sample_collected_by_laboratory_name**" can be used to track laboratory names more granular than the agency name. If multiple (or different) sections or divisions need to be specified, they can be listed and separated by a semi-colon.
4. It is useful to provide project names associated with sample collection as there is often further documentation that can be obtained which may provide details useful for including/excluding samples in analyses and/or understanding their limitations. Projects associated with sample collection should be included in the ""**sample_collection_project_name**" field.
5. A sample plan is a detailed outline of which measurements will be taken at what times, on which material, in what manner, and by whom. Sample plans are often associated with projects and describe project aims, sampling strategies, scope of sample types, methods, etc. If a sample plan is associated with a sample, the name of the plan can be included under "**sample_plan_name**". If the sample plan is known by an identifier, the identifier should be included under "**sample_plan_ID**". The inclusion of this information, if available, is highly recommended.
6. Contact information is useful should an analysis require further information about a sample. For many reasons, the individual that originally carried out sample collection may not be available for follow-up. It is therefore recommended that the email address designated for a position or lab (rather than for an individual person) be included in the "**sample_collector_contact_email**" field, to ensure accuracy of information and institutional memory.
7. **"Sample_collection_date"** is the date a sample was originally collected. Isolation_date is the date when an isolate was originally isolated from the sample. Sample_received_date and isolate_received_date should be used to denote when samples and isolates were received rather than originally collected/isolated.

8. Dates should always be denoted according to the ISO 8601 standard **"YYYY-MM-DD"**. To avoid formatting errors in the Excel collection template, include “00” if the month or day cannot be provided e.g. 2021-04-00 or 2021-00-00. To ensure that “00” has not been entered in error, if a date is incomplete (i.e. only to the year or month), the “**sample_collection_date_precision**” field should be filled in and the corresponding level of granularity of the date information selected from the picklist.
9. If the organization sequencing a sample is not the organization that collected the sample, sometimes collection dates are unknown. If the sample received date is known, enter that in the “sample received date” field and put a null value in the “**sample_collection_date”** field.
10. While standardized sample descriptions use highly useful for analyses, sample descriptions are often provided in different formats or free text. It is possible that some information in the original sample description may be lost during standardization (such as brand names or highly granular details). It is good practice to retain original sample descriptions along with standardized attributes. Original sample descriptions should be stored in the “ **original_sample_description**” field.
11. Geographical locations may be sensitive information. Check with your data steward before providing any information more granular than province/territory. Geographical information may appear in agency or laboratory names associated with sample collection. **Check that the lab name you include does not provide sensitive information about samples.**
12. Most often, latitude and longitude of sample collection fields are populated with a null value. **Do not extrapolate central coordinates for provinces, countries or laboratories and enter these in the latitude and longitude of sample collection fields.** While useful in analyses, such extrapolations implicate real locations and the information, if represented as true values, may be misleading. Only include true coordinates if you have permission to share them.
13. **How samples are chosen for collection** may have inherent biases due to sampling strategy. Provide the reason for sample collection in the “**purpose_of_sampling”** field. If known, you can provide sample plan and project names/identifiers to link materials to projects and additional information more easily.
14. To capture presampling processes and events which may affect samples and isolates, there are two fields called **“presampling_activity”** and **“presampling_activity_details”**, where curators can select a standardized category for experimental activities as well as providing further details using free text to make methods more explicit e.g. describing treatments provided in feed or genetic mutations that were introduced in the lab.
15. If your sample was generated from a single source, provide “Isolated from single source” under **“specimen_processing”**. If your sample represents different

materials or sampled sites that were pooled, enter “samples pooled” under sample_processing.

16. Sometimes samples have been tested or measured in different ways (e.g. chemical measurements, physical measurements) or additional records were kept that can provide additional contextual information (e.g. therapeutic or drug administration histories). Additional available data types may require insights by the data steward with regards to methods or interpretations. Data curators may indicate that additional data is available (beyond the contextual data provided in the template) by providing a list of data types selected from the picklist in the “**available_data_types**” field. Multiple data types can be provided, and should be separated by a semi-colon.
17. Food samples do not have a host, so only fill in the food-related fields. Environmental samples also do not have a host, so only fill in the environment-related fields. If a plant or animal was directly sampled, provide the common name for the host in “**host (common name)”** and the taxonomic name for the host in the “**host (scientific name)”** field (see Host Information section below).
18. Different samples may be taken from a host, such as an anatomical_part, anatomical_material or body_product. An “**anatomical_part”** is an anatomical location/structure (e.g. lung, knee, cloaca), whereas an “**anatomical_material”** is material that can be removed from an anatomical part (e.g. blood, tissue, cecal contents). A “**body_product”** is a material meant to be excreted/secreted from a host, and is usually a waste product (e.g. feces, urine, vomit).
19. The devices and processes used to collect a sample are often informative for interpreting results. Information about what devices were used to collect a sample (e.g. whirl pack, swab) should be provided in the “**collection_device”**, while types of techniques (e.g. rinsing, lavage, necropsy) used to collect a sample can be included in “**collection_method”**. Provide the name of the protocol for sample collection in the “**sample_plan_name”** field.
20. If your sample was from an environmental site that is inhabited by a certain population of plants or animals e.g. turkey farm, crop farm. In these cases, provide the proper noun in the environmental site field (in this case “farm”), and the modifier in the “**animal_or_plant_population”** field (in this case “turkey” or “crop”).
21. **Upstream information about food samples or food production animals** is often useful for interpretation of results and for linking pathogens to upstream sources. Two fields, “**food_product_origin (geo_loc (country))**” and “**animal_source_of_food**” can be used to track whether food products or their sources are from other countries (although the food was sampled in Canada), as well as the animal source of the food (e.g. cheese from a goat as opposed to a cow, chilli made with ground beef (animal source is a cow) vs ground turkey (animal source is turkey)). Use the common name of the animal in the “**animal_source_of_food**” field. When referring to the geographical origins of

multi-ingredient foods, prioritize the geographical origin of the primary ingredient. Notes about ingredient sources can be included in the "original_sample_description" field. When interpreting information provided by others in the "food_product_origin (geo_loc (country))" field, it is important to note that not all information for all ingredients may have been included. If an interesting signal is detected, it is good practice to follow-up with the data provider before you act to ensure your analysis is properly informed.

22. Food products have often been modified by processes (e.g. cooking, preservation, portioning) so that they have certain properties or modifications. Qualities or characteristics of food products should be added in the "**food_product_properties"** field (e.g. Ready-to-eat (RTE)). Label claims about a food product (e.g. Organic food claim or use) can also be added here. Whether a food was fresh or processed (e.g. frozen or preserved by drying, canning etc) is usually a critical piece of information for understanding how pathogens are introduced. It is highly recommended that this information be included when it is available. Dates indicating thresholds for food quality (e.g. best before, use by) should be included in the "**food_quality_date**" field. The date a food was packaged, as opposed to consumption dates, should be included in the "**food_packaging_date**" field.
23. A production stream is a production pathway incorporating the processes, material entities (e.g. equipment, animals, locations), and conditions that participate in the generation of a particular commodity. Classifying components of the same production stream is often useful in different analyses. If a sample is part of the production pathway of a particular commodity or food production animal, this can be indicated in the "**production_stream**" field by using the picklist provided.

7. Host Information

In this section, enter information about the ***host***.
The "host" is the organism that was directly sampled and from which a pathogen is isolated. Not all samples will involve hosts (e.g. food or environmental samples). Clinical samples involve human hosts (Homo sapiens). Wild and domesticated animals are also considered hosts and may be sampled for animal health diagnostics, regulatory purposes, or monitoring/surveillance. Individual plants can be considered hosts, however, plants can also be considered food or environmental materials. Whether a plant is considered a host or material is up to the discretion of the curator.

The host can be specified by its common name (e.g. Cow) or its scientific or taxonomic name (e.g. Bos taurus). Where possible, it is preferable to provide the non-abbreviated taxonomic name of the host. Sometimes the precise taxonomic name of the host will not be known, in which case common names should be provided (e.g. Tuna). More specific information

about the host can also be provided using the "**host (ecotype)**", "**host (breed)**", and "**host (food production name)**" fields. If a host was known to present with a disease, this information can be included in the host_disease field. The age range of the host can be provided in the "**host_age_bin**" field. Currently age bins are provided as attributes for banded birds. To include additional types of age bins (such as ranges of years appropriate for human host samples), contact Emma Griffiths (ega12@sfu.ca).

A host is directly sampled when substances are removed from its body (e.g. bodily fluids, tissue (animal or plant)), or if body products from that individual are sampled.

Field Name	Field Definition
host (common name)	The commonly used name of the host.
host (scientific name)	The scientific name of the host.
host (ecotype)	The biotype resulting from selection in a particular habitat, e.g. the A. thaliana Ecotype Ler.
host (breed)	A breed is a specific group of domestic animals or plants having homogeneous appearance, homogeneous behavior, and other characteristics that distinguish it from other animals or plants of the same species and that were arrived at through selective breeding.
host_age_bin	Age of host at the time of sampling, expressed as an age group.
host (food production name)	The name of the host at a certain stage of food production, which may depend on its age or stage of sexual maturity.
host_disease	The name of the disease experienced by the host.

8. How to Structure Your Sample Information from a Free Text Description

NCBI uses the "isolation_source" field to describe the material and/or site that was sampled and/or method used for collection. This results in different information types in the same field, which then become difficult to aggregate and compare. To improve how these information types are structured, 19 additional fields have been introduced to the GRDI data specification to capture different kinds of anatomical and environmental samples, food products and their properties, as well as collection methods, devices and storage conditions that could affect the sample (i.e. environmental_site,
water_depth, sediment_depth, air_temperature, water_temperature, weather_type, animal_or_plant_population, environmental_material, anatomical_material, body_product,

anatomical_part, food_product, food_product_properties, animal_source_of_food, food_packaging, sample_storage_method, sample_storage_medium, collection_device, collection_method). Not all of these fields need to be populated - only the fields that pertain to your sample. Provide the most granular information available. Use the pick lists provided. If desired terms are missing from the pick lists, use the New Term Request System and SOP to request the terms you need.

Below are several examples modelling how free text descriptions can be structured in the collection template.

e.g. human feces should be recorded:

Host (common name)	Host (scientific name)	body_product
Human [NCBITaxon_9606]	Homo sapiens [NCBITaxon_9606]	Feces [UBERON_0001988]

e.g. a swab in an abattoir should be recorded:

environmental_site	collection_device
Abattoir [ENVO_01000925]	Swab [NCIT_C17627]

e.g. organ from a turkey (Meleagris gallopavo) on a farm should be recorded:

Host (common name)	Host (scientific name)	anatomical_part	environmental_site
Turkey [NCBITaxon:9103]	Meleagris gallopavo [NCBITaxon:9103]	Organ [UBERON_0000062]	Farm [ENVO_00000078]

e.g. retail, ground chicken should be recorded:

environmental_site	food_product
Retail environment [NCIT_C99758]	chicken (ground or minced) [FOODON_03311826]

9. Strain and Isolate Information

In this section, enter information about the ***original strain or isolate***.

Here, a strain is considered an original lab reference strain or a strain from a type culture collection. An isolate is derived from a pure culture taken from a sampled material. A strain that has undergone several passages should also be considered an isolate as it may have accumulated mutations that distinguish it from the original reference strain.

If you have multiple sequences or tests for each isolate, you only need to enter isolate information once in this section, and then can refer to the corresponding isolate ID in the subsequent sections.

Curators should try to fill as many of the following fields as possible, as permitted according to organizational policies.

Field Name	Field Definition
microbiological_method	The laboratory method used to grow, prepare, and/or isolate the microbial isolate.
strain	The strain identifier.
isolate_ID	The user-defined identifier for the isolate, as provided by the laboratory that originally isolated the isolate.
alternative_isolate_ID	An alternative isolate_ID assigned to the isolate by another organization.
progeny_isolate_ID	The identifier assigned to a progenitor isolate derived from an isolate that was directly obtained from a sample.
IRIDA_isolate_ID	The identifier of the isolate in the IRIDA platform.
IRIDA_project_ID	The identifier of the Project in the iRIDA platform.
isolated_by	The name of the agency, organization or institution with which the individual who performed the isolation procedure is affiliated.
isolated_by_laboratory_name	The specific laboratory affiliation of the individual who performed the isolation procedure.
isolated_by_contact_name	The name or title of the contact responsible for follow-up regarding the isolate.
isolated_by_contact_email	The email address of the contact responsible for follow-up regarding the isolate.
isolation_date	The date on which the isolate was isolated from a sample.
isolate_received_date	The date on which the isolate was received by the laboratory.
organism	The taxonomic name of the organism.

taxonomic_identification_process	The type of planned process by which an organismal entity is associated with a taxon or taxa.
taxonomic_identification_process_details	The details of the process used to determine the taxonomic identification of an organism.
serovar	The serovar of the organism.
serotyping_method	The method used to determine the serovar.
phagetype	The phagetype of the organism.

Strain and Isolate Information curation tips:

1. **"isolate_ID"** refers to the ID of the **isolate obtained from the sample**. If you received an isolate from another lab and assigned it an ID, put the new ID in the "**alternative_isolate_ID"** field. Make sure to include the original isolate_ID to establish the chain of custody, and that isolates are properly tracked and differentiated between labs.
2. Alternative isolate IDs should be provided in the in a prescribed format which consists of the ID followed by square brackets (no space in between the ID and bracket) containing the short form of ID provider's agency name i.e. ID[short organization code]. Agency short forms include the following:
   Public Health Agency of Canada: **PHAC**
   Canadian Food Inspection Agency: **CFIA**
   Agriculture and Agri-Food Canada: **AAFC**
   Fisheries and Oceans Canada: **DFO**
   Environment and Climate Change Canada: **ECCC**
   Health Canada: **HC**
   An example of a properly formatted alternative_isolate_identifier would be e.g. **XYZ4567[CFIA]**
   Multiple alternative isolate IDs can be provided, separated by semi-colons.
   If an agency name does not appear in this list and has provided an identifier that should be tracked, contact Emma Griffiths (ega12@sfu.ca).
3. If the isolate was directly obtained from a sample, link the isolate_ID to the sample information. If the isolate has been passaged, the isolate is considered progeny of the original isolate. Put the ID of a progeny isolate in the "**progeny_isolate_ID"** field. Do not link sample information to the progeny as progeny are considered to be different organisms than that their parental isolates.
4. Best practices for establishing and using sample and isolate identifiers have been described in a companion document (**GRDI Contextual Data Guidance - Best Practices for Sample and Isolate Identifiers**). Identifiers should be alphanumeric and sufficiently complex to avoid ID clash (e.g. A1 is not sufficiently complex). Identifiers should avoid the inclusion of contextual data which could reveal private or identifiable information when shared (e.g. JonesFarm_egg_salmonella_toronto_2021_1a).

5. Agency names should be provided in the "**isolated_by**" field. "**isolated_by_laboratory_name**" can be used to track laboratory names more granular than the agency name. If multiple (or different) sections or divisions need to be specified, they can be listed and separated by a semi-colon.
6. Contact information is useful should an analysis require further information about a sample. For many reasons, the individual that originally isolated the organism may not be available for follow-up. It is therefore recommended that the email address designated for a position or lab (rather than for an individual person) be included in the "**isolated_by_contact_email**" field, to ensure accuracy of information and institutional memory.
7. If you know the name and version of the method or protocol used to generate the isolate, include this information in the "**microbiological_method"** field. The sample_plan and microbiological_method are different. The sample_plan outlines how samples are to be collected, the microbiological method describes the microbiological techniques/processes/steps used to isolate pure colonies and isolates.
8. Provide the "**IRIDA_project_ID"** and "**IRIDA_sample_ID"** used to upload the isolate information to IRIDA. Note that in IRIDA, sample_ID = isolate_ID.
9. The species of the organism should be entered in the **organism** field. Information about subspecies (denoted subsp.) should be included if available, however strain level nomenclature should be omitted. Organisms that have been taxonomically characterized to the genus level should be written without species information e.g. *Acinetobacter sp.* should be simply written *Acinetobacter.* A pick list has been provided to help standardize organismal nomenclature.
10. The methods used to characterize and isolate an organism can influence the results of an analysis as well as a taxonomic determination. It is strongly recommended that the method used to identify the organism be included in the "**taxonomic_identification_process"** field. The pick list provides high level testing types that can be used to help understand discrepancies or document uncertainty during identification (e.g. if an organism is typed using a phenotypic assay but the genomic sequence analysis suggests an alternative taxonomic determination, then a review of the original determination is likely warranted). Further details pertaining to assays and instruments can be provided if available in the "**taxonomic_identification_process_details"** field.
11. Sometimes organisms are characterized using methods other than sequencing (e.g. culture, PCR). When a preliminary characterization differs from the sequencing results, the sequencing results should be double checked and compared to the rest of the contextual data. Sequencing results are often (though not always) more specific than other microbiological or molecular methods. If the determined identity of the organism is believed to be different after sequencing, change the "organism" name to the most up-to-date information and indicate the method used in the "**taxonomic_identification_process"** field. A note can be

placed in the ”**taxonomic_identification_process_details”** field to track the change for auditability purposes.

12. If the serovar was determined via traditional serotyping methods, put “Traditional serotyping” in the **serotyping_method** field. If the serovar was determined via i*n silico* methods, provide the **name and version number of the software used**.

10. Sequence Information

In this section, enter information about the ***sequence***.
Isolates may have different associated sequences (e.g. isolates may be technical replicates, or be produced by different platforms/instruments). A sequencing library is a pool of DNA fragments with adapters attached, and is the starting point for generating a sequence. Associated library_IDs should be provided and linked to isolates and samples wherever possible. Sequence filenames (FASTQ or FAST5, and/or FASTA) should also be included.

Curators should try to fill as many of the following fields as possible, as permitted according to organizational policies.

Field Name	Field Definition
library_ID	The identifier of the library sequenced.
sequenced_by	The name of the agency, organization or institution responsible for sequencing the isolate's genome.
sequenced_by_laboratory_name	The specific laboratory affiliation of the responsible for sequencing the isolate's genome.
sequenced_by_contact_name	The name or title of the contact responsible for follow-up regarding the sequence.
sequenced_by_contact_email	The email address of the contact responsible for follow-up regarding the sequence.
purpose_of_sequencing	The purpose of sequencing.
sequencing_project_name	The name of the project/initiative/program for which sequencing was performed.
sequencing_platform	The platform technology used to perform the sequencing.
sequencing_instrument	The instrument type used to perform the sequencing.
library_preparation_kit	The name of the DNA library preparation kit used to generate the library being sequenced.

sequencing_protocol	The protocol or method used for sequencing.
r1_fastq_filename	The user-specified filename of the r1 FASTQ file.
r1_fastq_filename	The user-specified filename of the r2 FASTQ file.
fast5_filename	The user-specified filename of the FAST5 file.
assembly_filename	The user-defined filename of the FASTA file.
publication_ID	The identifier for a publication.

Sequence Information curation tips:

1. Library identifiers are needed for submission to NCBI. If available, provide these under "**library_ID"**. Library IDs can be created based on sample IDs e.g. the sample ID ABCD0123 is assigned a library_ID "ABCD0123_fastq".
2. A sample may be collected for one purpose, but sequenced for another (e.g. collected for surveillance, but sequenced in a research project such as the AMR-GRDI). Sampling strategies and criteria for selection for sequencing can bias results. The reason why an isolate was selected for sequencing can be recorded in two fields, "**purpose_of_sequencing"** and "**purpose_of_sequencing_details"**. Standardized tags are available in the purpose_of_sequencing field, and free text explanations can be included in purpose_of_sequencing_details.
3. Different sequencing chemistries have different biases and can affect results. Record the company name of the platform manufacturer under "**sequencing_platform"** (e.g. Illumina). For "**sequencing_instrument"**, provide the instrument and model number, if known (e.g. NextSeq 550). For "**sequencing_method"**, provide the protocol name and version number used for sequencing (this can be a link to a protocol online).

11. Public Repository Information

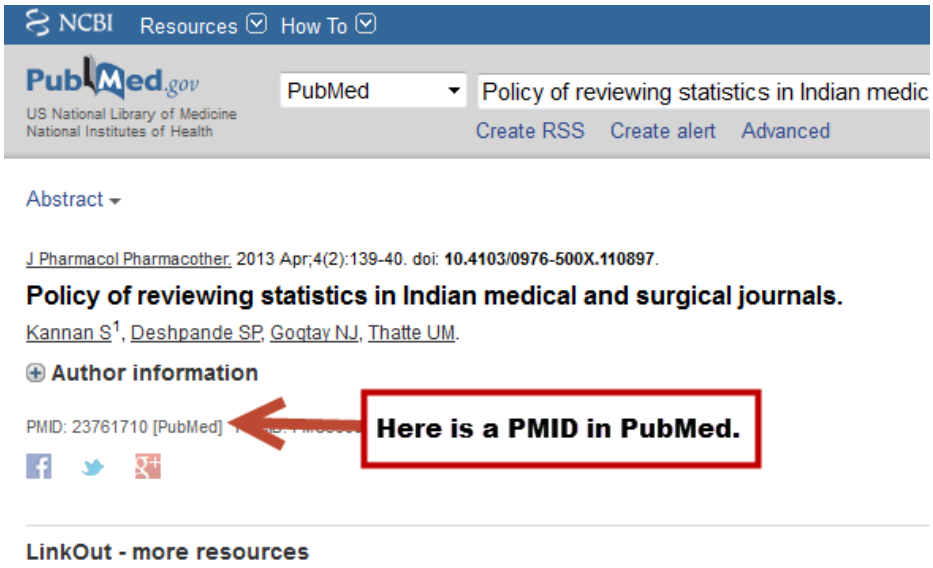
In this section, enter information about the ***sequence submitted to a public repository***. Open data (publishing sequences and contextual data in public repositories) encourages transparency and maximizes the value of sequence data by enabling wider access and reuse for different downstream analyses. If data was submitted to a public repository like the National Center for Biotechnology Information (NCBI), sequence accession numbers, BioSample accession numbers, and publication identifiers should be included.

Field Name	Field Definition
sequence_submitted_by	The name of the agency that submitted the sequence to a database.

sequence_submitted_by_contact_name	The name or title of the contact responsible for follow-up regarding the submission of the sequence to a repository or database.
sequence_submitted_by_contact_email	The email address of the agency responsible for submission of the sequence.
publication_ID	The identifier for a publication.
attribute_package	The attribute package used to structure metadata in an INSDC BioSample.
bioproject_accession	The INSDC accession number of the BioProject(s) to which the BioSample belongs.
biosample_accession	The identifier assigned to a BioSample in INSDC archives.
SRA_accession	The Sequence Read Archive (SRA), European Nucleotide Archive (ENA) or DDBJ Sequence Read Archive (DRA) identifier linking raw read data, methodological metadata and quality control metrics submitted to the INSDC.
GenBank_accession	The GenBank/ENA/DDBJ identifier assigned to the sequence in the INSDC archives.

Public Repository Information curation tips:

1. When submitting to NCBI, there are two types of “attribute package”. An **attribute_package** is a collection of fields customized for the type of sample you are describing. The “Clinical” attribute package should be applied to samples collected from a specific human, animal, or plant (e.g. feces). The “Env” attribute package should be used to describe environmental samples including food, feed, facilities, farms, water sources, manure etc. The only difference between the Clinical and Env packages is that the Clinical attribute package requires the Host and Host Disease fields.
    i. If the sample is from a specific human, animal or plant, put “Pathogen.cl”.
    ii. If the sample is from an environmental sample including food, feed, production facility, farm, water source, manure etc, put “Pathogen.env”.
2. Publication identifiers are commonly DOIs and PubMed IDs (PMID). If the sequence is associated with a published work which can provide additional information, provide a “**publication_ID”**. The PMID can be found at the bottom of a PubMed record as indicated below:



12. AMR Testing Information

In this section, enter information about ***AMR testing results***.
Phenotypic testing results are important for correlating with genomic analytical results.
An antibiogram is a profile of antimicrobial susceptibility testing results of a specific isolate to a set of antimicrobial compounds. The measurement of resistance is crucial to capture, rather than just the interpreted result, as interpretation criteria (such as testing standard e.g. CLSI) may affect results. Be sure to capture the Minimum Inhibitory Concentration (MIC) value and units, as well as the methods used to determine the value.

Provenance information specifying who performed the susceptibility tests, when the tests were performed, and how to contact them for follow-up, should be provided. We have provided structured fields per drug rather than per panel, as different panels can contain different combinations of compounds for testing. Make sure to include information to track the panel and any instrumentation used for automating testing.

Note: Some information is linked to the antimicrobial agent being tested. Agent-specific information (such as measurements, breakpoints and the phenotype) has been structured in the template so that each agent is matched with 14 different fields. These fields are repeated for every agent. The standardized field names are defined once in the Reference Guide (their

permutations in the template are not). If the agent of interest is not already represented, use the New Term Request System to request the addition of other antimicrobial agents (and their corresponding fields).

Example:
antimicrobial_agent_name: amikacin

The amikacin measurement and threshold data should be entered in the template in the fields:
amikacin_resistance_phenotype
amikacin_measurement
amikacin_measurement_units
amikacin_measurement_sign
amikacin_laboratory_typing_method
amikacin_laboratory_typing_platform
amikacinl_laboratory_typing_platform_version
amikacin_vendor_name
amikacin_testing_standard
amikacin_testing_standard_version
amikacin_testing_standard_details
amikacin_susceptible_breakpoint
amikacin_intermediate_breakpoint
amikacin_resistant_breakpoint

If you have multiple tests for each isolate, be sure to link the results to the isolate, and if known, the sample by including all the appropriate identifiers in the AMR Testing Information tab.

Curators should try to fill as many of the following fields as possible, as permitted according to organizational policies.

Field Name	Field Definition
AMR_phenotype_testing_by	The name of the organization that performed the antimicrobial resistance testing.
AMR_phenotype_testing_by_laboratory_name	The name of the lab within the organization that performed the antimicrobial resistance testing.
AMR_phenotype_testing_by_contact_name	The name of the individual or the individual's role in the organization that performed the antimicrobial resistance testing.
AMR_phenotype_testing_by_contact_email	The email of the individual or the individual's role in the organization that performed the antimicrobial resistance testing.
AMR_phenotype_testing_date	The date the antimicrobial resistance testing was performed.

antimicrobial_resistance_phenotype	The antimicrobial resistance phenotype, as determined by the antibiotic susceptibility measurement and testing standard for this antibiotic
antimicrobial_measurement	The measured value of antimicrobial resistance.
antimicrobial_measurement_units	The units of the antimicrobial resistance measurement.
antimicrobial_measurement_sign	The qualifier associated with the antibiotic susceptibility measurement
antimicrobial_laboratory_typing_method	The general method used for antibiotic susceptibility testing.
antimicrobial_laboratory_typing_platform	The brand/platform used for antibiotic susceptibility testing
antimicrobial_laboratory_typing_platform_version	The specific name and version of the plate, panel, or other platform used for antibiotic susceptibility testing.
antimicrobial_vendor_name	The name of the vendor of the testing platform used.
antimicrobial_testing_standard	The testing standard used for determination of resistance phenotype
antimicrobial_testing_standard_version	The version number associated with the testing standard used for determination of resistance phenotype
antimicrobial_testing_standard_details	The additional details associated with the testing standard used for determination of resistance phenotype
antimicrobial_susceptible_breakpoint	The maximum measurement, in the units specified in the “AMR_measurement_units” field, for a sample to be considered “sensitive” to this antibiotic
antimicrobial_intermediate_breakpoint	The intermediate measurement(s), in the units specified in the “AMR_measurement_units” field, where a sample would be considered to have an “intermediate” phenotype for this antibiotic
antimicrobial_resistant_breakpoint	The minimum measurement, in the units specified in the “AMR_measurement_units” field, for a sample to be considered “resistant” to this antibiotic

13. Risk Assessment Information

In this section, enter information about samples and isolates to enhance ***risk assessment***. According to the Codex Alimentarius Commission, a food safety risk assessment is a scientific evaluation of known or potential adverse health effects resulting from human exposure to foodborne hazards. A risk assessment is often quantitative in nature, and requires measurements of pathogens at different stages of production, processing, sale, and consumption. To facilitate collection of this information, fields are provided to capture reference material (e.g. sample plans) and metrics (e.g. prevalence data) enabling risk assessment.

If available, curators should try to fill as many of the following fields as possible, as permitted according to organizational policies. If unavailable, leave the fields empty.

It should be noted that these fields and this section are under active development.

Field Name	Field Definition
prevalence_metrics	Metrics regarding the prevalence of the pathogen of interest obtained from a surveillance project.
prevalence_metrics_details	The details pertaining to the prevalence metrics from a surveillance project.
stage_of_production	The stage of food production.
experimental_intervention	The category of the experimental intervention applied in the food production system.
experiment_intervention_details	The details of the experimental intervention applied in the food production system.

14. How to Request New Standardized Terms and Data Validation

Data standards are living specifications and according to best practices, should evolve over time in order to be fit-for-purpose. To better harmonize information collected in the GRDI-AMR template, pick lists of ontology-based vocabulary have been provided for many of the fields based on real datasets and user requests. These pick lists reflect data needs at a snapshot in time and will need to be updated periodically and version-controlled.

Versioned templates and associated reference guides and SOPs are available at the GRDI-AMR GitHub repository under a MIT Data Use License (unrestricted use).

While ontology look-up services are available online for users to identify additional terms they would like to use, pilot implementations suggest that unrestricted term additions create variability and uncertainty in the specification, as well as unnecessary burden on curators. As a result, a New Term Request System has been put in place so that users can request new fields and terms which will be processed by the Centre for Infectious Disease Genomics and One Health (CODGOH, Simon Fraser University) curation team on an ongoing basis.

If a desired term cannot be found in a pick list, new terms can be requested by following the New Term Request System SOP. Any question about this process should be directed to Dr. Emma Griffiths at ega12@sfu.ca.
Data validation is an important best practice of data management. The GRDI-AMR template currently offers validation support by enforcing pick lists and required formats (such as ISO 8601 date formats). While a user can enter non-standardized values into fields where data entry rules are enforced, a warning will appear to flag validation issues. Additional data management tools that offer improved validation and data transformation capabilities will soon be available.

15. Uploading the Summary Sheet to IRIDA

The final step of curating contextual data for the GRDI is uploading it to IRIDA (irida.ca). You or your team members should have already created a project specific for the GRDI.

To upload harmonized contextual data to your Project in IRIDA, use the IRIDA Metadata Uploader according to the following steps.

1) Select all of the column titles and values in the "Merged Data" sheet (second tab in the template file). Then copy this selection to your clipboard by right clicking on it and selecting "Copy" or by typing Ctrl+C. *Note:The last column of the Merged sheet is "OO".*
2) Open a new file in Excel, and right click on the first cell (A1). Under "Paste Options", select "Values". The metadata should now appear in this new file.
3) In the window with the copied metadata, under "File", select "Save As". Make sure that the "Save as type" is "Excel Workbook" or ".XLSX". Save the file with the name and location of your choice.
4) Use the IRIDA Metadata Uploader to import your contextual data into the GRDI Project using the instructions provided here:
https://phac-nml.github.io/irida-documentation/user/user/sample-metadata/

Note:
The IRIDA uploader only accepts Excel files, not csv files.
The IRIDA uploader only looks in the first worksheet in an Excel file

The IRIDA uploader reads the empty lines as invalid samples, but it separates those from the valid samples.

16. Deprecating and Normalizing Ontology Values & IDs

There are many reasons why a term may be rendered “obsolete” and consequently deprecated in favour of a different term. It may be that it was rehomed in a more appropriate ontology or that it was deemed redundant with an existing term. During specification development, we usually need to generate new terms quickly - which is relatively easy to do in ontologies we manage, but more difficult when requesting terms in ontologies within the greater OBO Foundry community. Long term, the formal deprecation of terms enables downstream ontologies that import them to pass this information on to database systems so they are in a position to update their own contents and therefore support federated querying without the need to map terms.

To help users identify a term match and update their conceptual data when a specification term has changed, we have included “Deprecated Label” and “Deprecated ID” columns in all reference guides. This is to make it easier for users to “match” old terms with the new version without having to query an ontology. Please note that this is only done when strictly necessary, as we avoid deprecating terms whenever possible. If a term changes in its fundamental meaning (e.g., the core concept being described), a new picklist term will be created.

For more information reference:
https://github.com/cidgoh/GRDI_AMR_One_Health/wiki/Deprecating-&-Normalizing-Specification-Values-&-IDs

Appendix A: Document Revision History

Version	Date	Writer	Description of Change
1.0	May 6 2020	Emma Griffiths, Julie Shay	Initial release
1.5	November 2 2020	Emma Griffiths	Modified for round 2 of curation, only updates included (original instructions removed)

2.0	February 22 2020	Emma Griffiths	Completely rewritten based on curator feedback and updated in parallel with collection template. Field-level guidance largely stored in Reference Guide.
2.1	July 9 2021	Emma Griffiths	Refined text and organized into sections (included Table of Contents). Added instructions for upload to IRIDA.
2.2	September 7 2021	Emma Griffiths	Edited text to reflect curator feedback.
3.0	July 11 2022	Emma Griffiths, Rhiannon Cameron	New fields and sections added. AMR information restructured. Pick lists in template expanded. Instructions for new term requests included.
4.2	November 14 2022	Emma Griffiths, Rhiannon Cameron	Added new fields and instructions to “Strain and isolation information”.
5.3	December 02 2022	Rhiannon Cameron	Added “16. Deprecating and Normalizing Ontology Values & IDs”. Removed “AMR_measurement” field.
6.4	December 15 2022	Emma Griffiths	Added food quality and packaging date fields to “Sample Collection and Processing Information” and updated user instructions.
7.6	May 26 2023	Emma Griffiths	Added additional environmental fields, “available data types”, and “production stream” to “Sample Collection and Processing Information”. Expanded guidance for formatting alternative sample and isolate identifiers. Expanded guidance for providing agency vs lab names, as well as contact information. Added “host age bin” to “Host Information”.

			Expanded guidance for updating organism names when sequencing results reveal different taxonomic designations, as well as instructions for naming organisms with known taxonomic designations to the genus. Expanded guidance for restructured AMR profiling information.
7.7	Sep 29 2023	N/A	Version compatibility update
8.8	Oct 25 2023	N/A	Version compatibility update
8.9	Dec 1	N/A	Version compatibility update