# Standards Architecture for the Pathogen Genomic Surveillance Enterprise:

# Workshop Learnings and Outcomes

Marc Salit & Arend Sidow, Viridae, Inc.
in partnership with The Rockefeller Foundation

# Consultation with CSWG & SPHERES on Standards Architecture for Genomic Surveillance

- Pathogen Genomic Surveillance working definition:
  - integrative analysis of viral whole-genome sequencing from individual human-derived samples
- Vision of Standards Architecture
- RF Workshop Takeaways

# Why a Standards Architecture?

- Reliable surveillance needs systematic 'enterprise'
  - robust, integrated, interoperable, verifiable, trustworthy, and reliable
- Enterprise is composed of multiple distinct operations
  - conducted by multiple organizational entities
- "Standards Architecture" is the full suite of standards to enable systematic operation

# What are the roles of standards in enterprise performance?

Provide a way to evaluate performance

- "how well can I trust these data?"
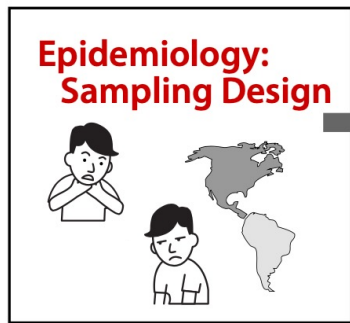- what's the confidence interval?

Provide the basis of interoperability

- "I can share these data…" and someone else can use them, or compare to them
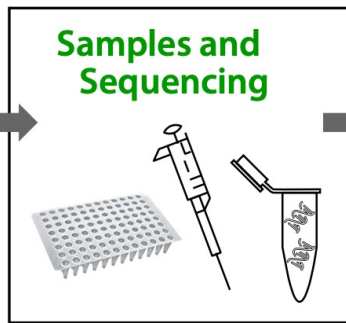
# Invited community to consider *Standards Architecture*

- Shared and reviewed 'strawperson' in 4 open, public working group meetings
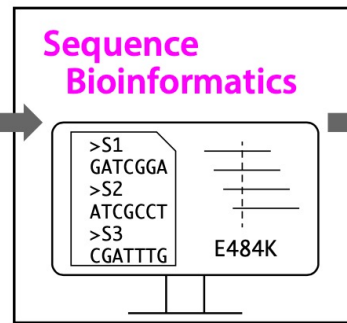- Heard substantive feedback, recommendations, discussion of key matters in each group
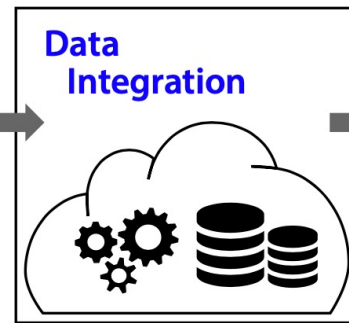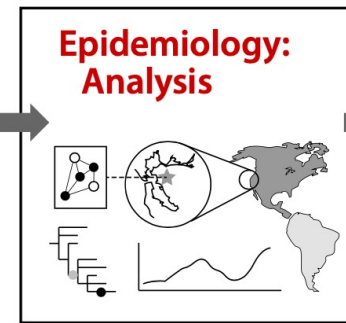


**Process Category**

**Epidemiology: Sampling Design** → **Samples and Sequencing** → **Sequence Bioinformatics** → **Data Integration** → **Epidemiology: Analysis** → **Actionable Public Health Knowledge**

Which strain at which prevalence?
What is the geographic spread?
Demographic variations?
How did transmission occur?
Speed of transmission?
Vaccine escape?
Diagnostic performance?
Therapeutic implications?

WG 4    WG 1    WG 2    WG 3    WG 4

# Diverse Working Group Engagement

- Public/Private/Academic/Non-profit

- Public Health labs, research institutions, clinical labs, technology developers, professional societies, funders

- Laboratorians, informaticians, tool makers, scientists, epidemiologists, standards developers

People came and stayed



Time in meeting

WG  ☐ 1  ☐ 2  ☐ 3  ☐ 4
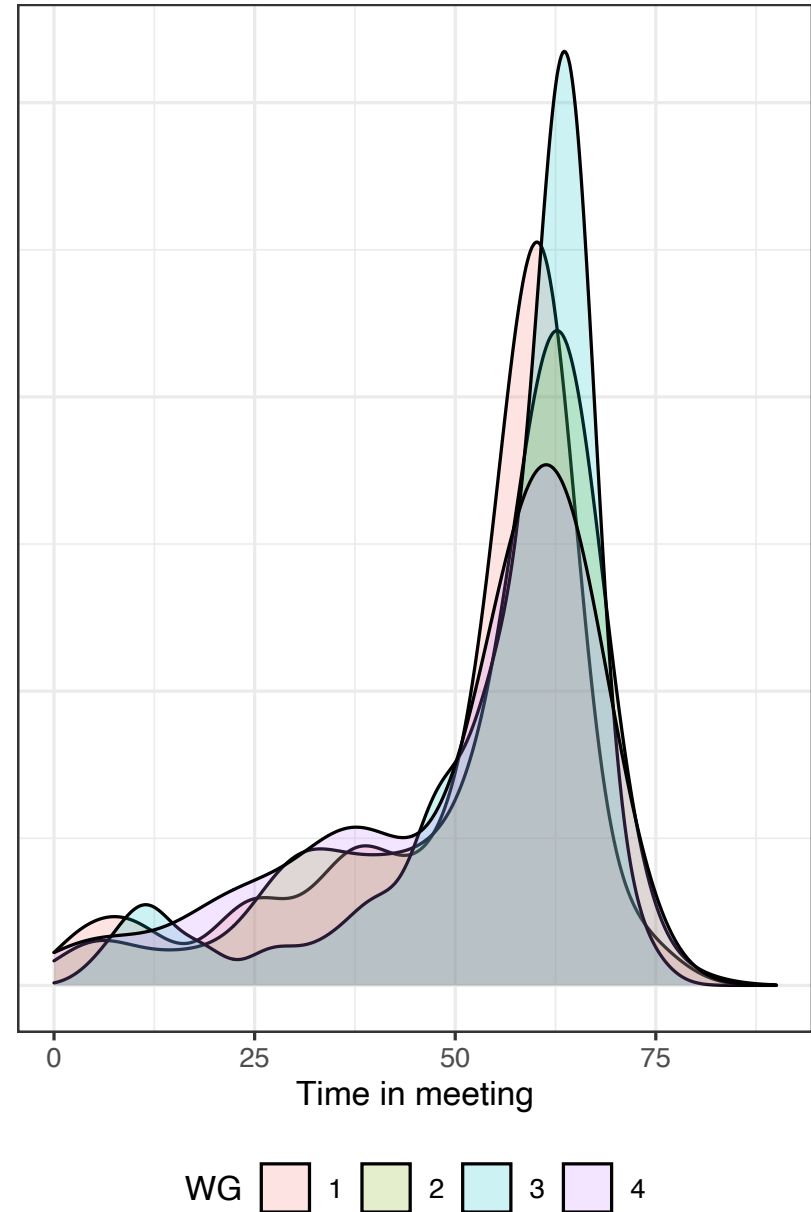
# Pathogen Genomic Surveillance Process Model
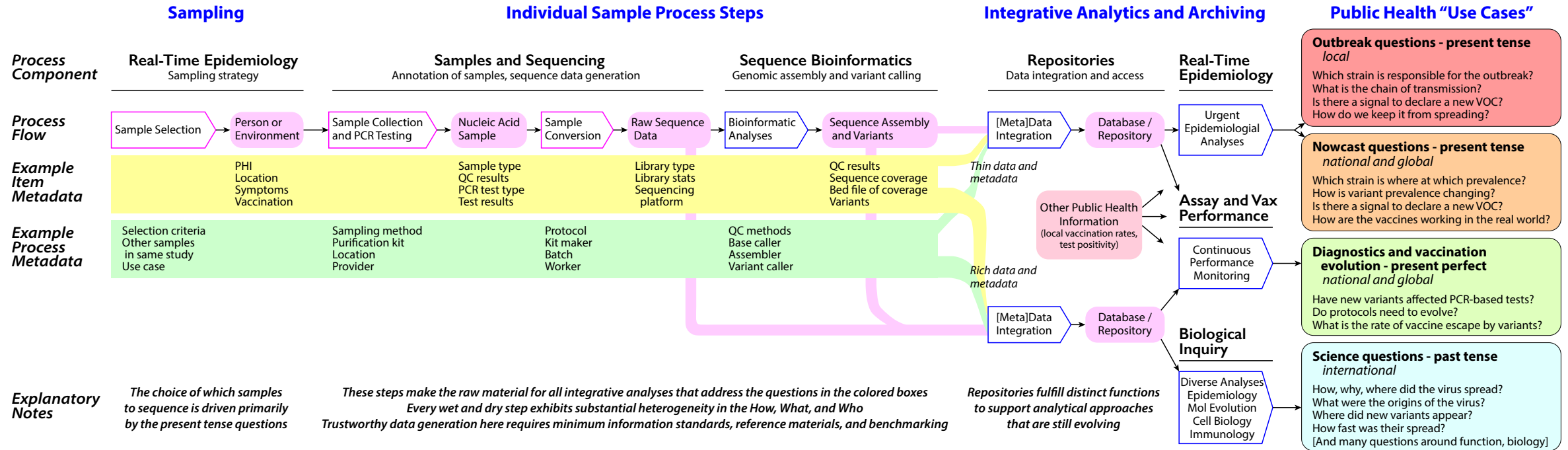
# Overarching Takeaway I:

# Implications of *Diverse Use Cases*

- ***Diversity of Use Cases*** drives diversity of <u>Sampling Strategy</u>, <u>Sharing</u>, <u>Analysis</u>
  - global questions of strain prevalence, spread of variants
  - local questions of outbreak identification
  - novel variant / strain detection
  - relevance/willingness to share data
- Takeaways
  - *Metadata Annotation* – one size will not fit all
  - *Standards Development* should be cognizant of critical use cases

# Overarching Takeaway II:

## Implications of *Diverse Jurisdictions*

- Local, regional, national, international legal and cultural contexts create different requirements and constraints for
  - data sharing (ownership issues)
  - metadata annotation (privacy issues)

- Takeaway
  - Standards Development needs to accommodate jurisdictional contexts

# Overarching Takeaway III:

## *Diversity of Sample Handling and Measurement*

- Wide variety of methods and technologies
  - processing samples, sequencing, and sequence bioinformatics
- Takeaways
  - need *technology-agnostic* standards
  - need repository/library of *validated* tools, protocols, pipelines
  - need authoritatively characterized *Reference Materials and Data*
    - for optimization and validation of methods and analysis
    - for benchmarking and proficiency testing

# Overarching Takeaway IV:

## *Metadata annotation, aggregation, integration is hard*

- Metadata annotation is costly
  - not always budgeted
  - burden carried by "upstream" labs, not those analyzing

- Make and deploy better tools
  - tooling to annotate is immature
  - metadata sharing from lab-to-lab or process-to-process is hard
    - e.g., interoperability of *Lab Information Systems* and *Electronic Health Records* is uncommon

## Overarching Takeaway V:

## *Pathogen Genomic Surveillance* is an evolving discipline

- Pathogen Genomic Surveillance ecosystem is evolving
  - sample handling, sequencing, sequence bioinformatics more mature than epidemiological analyses

- Standards, shared resources, and the *Standards Architecture*
  - must be designed to evolve
  - must have continued engineering to be maintained

- *Can't hide behind this and eschew rigor*

# Potential Standards

Set of widely available, authoritatively characterized reference samples

Reference data for sequencing and sequence bioinformatics benchmarking

Open, public, secure benchmarking platform

Widely adopted metadata standards

Benchmarking challenge for integrative analysis