Implementing ISO 23418:2022 for data management in laboratories

Dr. Peter Evans[1] & Dr. Emma Griffiths[2]

[1]FSIS USDA, USA & [2]Simon Fraser University, Canada

EURGen-RefLabCap

March 12 2024

Outline

1. History and goals of ISO 23418

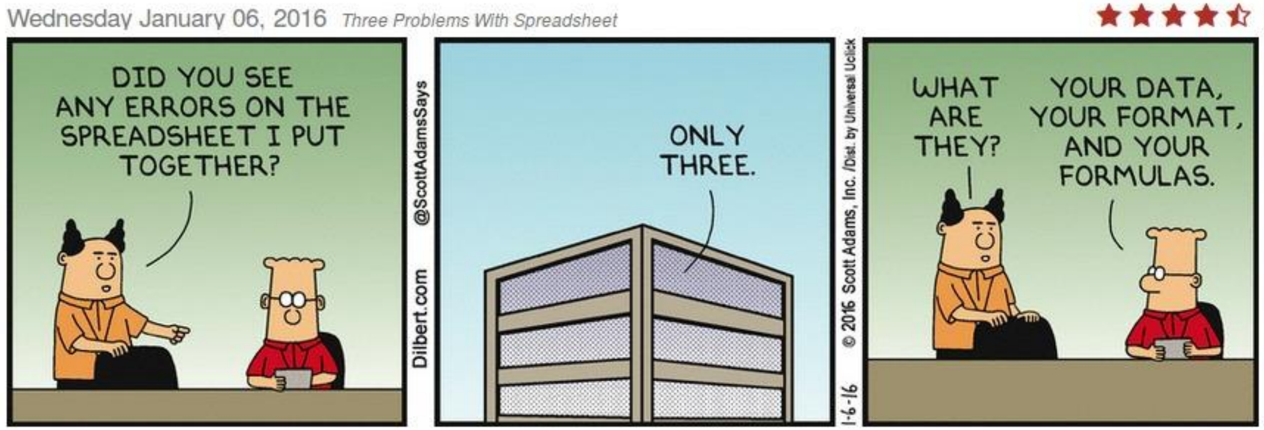
2. Contextual data ("metadata") overview and best data standard design practices

3. Canadian implementation – modularization, customization, reuse

4. Tools & databases

5. What's next for ISO

6. Wrap up



ISO Principles

General

- Identified market need: improve quality, consistency, reproducibility
- Consensus
- Several rounds of international review, feedback, voting

Data Management and Sharing

1. **Clear meaning** (human and machine readable)
2. **Interoperability** (different datasets, systems, processes)
3. **Harmonization** (no organization-specific terminology should take precedence)
4. **Flexibility** (recognizing different needs in different lab settings)
5. **Maximizing utility of data** (prioritizing information types/structures)
6. **Best semantic practices**

Challenges of Status Quo in Standards Development

Expert consultation

Use case-specific vocabulary

- Narrowly scoped (limits interoperability)
- Organization-specific terminology
- Lack of semantic best practices (no rules, impacts machine-readability)
- Abbreviations, inconsistent structure, word bombs, highly composite terms

ISO terminology like a "common currency" for communication in genomics and beyond

Why don't we just use "X" vocabulary?

- Organization-specific vocabularies
- Clinical and regulatory terminologies (SNOMED, LOINC, FHIR HL7, CDISC)
- Public repository/public database requirements
- Sectors (One Health – animal, human, environment)
- Industry
- Research

Ontologies: Built for harmonization and data linkage

Controlled (standardized) vocabulary

Hierarchy + logic (linked data, enable classification for analyses)

Universality

- Meanings disambiguated with URIs
- Labels/Synonyms (organization-specific/interoperability)
- Principles and practices to enable reuse (BFO, RO)

Community

- Community of practice (OBO Foundry, >200 interop ontologies)
- **Registries/Portals** (EBI OLS, Ontobee, BioPortal)
- **Languages/Tools** (Protégé, LinkML, Robot, OntoFox)

Lager beer [BeerO:1234]

FAIR

Hausenblas & Kim (2012)
Berners-Lee (2009)

5-star Open Data Plan

★ Make your stuff available on the Web (whatever format) under an open license

★★ Make it available as structured* data (e.g. Excel instead of an image scan of a table)

★★★ Make it available in (2+) non-proprietary open format (e.g., CSV instead of Excel)

★★★★ Use **URI**s to denote things, so that people can point to your stuff

★★★★★ Link your data to other data to provide context

FoodOn:455678

VS

ENVO:009747

Standards: ISO 23418:2022

Microbiology of the Food Chain — Whole genome sequencing for typing and genomic characterization of foodborne bacteria — General requirements and guidance

Contextual Data Fields
Sample Collection Lab Contact Information Geographic Location of Sample Collection Collection Date Sample Type Food Product Food Processing Environmental Material Environmental Location Collection Device Collection Method Microbiology Lab Contact Information Organism Strain Isolate Serotype Isolation Media Isolate Passage History AMR & Virulence phenotypes

ISO standard provides tables and annexes to describe...

1. Information about the **sample**
2. Information about the **isolate**
3. Information about the **sequence**

ISO slim (package of fields and terms) available:
https://github.com/GenEpiO/iso2017

Fields and terms sourced and adapted from:

- Agency documentation
- Public repository submission forms
- Domain expert consultations
- Existing standards and ontologies

How organizations implement ISO 23418 for metadata management is up to them.

- Makes **recommendations**, not laws
- Depends on organization’s infrastructure, capacity, goals, roles

- What we can do today, is give you **options**

- **Examples** of successful implementations

No one size fits all solutions

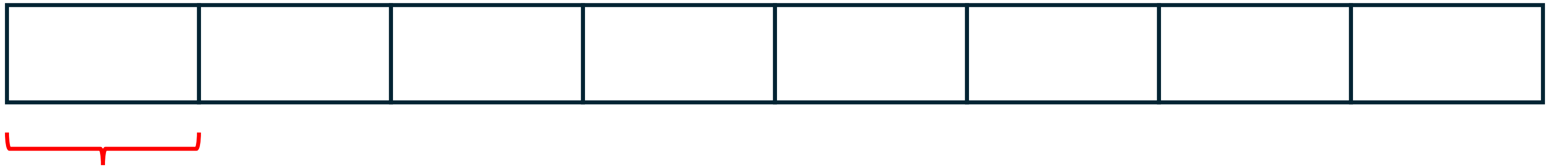
Emma will describe how ISO 23418 has been used in Canada

Canada: Federated system, patchwork of jurisdictional powers

- **10 provinces, 3 territories** (Federal, Provincial/Territorial, Municipal jurisdictions)
- **Many federal agencies** with different departments e.g. **Health Canada, Public Health Agency of Canada, Canadian Food Inspection Agency**, **Agriculture & Agri-Food Canada, Environment and Climate Change Canada, Department of Fisheries & Oceans**
- **Many labs accredited for ISO 17025**
- Most microbiological assays and PCR, increasingly genomics
- Most capacity at federal level, increasing at provincial level as well as local level (e.g. hospitals)
- Most labs have own data management solutions, data sharing difficult, international standards help

Putting ISO 23418 into Practice: Pathways to Implementation

Modular framework and core content (ISO 23418:22)



Modules expanded, populated with fields/terms from community-driven ontologies

- Flexible, extensible, customizable, interoperable
- Apply to different use cases/pathogens/programs (**Federal/Provincial**)
  - SARS-CoV-2 (pandemic)
  - MPOX (epidemic)
  - Wastewater
  - One Health AMR
  - *LIMS modernization (NML e-reqs, intake/storage)
- Different technical implementations

(*not accredited)

Thematic Modules:

Database identifiers

Sample collection and processing

Isolate information

Sequencing information

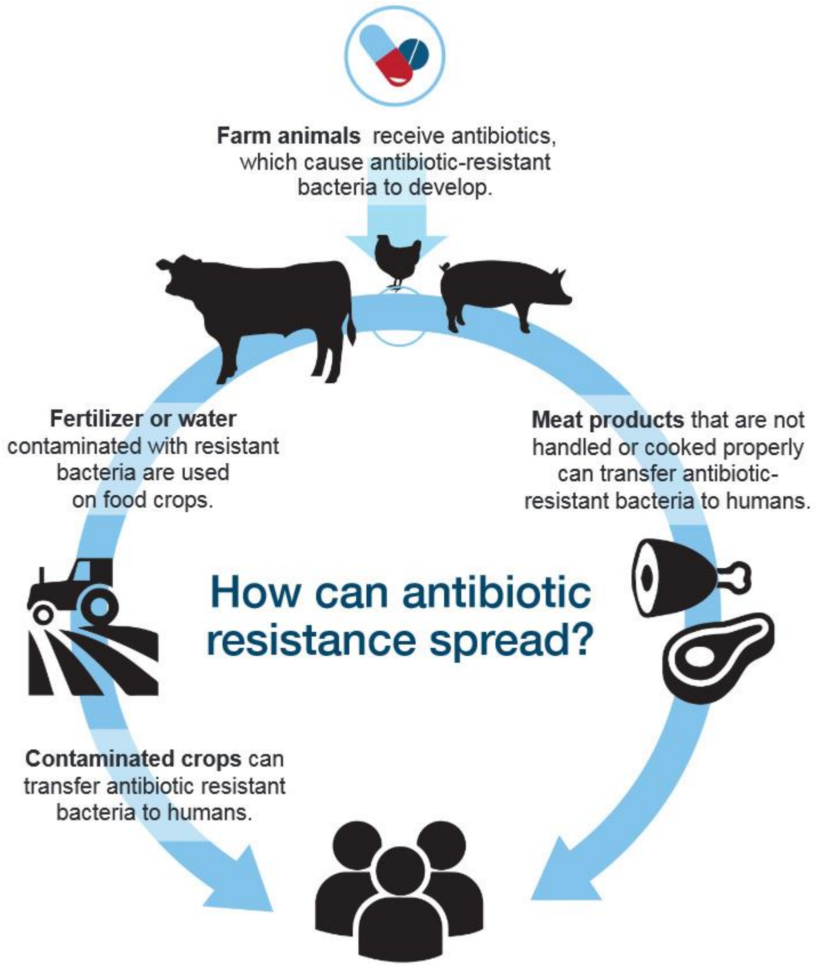
Bioinformatics & QC metrics

AMR testing information

Provenance & acknowledgements

GRDI-AMR standard: ISO-based specification for One Health Antimicrobial Resistance (AMR)

GRDI-AMR: Genomics and Research Development Initiative to support Canada's federal AMR action plan



- Based on **ISO framework**
- Scope: Bacteria. **WGS across sectors, commodities, environments, hosts**

- Goal: use genomics and harmonized contextual data to understand foodborne **AMR** in **food supply and environment,** identify interventions

- **Canadian implementation: Federal Interagency** (PHAC, CFIA, AAFC, ECCC, DFO, HC etc)
- **also international sharing Uganda, Canada-UK sharing**

https://github.com/cidgoh/GRDI_AMR_One_Health

Adapting ISO Framework for One Health AMR: customized modules and content

Domain Content

- Repository accession numbers and identifiers
- **Sample collection and processing**
    - Food **products**
    - Food **processing**
    - Host/food **geo-loc origin** vs **sampling location**
    - **Environments** (abattoir, farm, natural enviros, fisheries)
    - Environmental **materials** (chicken litter, sediment, water, soil)
    - **Anatomical** parts/sites (feces, organ contents)
    - **Presampling activities** (fertilizer, vaccination, decontamination)
    - **Sampling/sequencing strategies** (bias/limitations)
- **Isolate information**
- **Host information (animals, plants, humans)**
- **Sequencing methods**
- **Bioinformatics and quality control metrics**
- **AMR phenotype testing**
- **Risk assessment**
- **Provenance and attribution**

Standardized null values (INSDC)

Standardized fields & Picklists (can be updated)

Support docs (ref guide/SOP)

Operationalized using data curation tools

****being integrated across federal genomics ecosystem**

Technical Implementations – Tools & Databases

Different ways to implement the standard for data management.

1. Spreadsheet-based templates and tools

Implement ISO-compliant standard (as-is)

e.g.
DataHarmonizer

2. Existing Systems

- Mapping
- Automated transformations, development of interchange formats, focus on interoperability

e.g. mapping/interchange:
- National Microbiology Laboratory LIMS (Public Health)
- CIPARS (Canadian Integrated Program for AMR Surveillance)
- INSDC BioSample packages

3. New Systems

Implement ISO-compliant standard (as-is)

e.g.
- CFIA Genomics db
- Virtual Microbial Resource (graph db)

Spreadsheet-based templates: The DataHarmonizer

- Javascript application
- Download locally, all prov public health labs have local copy
- Extensible, as many templates as needed (under active dev)
- Colour-coding, picklists, curation features, validation

View all fields
View required fields
Move to desired field

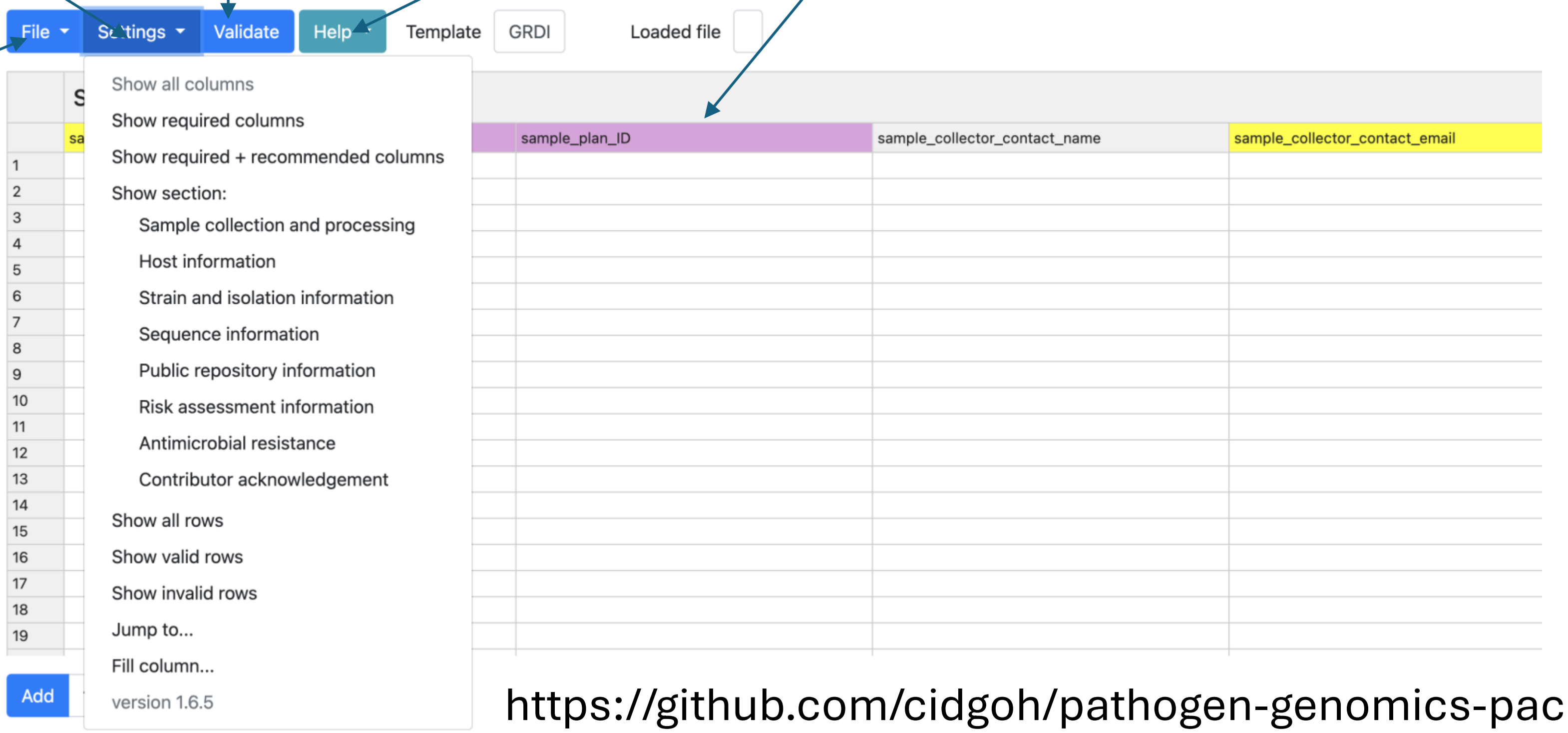
Validate
(check for errors
or missing info)

Learn your
way
around the
system

Double click on
field labels for
guidance on
how to fill them

*Data upload format: xlsx, xls, tsv and csv

Save
Open existing file
Export to chosen format



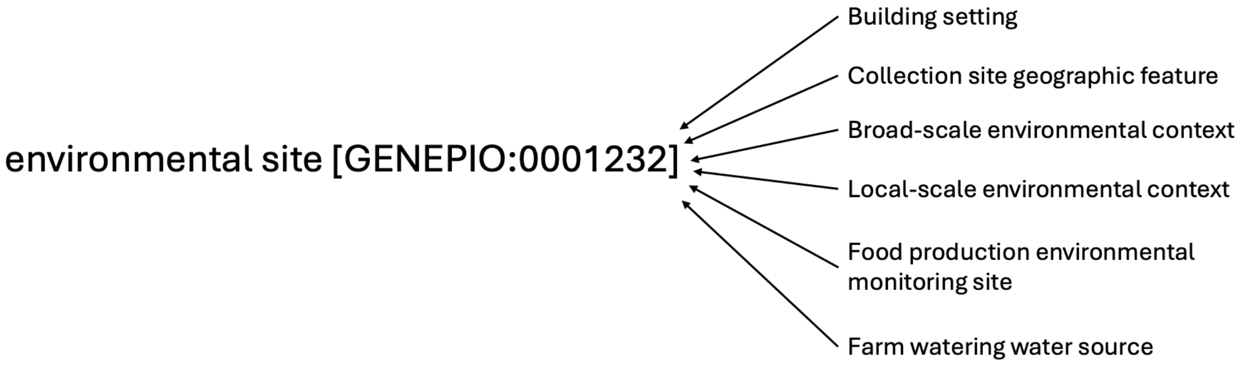
https://github.com/cidgoh/pathogen-genomics-package

Adapting existing systems: Creating **bridges across systems** with **mapping and exchange formats/tools.**

Examples of mapping ISO vocab to other dictionaries/schemas to create “common terminology currency”.

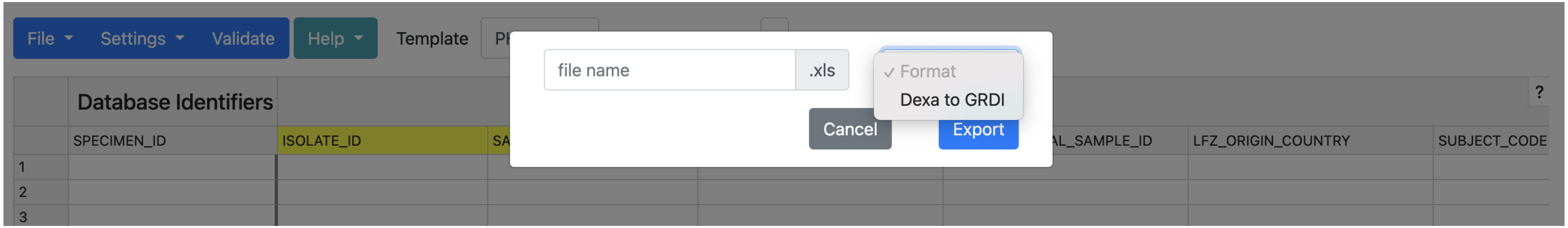
LAB LIMS	**STANDARD**
TEXT_ID	specimen collector sample ID
CUSTOMER	sample collected by
HC_COUNTRY	geo_loc_name (country)
PH_TRAVEL	destination of most recent travel (city)
PH_TRAVEL	destination of most recent travel (state/province/territory)
PH_TRAVEL	destination of most recent travel (country)
PH_TRAVEL	most recent travel departure date
PH_TRAVEL	most recent travel return date

Canada Federal LIMS: Standard



US One Health NCBI BioSample Package: Standard

Mapping and interchange formats enable automated transformations (***ref lab*** formats → ***community*** formats → ***downstream*** formats).



e.g.

CIPARS (DEXA) → GRDI → NCBI BioSample

CIPARS (DEXA) → GRDI → NML LIMS format

GRDI standard acts as linker “ground truth”.

Enter data once, export for different uses!

***data is not stored in the DataHarmonizer, only processed**

** Easier to develop tools/dbs for standardized, widely used schemas

Microbial Genomics, 2022, 10.1099/mgen.0.000908) https://github.com/cidgoh/pathogen-genomics-package

New databases & agency integration: The Canadian Food Inspection Agency Genomics Database

CFIA·ACIA

Azure SQL

- SQL database in an Azure cloud with an Access db front end
- ISO-based GRDI specification forms main schema
- Synonyms captured in tables, some additional customization
- Goal: link directly to agency’s cloud computing so metadata can be used in different ways

Virtual Microbial Resource

CFIA·ACIA

Public Health Agency of Canada
Agence de la santé publique du Canada

- *"Tinder for microbial collections"* (making data matches between organizations)
- **Graph database** for hosting data, analysis → **schema based on ISO-based GRDI specification**

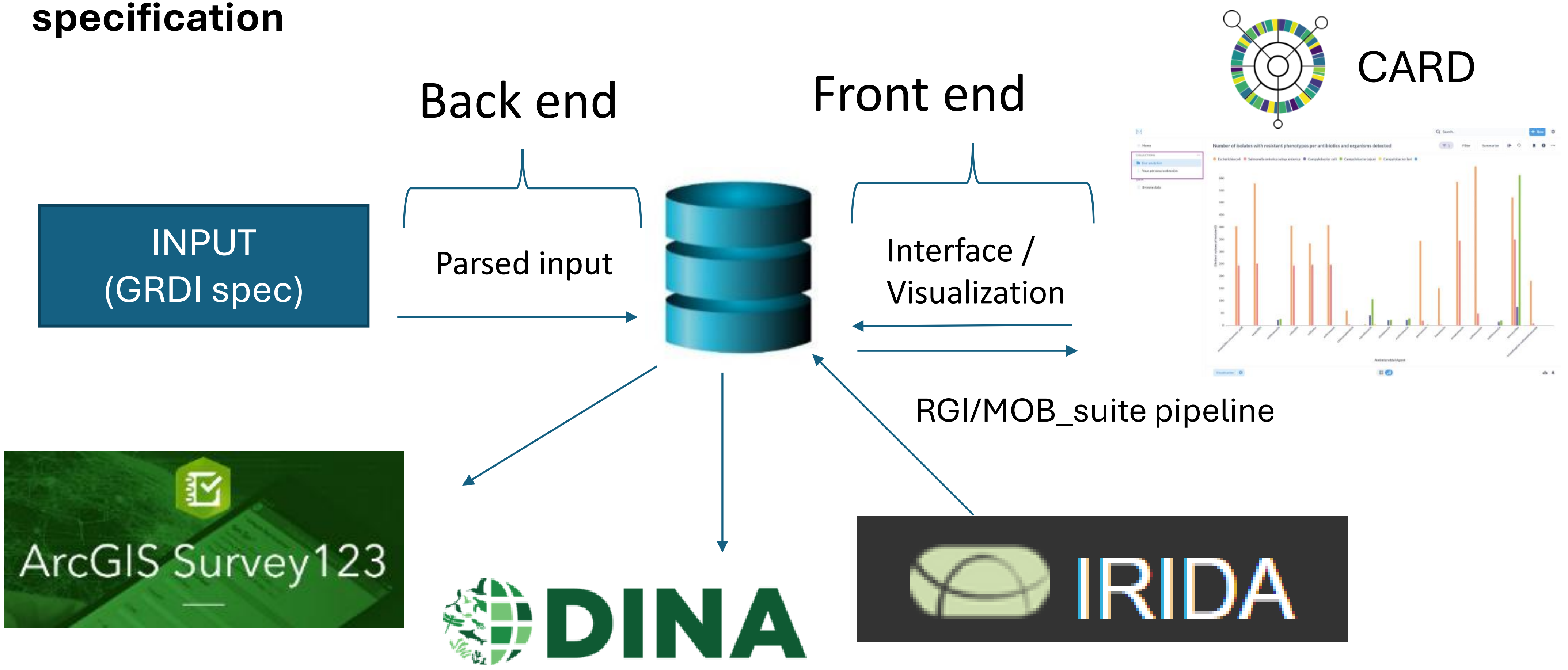


Figure courtesy of Gabriel Wajnberg (CFIA)

Public Health Alliance for Genomic Epidemiology

International adoption of the ISO framework

- Global, volunteer organization
- >200 members, 90 organizations, 30 countries

Scope:

- Reproducibility, interoperability, portability, capacity for public health bioinformatics

Working Groups:

1. Data Structures
2. Infrastructure
3. Pipelines & Visualization
4. Training & Workforce
5. Ethics & Data Sharing

Reduce the **barrier to entry** for routine sequencing	Promote **innovation, collaboration and development** from public/private sector
Promote **standardization, portability and reproducibility** of assays and workflows	Foster the **development/resiliency of the global public health bioinformatic workforce**
Advance the use of **open data and open source** in public health	Enable global public health to **rapidly adapt to changing priorities and emerging threats**
Improve **surveillance and outbreak response** capabilities	Empower more labs to **analyze/govern their own data, regardless of resource status**

https://www.github.com/pha4ge

https://www.pha4ge.org

@pha4ge

@pha4ge@@mstdn.science

BILL & MELINDA GATES foundation

Enshrined ISO-based Framework in International Specifications

Customized framework:
PHA4GE SARS-CoV-2 contextual data specification
https://github.com/pha4ge/SARS-CoV-2-Contextual-Data-Specification

PHA4GE Wastewater contextual data specification
https://github.com/pha4ge/Wastewater_Contextual_Data_Specification

New Modules:
PHA4GE PCR primer amplicon scheme specification
https://github.com/pha4ge/primer-schemes

PHA4GE QC tag specification
https://github.com/pha4ge/contextual_data_QC_tags

PHA4GE hAMRonization specification (AMR detection across widely used tools)
https://github.com/pha4ge/hAMRonization

Public Health Alliance for
Genomic Epidemiology

Rewiring & Modernizing LIMS at the NML (national reference lab)

PAGERR – Pathogen Agnostic Genomic Electronic Requisition and Reporting

- **Extra modules** added to ISO-based framework
- **Ontology approach**
- **Streamline** data intake and storage
- **Harmonize** across disease/pathogen areas
- Better **data integration/analysis**

"Upgrading the plumbing and re-wiring the NML as we support the expansion of genomics by implementing e-requisitions and reporting to modernize infectious disease detection and surveillance."

Summary: ISO 23418 provides a quality framework for your contextual data

- Improves **auditability** (e.g. chain of custody)
- **Provenance** and **acknowledgement**
- Streamlines **re-use** and **data sharing**
- **Reduces uncertainty**
- Creates **expectations** for structure, **requirements**, and **completeness**
- Can **reuse** curation **training/skills, tools, also agreements**
- **Future-proofs data**

What's Next for ISO?

ISO TC 34 / SC 9
Ad'hoc Group G 5
"Antimicrobial resistance brainstorming"

Mandate:

Investigate the need and feasibility to launch standardization work on AMR of bacteria, based on sequencing with a One Health perspective

Invited experts:

- ISO TC 34 / SC 16 (is Standardization of biomolecular testing methods applied to foods, feeds, seeds and other propagules of food and feed crops)
- ISO TC 212 (Medical laboratories and in vitro diagnostic systems)
- ISO TC 276 (Biotechnology)

Acknowledgements

ISO TC34/SC9/WG25

US Department of Agriculture

US Food & Drug Administration

Centre for Infectious Disease Genomics and One Health (SFU)

Public Health Agency of Canada

Canadian Food Inspection Agency

GRDI-AMR

Public Health Alliance for Genomic Epidemiology (PHA4GE)

Thank you for listening!