T3connect

How to design and deliver pathogen genomics training for health and research professionals

Module 3A: Specimens: Collection, Preservation, Processing and metadata

Dr Wilber Sabiiti and Dr Charlene Rodrigues

Date: 7th March 2023

wellcome connecting science

Centre for Genomic Pathogen Surveillance

BIG DATA INSTITUTE

UNIVERSITY OF OXFORD

Learning outcomes

1. List the do’s and don’ts in specimen collection and processing
1. Distinguish specimen requirements in outbreak and surveillance investigations.
1. Map the specimen collection and processing pathway for genomics.
1. Identify the types of metadata needed in different scenarios.
1. Identify ways to effectively teach data/specimen collection and processing.

wellcome connecting science

T3connect

Centre for Genomic Pathogen Surveillance

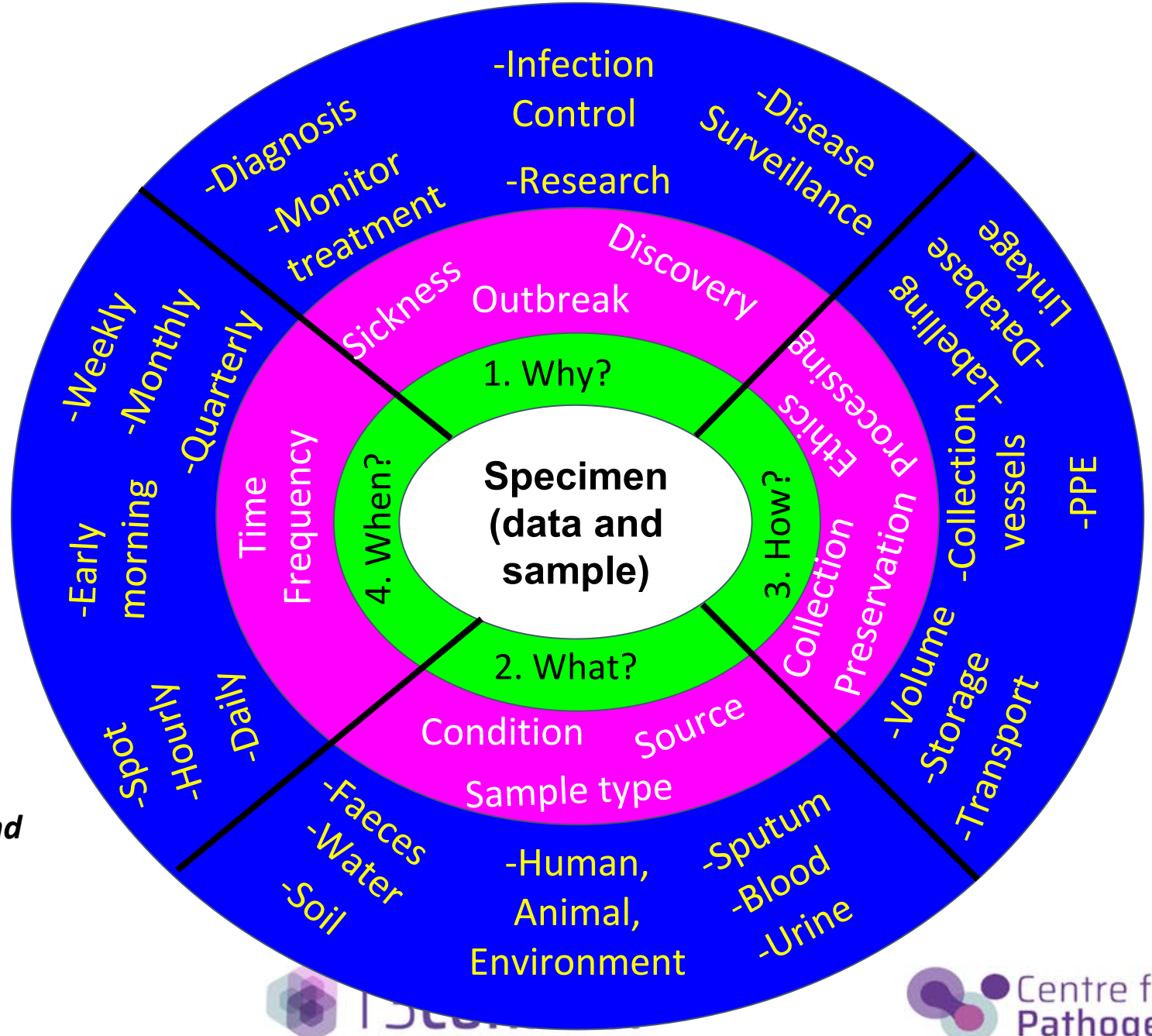
Part 1.
What, why, and how of specimen and data collection and processing

wellcome connecting science

T3connect

Centre for Genomic Pathogen Surveillance

Part 1 Best practice in sampling and data collection

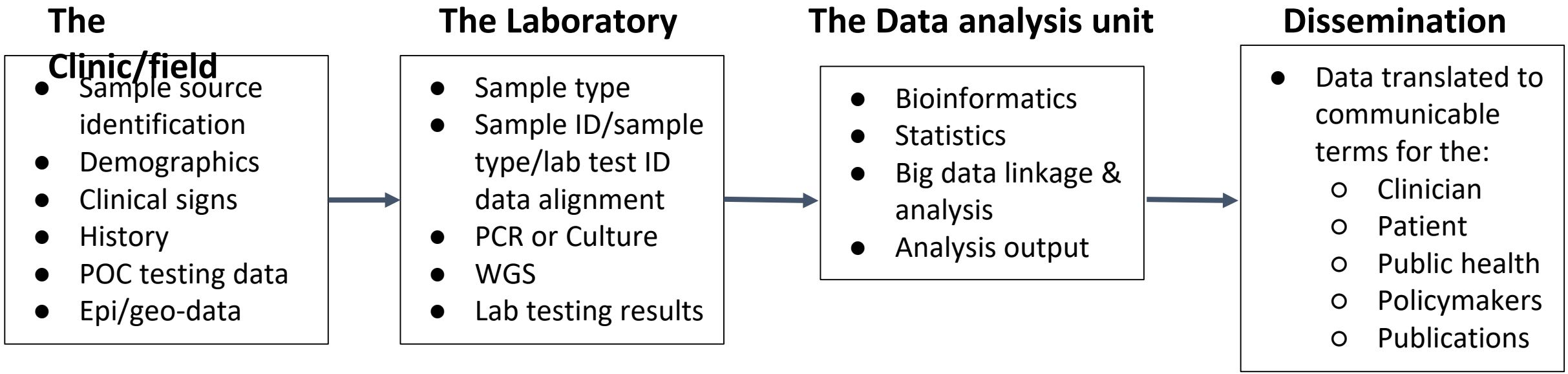


Data collection starts before the actual sample is taken and goes all the way to data analysis & dissemination.

wellcome connecting science

Centre for Genomic Pathogen Surveillance

Part 1: Sample & Data collection pathway



Data collection starts before the actual sample is taken and goes all the way to data analysis & dissemination.

wellcome connecting science

T3connect

Centre for Genomic Pathogen Surveillance

Part 1. Lessons from the Video

What good specimen practice procedures are observed in the video?

https://youtu.be/yRSEB3Gevvo

- What are the dos and don’ts in the video?
- Can you identify the why, what, how and when?

wellcome connecting science

T3connect

Centre for Genomic Pathogen Surveillance

Part 1. Sampling for pathogen genomics

- Define reason (s) for sampling
  - Outbreak
  - surveillance
  - Infection control
- Define the sample type
  - Pathogen type
  - Clinical sample or culture isolate
- Define methodology for sampling
- Define the sampling time frame

wellcome connecting science

T3connect

Centre for Genomic Pathogen Surveillance

Part 1. Parameters that affect genomic sample quality

- Source sample
  - Type - is it the right type?
  - Volume - is the quantity sufficient
  - Cross contamination - wrong pathogen
- Nucleic acid quality
  - Extraction process - DNA or RNA?
  - Chemical contamination - ethanol, lysis buffers, preservatives
  - Degradation - RNases, freeze & thaw etc

wellcome connecting science

T3connect

Centre for Genomic Pathogen Surveillance

End of Part 1. Reflect - concepts

What concepts do we need to cover?

Why do we need to cover them?

How are we going to teach them?

Part 2.
Breaking down concepts into sub-topics

wellcome connecting science

T3connect

Centre for Genomic Pathogen Surveillance

Part 2. Specimen and data collections and processing - Teaching strategies

Domain concepts	Content details	Training strategies and Assessment
Identifying which data/samples are needed	Content involves demonstrating to trainees that you need to consider the **clinical question** to work out what you need to collect to answer the question Need to recognise there are **different types** of samples (different sites, longitudinal, sterile vs non sterile) Need to recognise the value of **metadata** for interpretation of the WGS to address the clinical question	1. Case studies of different clinical scenarios needing to be addressed with WGS and work through the samples types you would ideally like to scientifically address the question, then compare to what can practically be obtained and assess if suitable. (demonstrated in the course) 2. Metadata exercise - interpreting different outbreak responses depending on which metadata you use (demonstrated in the course) 3. Compare metagenomic study of sterile site vs non sterile site
Decide how to/issues with storage	Content involves identifying the different steps at which you need to **process** the samples: archive/store, dependent on global setting, clinical question, stability of samples, hospital vs community	1. Follow through a sample journey from taking sample through to WGS and how this differs in different case studies
Determine the most appropriate pre-WGS processing	Content involves an overview of the sample **preparation** needed prior to WGS: culture, subculture, DNA extraction, quality check, quantification of DNA, normalisation, library preparation **WGS platforms**: Long read vs short read WGS	1. Follow through a sample journey from taking sample through to WGS and how this differs in different case studies
Identifying suitable metadata	Content requires an understanding that genomes do not exist in isolation, the **provenance** is key and needs to be included in sample collection and shared wherever possible including public repositories	1. Metadata exercise - interpreting different outbreak responses depending on which metadata you use (demonstrated in the course) 2. Provide a phylogenetic tree/fasta files of similar isolates, with no metadata and determine what you can determine from this, and then what more information you would like to be able to draw any conclusions about these data
Establish ethical issues/consents needed	Review the clinical question and whether additional **consents/ethical approval** are required for the study	1. Lecture by ethical experts in this area (demonstrated in the course) 2. Discussion forum to share challenges faced in this area

Part 3. Activities

1. TB case study (20 min)

1. Metadata activity (20 min)

wellcome connecting science

T3connect

Centre for Genomic Pathogen Surveillance

Part 3. Further challenges in sampling for genomics – TB case study in Birmingham

TB case study (20 min) - Students to read the paper before the session.

a. A retrospective investigation of TB outbreak in Birmingham UK.
b. Analysed samples collected over 10 years.
c. Community transmission of TB established after analysis.

Consider:

a. *What gaps did the study have in the context of sample and data collection?*
b. *What would you have done differently?*
c. *How do you think you would teach the lessons learned from this case study?*

Part 3. Further challenges in sampling for genomics – TB case study in Birmingham

- **Study**
  - Non-real time transmission study, 2009-19
  - Purpose of sampling seem not to have been defined before sampling

- **Source sample**
  - Culture isolates – 64 missing in the retrospective archive
  - Postcode information missing for some patients
  - Person-to-person contact data missing
  - Only one isolate sequencies – misses strain diversity
  - TB culture – long time to positivity, limits rapid genomic-based response
  - *Direct sequencing from the clinical sample may shorten time to result and increase strain diversity*
  - *Metagenomics – waster for example may simplify surveillance*

- **Nucleic acid quality**
  - Freezing compromises bacterial recovery & DNA quality?

wellcome connecting science

T3connect

Centre for Genomic Pathogen Surveillance

Part 3. Metadata - why do we need it?

Metadata - why do we need it?
(ppt slides handouts)

Divide into 5 groups

You are each given different metadata fields:

- what could this tell you about the species of the genomes?
- what are the implications for this patient/s?
- What are the other public health implications?

Part 4.
Reflection on strategies used for this sessions

wellcome connecting science

T3connect

Centre for Genomic Pathogen Surveillance

Next Steps after 3A Specimen and data collection & preservation

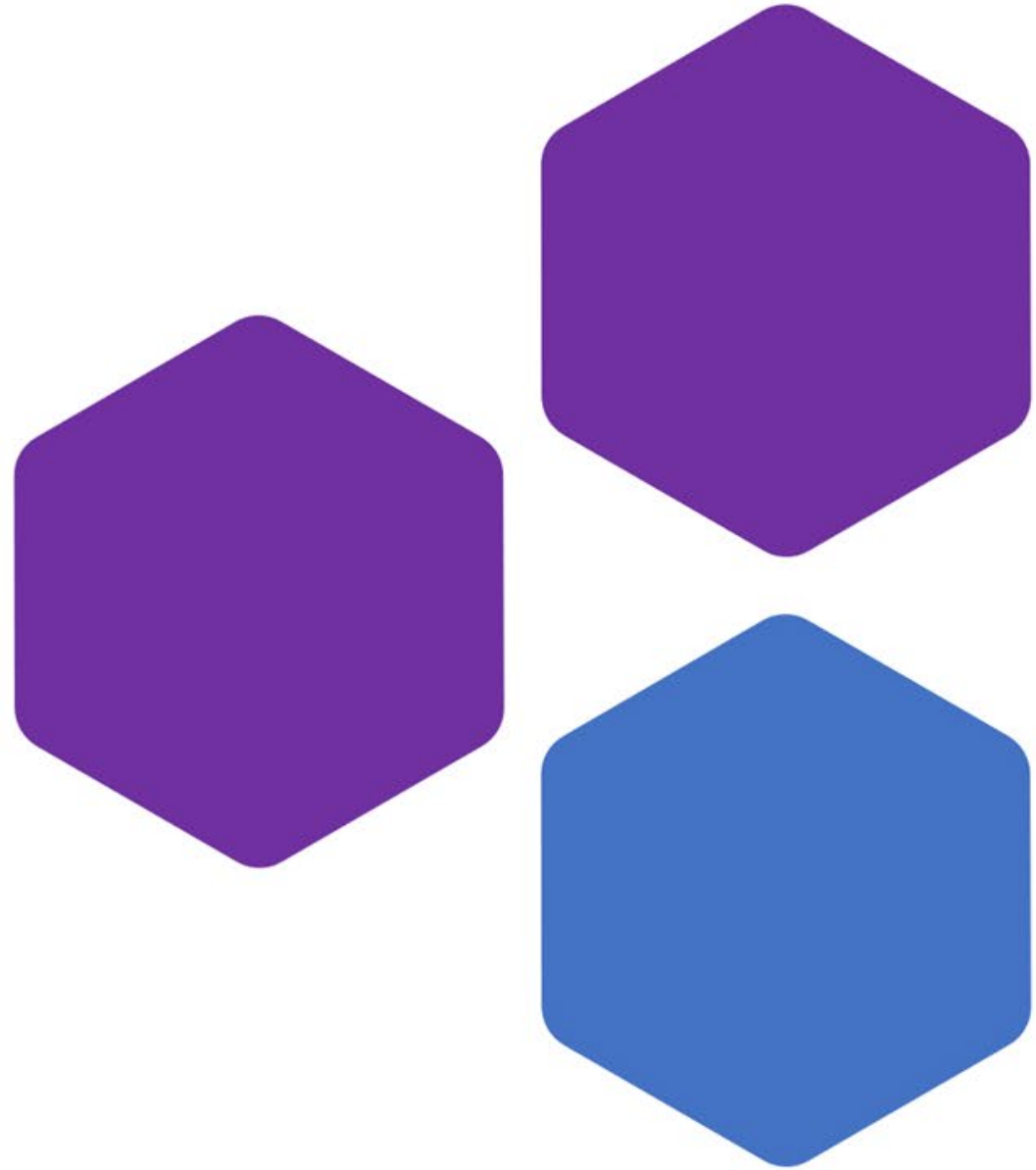
- Sample Genomic Analysis, Data Generation, and QC (Module 3B)
  - Nucleic acid extraction
  - Library prep & sequencing
  - Genome sequence analysis
- Data analysis & integration (Module 3C)
  - Labelling & barcoding is crucial
  - Data linkage
- Data interpretation & application (Module 3D)
  - Feedback to clinic or policy makers
  - Implications & actions

wellcome connecting science

T3connect

Centre for Genomic Pathogen Surveillance

Thank you



T3connect