GenEpiO and British Columbia Centre for Disease Control Tuberculosis data projects

- BC-Oxford TB Informatics Collaboration
  - March 2016: Variant calling Hackathon hosted by Public Health England to Compare their Tuberculosis pipeline with competitors.
  - Fall 2016: Jennifer Gardy leading the prototype COMPASS-TB report which optimizes the presentation of TB clinical and genomics information
  - October 2016: CRyPTIC/CPTR Data and Reporting Standardisation Meeting
- GenEpiO terms associated with the isolates and metadata associated with the reporting of lab test results

Damion Dooley
Scientific Programmer
Hsiao Lab, BCCDC Public Health Laboratory and
Department of Pathology and Laboratory Medicine
University of British Columbia

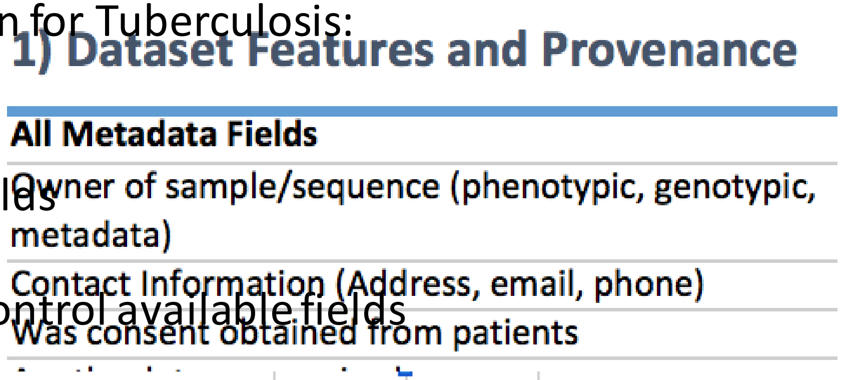
Dr. Emma Griffiths
Postdoctoral Researcher
Fiona Brinkman Lab
Simon Fraser University

DRAFT Comparison of database fields

CRyPTIC: Comprehensive Resistance Prediction for Tuberculosis: an International Consortium

Oxford-TB: Public Health England database fields

BCCDC: British Columbia Centre for Disease Control available fields



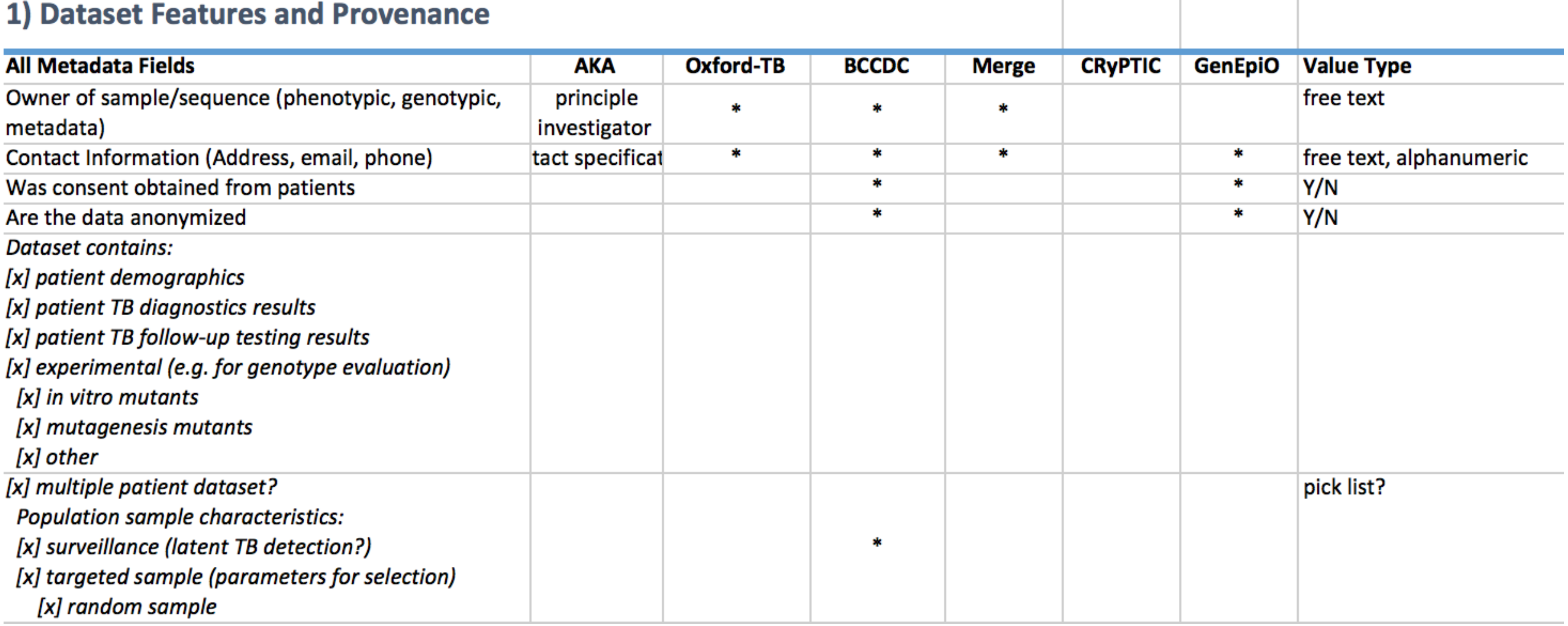
1) Dataset Features and Provenance

All Metadata Fields

Owner of sample/sequence (phenotypic, genotypic, metadata)

Contact Information (Address, email, phone)

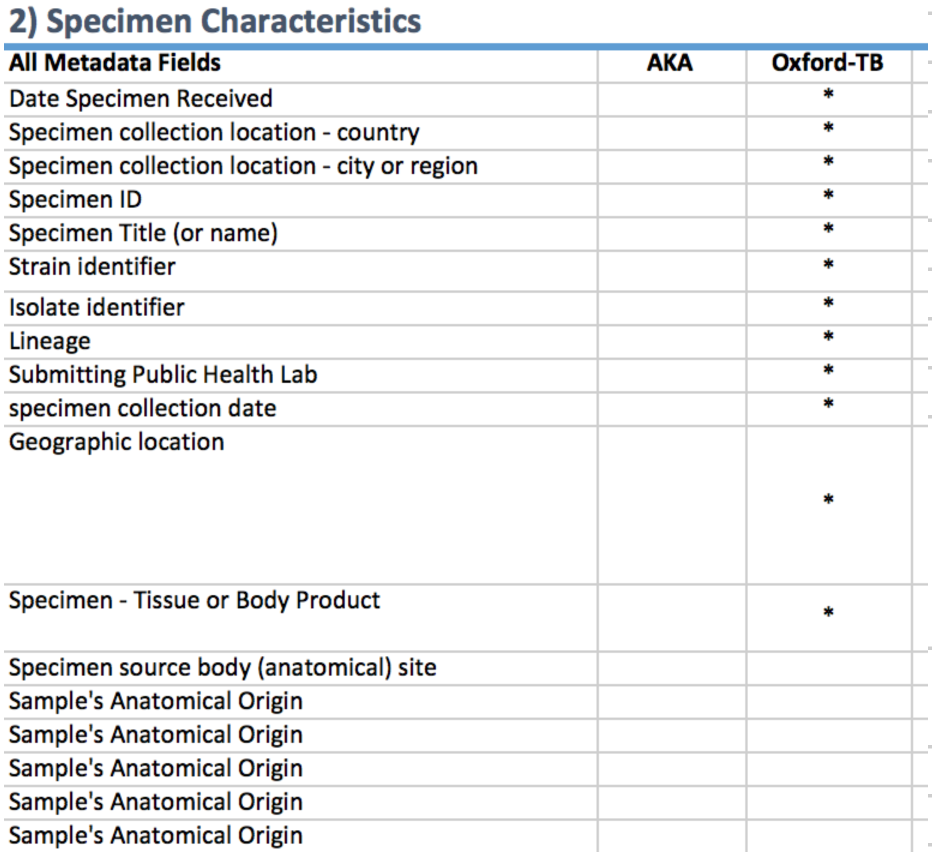
Was consent obtained from patients



1) Dataset Features and Provenance

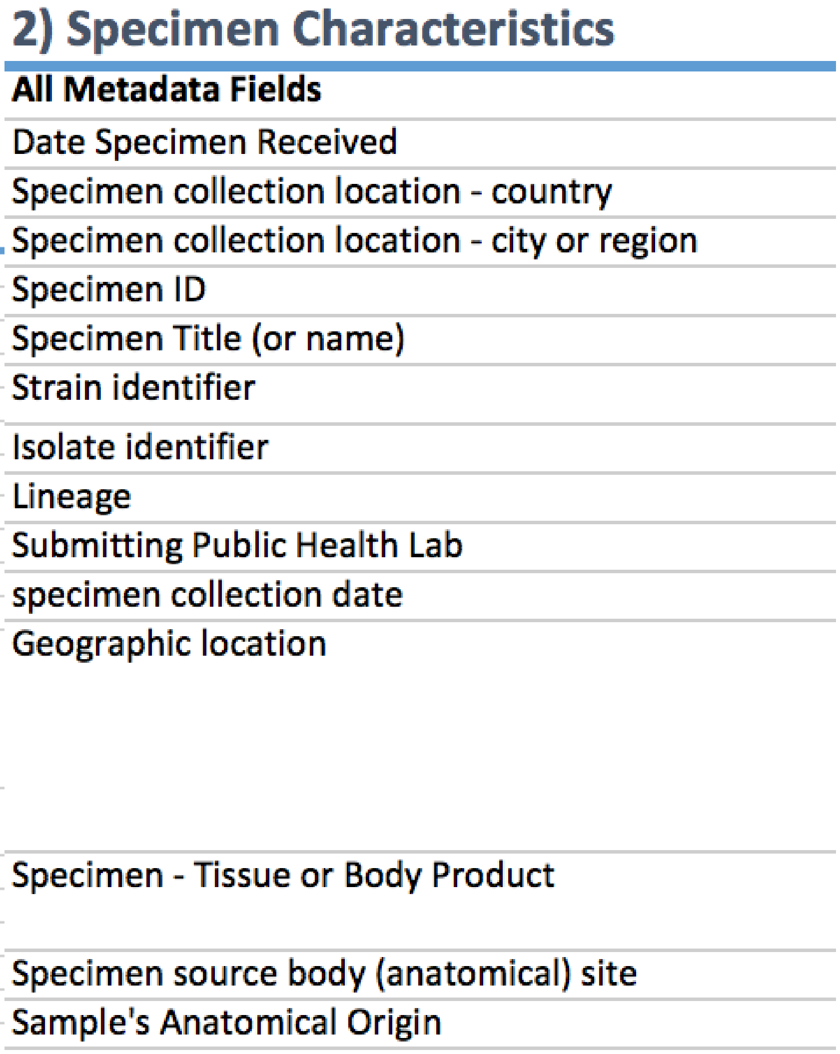
All Metadata Fields	AKA	Oxford-TB	BCCDC	Merge	CRyPTIC	GenEpiO	Value Type
Owner of sample/sequence (phenotypic, genotypic, metadata)	principle investigator	*	*	*			free text
Contact Information (Address, email, phone)	tact specificat	*	*	*		*	free text, alphanumeric
Was consent obtained from patients			*			*	Y/N
Are the data anonymized			*			*	Y/N
Dataset contains: *[x] patient demographics* *[x] patient TB diagnostics results* *[x] patient TB follow-up testing results* *[x] experimental (e.g. for genotype evaluation)* *[x] in vitro mutants* *[x] mutagenesis mutants* *[x] other*							
[x] multiple patient dataset? *Population sample characteristics:* *[x] surveillance (latent TB detection?)* *[x] targeted sample (parameters for selection)* *[x] random sample*			*				pick list?

DRAFT Comparison of database fields



2) Specimen Characteristics

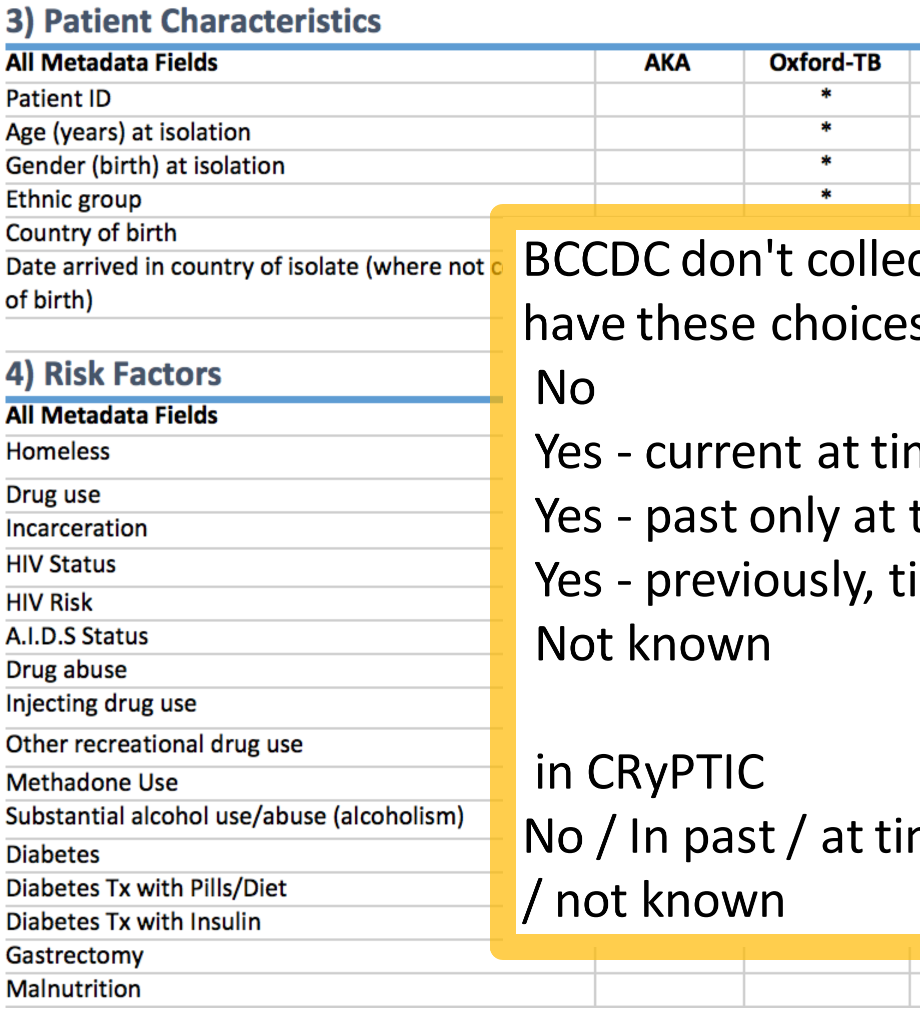
All Metadata Fields	AKA	Oxford-TB
Date Specimen Received		*
Specimen collection location - country		*
Specimen collection location - city or region		*
Specimen ID		*
Specimen Title (or name)		*
Strain identifier		*
Isolate identifier		*
Lineage		*
Submitting Public Health Lab		*
specimen collection date		*
Geographic location		*
Specimen - Tissue or Body Product		*
Specimen source body (anatomical) site		
Sample's Anatomical Origin		
Sample's Anatomical Origin		
Sample's Anatomical Origin		
Sample's Anatomical Origin		
Sample's Anatomical Origin		



2) Specimen Characteristics

All Metadata Fields
Date Specimen Received
Specimen collection location - country
Specimen collection location - city or region
Specimen ID
Specimen Title (or name)
Strain identifier
Isolate identifier
Lineage
Submitting Public Health Lab
specimen collection date
Geographic location
Specimen - Tissue or Body Product
Specimen source body (anatomical) site
Sample's Anatomical Origin

DRAFT Comparison of database fields

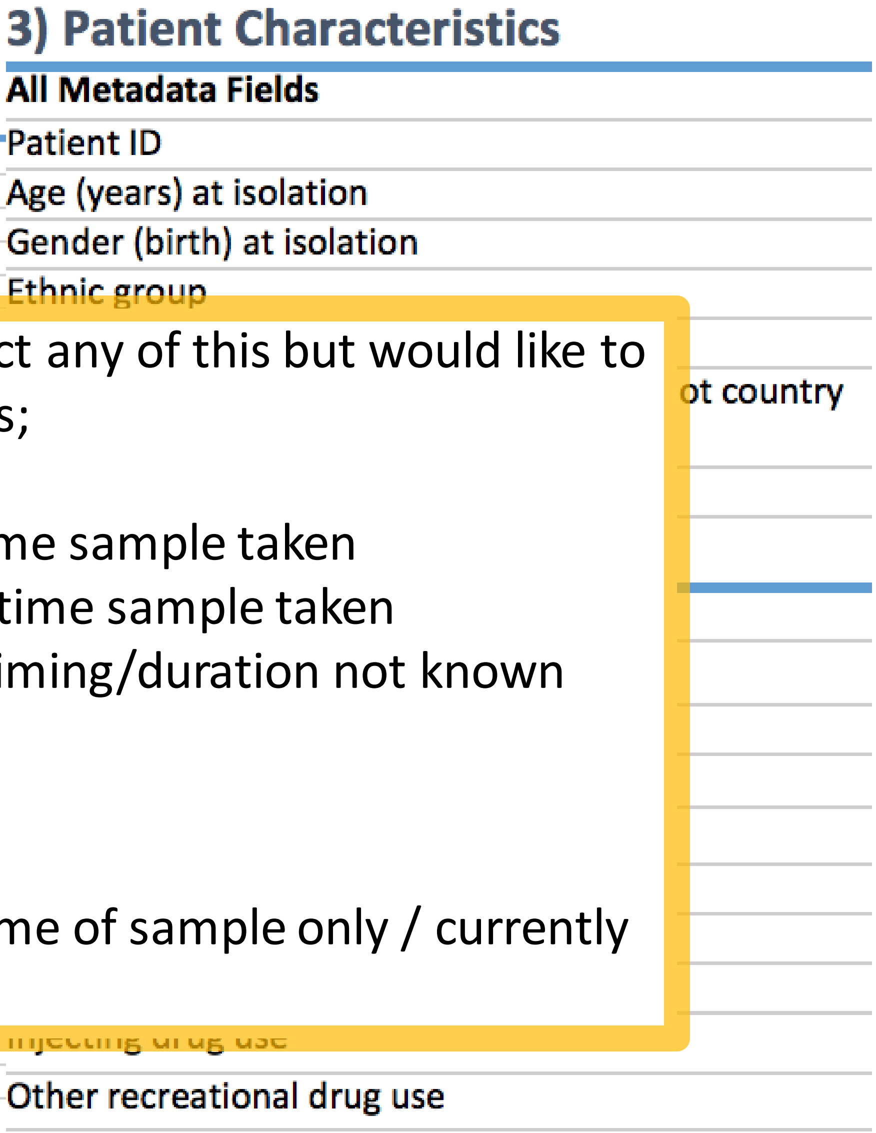


3) Patient Characteristics

All Metadata Fields	AKA	Oxford-TB
Patient ID		*
Age (years) at isolation		*
Gender (birth) at isolation		*
Ethnic group		*
Country of birth		
Date arrived in country of isolate (where not c of birth)		

4) Risk Factors

All Metadata Fields
Homeless
Drug use
Incarceration
HIV Status
HIV Risk
A.I.D.S Status
Drug abuse
Injecting drug use
Other recreational drug use
Methadone Use
Substantial alcohol use/abuse (alcoholism)
Diabetes
Diabetes Tx with Pills/Diet
Diabetes Tx with Insulin
Gastrectomy
Malnutrition



3) Patient Characteristics

All Metadata Fields
Patient ID
Age (years) at isolation
Gender (birth) at isolation
Ethnic group
ot country
Injecting drug use
Other recreational drug use

BCCDC don't collect any of this but would like to have these choices;

No

Yes - current at time sample taken

Yes - past only at time sample taken

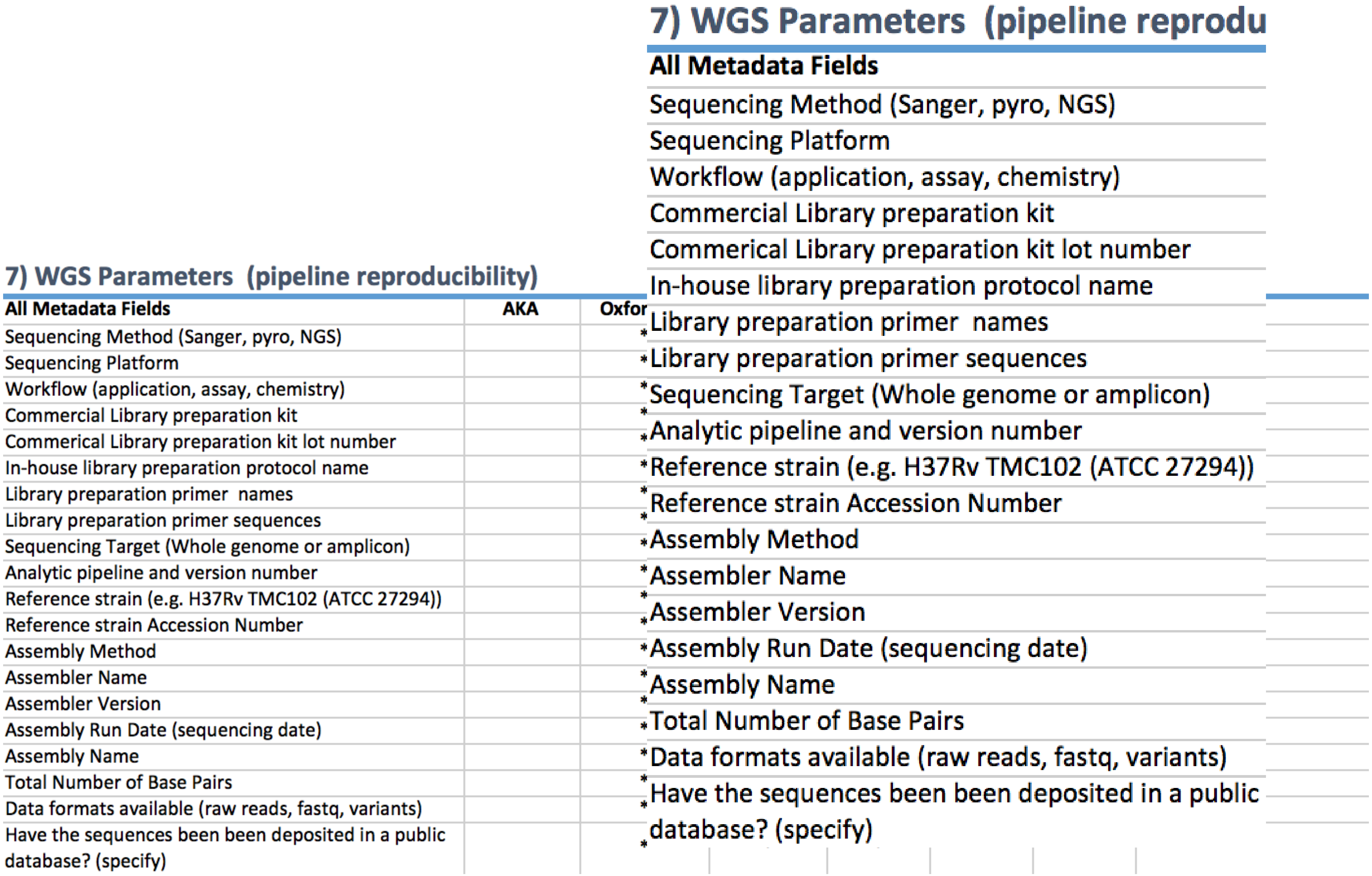
Yes - previously, timing/duration not known

Not known

in CRyPTIC

No / In past / at time of sample only / currently / not known

DRAFT Comparison of database fields



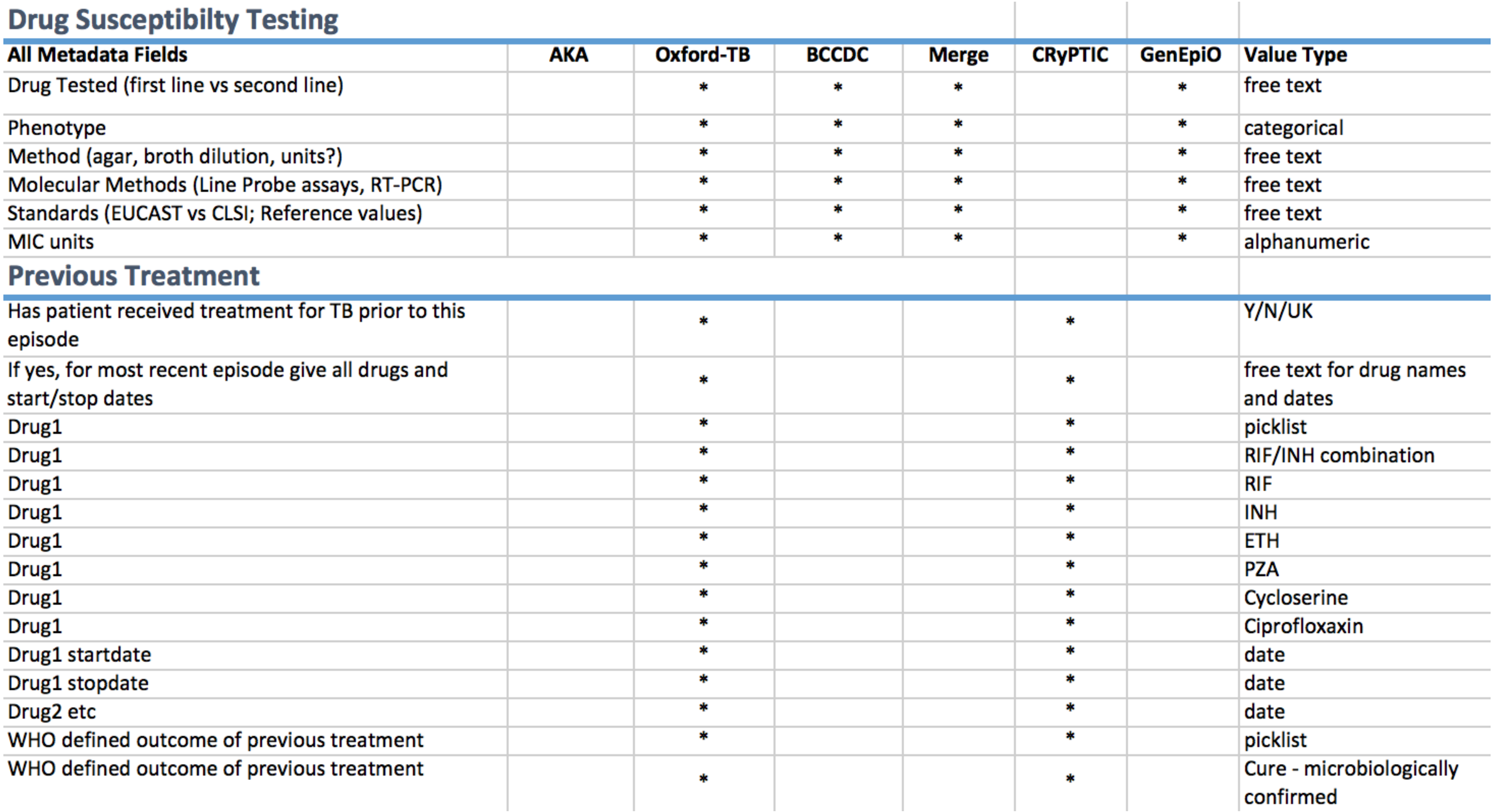
7) WGS Parameters (pipeline reproducibility)

All Metadata Fields	AKA	Oxfor
Sequencing Method (Sanger, pyro, NGS)		*
Sequencing Platform		*
Workflow (application, assay, chemistry)		*
Commercial Library preparation kit		*
Commerical Library preparation kit lot number		*
In-house library preparation protocol name		*
Library preparation primer names		*
Library preparation primer sequences		*
Sequencing Target (Whole genome or amplicon)		*
Analytic pipeline and version number		*
Reference strain (e.g. H37Rv TMC102 (ATCC 27294))		*
Reference strain Accession Number		*
Assembly Method		*
Assembler Name		*
Assembler Version		*
Assembly Run Date (sequencing date)		*
Assembly Name		*
Total Number of Base Pairs		*
Data formats available (raw reads, fastq, variants)		*
Have the sequences been been deposited in a public database? (specify)		*

7) WGS Parameters (pipeline reprodu

All Metadata Fields
Sequencing Method (Sanger, pyro, NGS)
Sequencing Platform
Workflow (application, assay, chemistry)
Commercial Library preparation kit
Commerical Library preparation kit lot number
In-house library preparation protocol name
Library preparation primer names
Library preparation primer sequences
Sequencing Target (Whole genome or amplicon)
Analytic pipeline and version number
Reference strain (e.g. H37Rv TMC102 (ATCC 27294))
Reference strain Accession Number
Assembly Method
Assembler Name
Assembler Version
Assembly Run Date (sequencing date)
Assembly Name
Total Number of Base Pairs
Data formats available (raw reads, fastq, variants)
Have the sequences been been deposited in a public database? (specify)

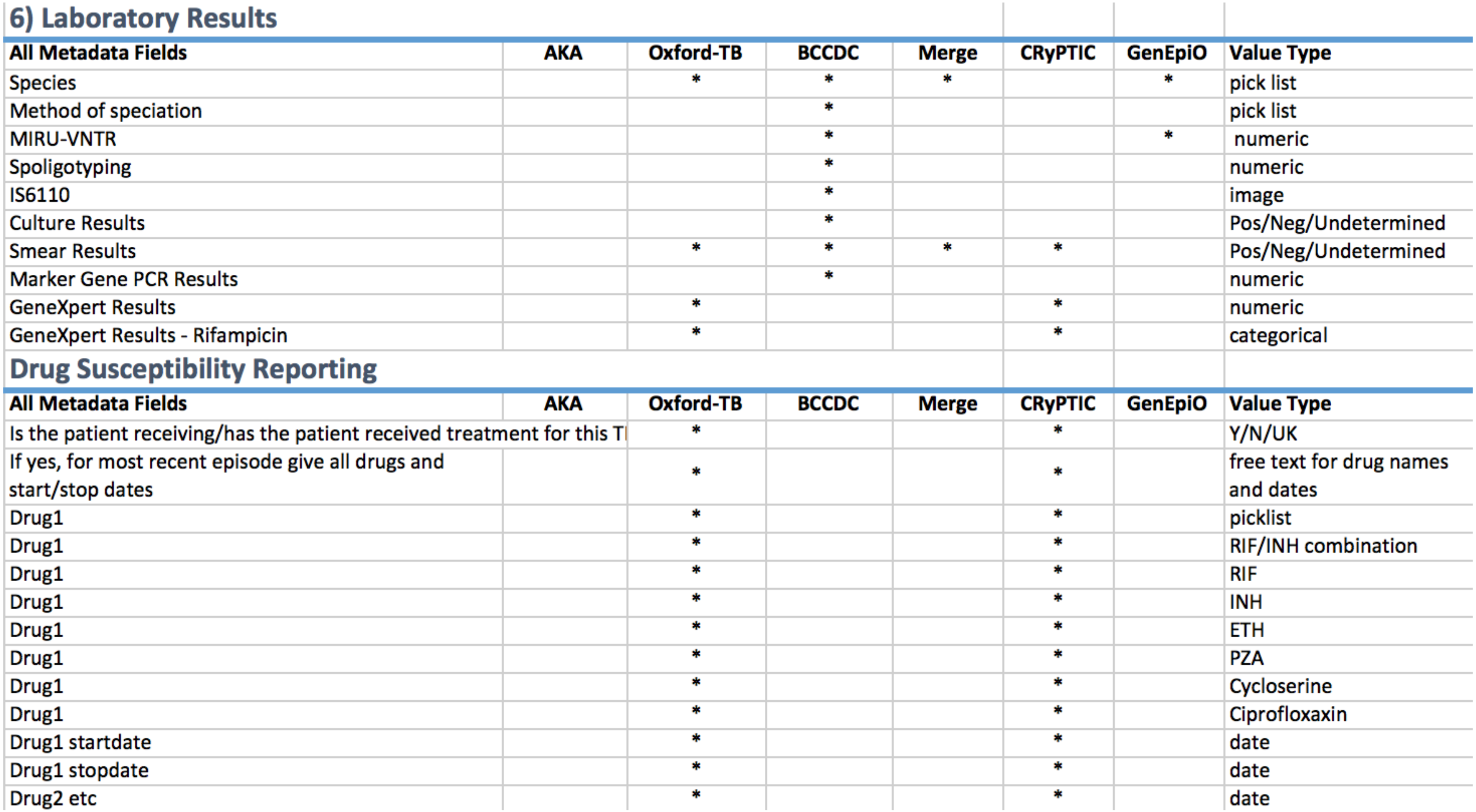
DRAFT Comparison of database fields



Drug Susceptibilty Testing

All Metadata Fields	AKA	Oxford-TB	BCCDC	Merge	CRyPTIC	GenEpiO	Value Type
Drug Tested (first line vs second line)		*	*	*		*	free text
Phenotype		*	*	*		*	categorical
Method (agar, broth dilution, units?)		*	*	*		*	free text
Molecular Methods (Line Probe assays, RT-PCR)		*	*	*		*	free text
Standards (EUCAST vs CLSI; Reference values)		*	*	*		*	free text
MIC units		*	*	*		*	alphanumeric
Previous Treatment							
Has patient received treatment for TB prior to this episode		*			*		Y/N/UK
If yes, for most recent episode give all drugs and start/stop dates		*			*		free text for drug names and dates
Drug1		*			*		picklist
Drug1		*			*		RIF/INH combination
Drug1		*			*		RIF
Drug1		*			*		INH
Drug1		*			*		ETH
Drug1		*			*		PZA
Drug1		*			*		Cycloserine
Drug1		*			*		Ciprofloxaxin
Drug1 startdate		*			*		date
Drug1 stopdate		*			*		date
Drug2 etc		*			*		date
WHO defined outcome of previous treatment		*			*		picklist
WHO defined outcome of previous treatment		*			*		Cure - microbiologically confirmed

DRAFT Comparison of database fields



6) Laboratory Results

All Metadata Fields	AKA	Oxford-TB	BCCDC	Merge	CRyPTIC	GenEpiO	Value Type
Species		*	*	*		*	pick list
Method of speciation			*				pick list
MIRU-VNTR			*			*	numeric
Spoligotyping			*				numeric
IS6110			*				image
Culture Results			*				Pos/Neg/Undetermined
Smear Results		*	*	*	*		Pos/Neg/Undetermined
Marker Gene PCR Results			*				numeric
GeneXpert Results		*			*		numeric
GeneXpert Results - Rifampicin		*			*		categorical

Drug Susceptibility Reporting

All Metadata Fields	AKA	Oxford-TB	BCCDC	Merge	CRyPTIC	GenEpiO	Value Type
Is the patient receiving/has the patient received treatment for this T		*			*		Y/N/UK
If yes, for most recent episode give all drugs and start/stop dates		*			*		free text for drug names and dates
Drug1		*			*		picklist
Drug1		*			*		RIF/INH combination
Drug1		*			*		RIF
Drug1		*			*		INH
Drug1		*			*		ETH
Drug1		*			*		PZA
Drug1		*			*		Cycloserine
Drug1		*			*		Ciprofloxaxin
Drug1 startdate		*			*		date
Drug1 stopdate		*			*		date
Drug2 etc		*			*		date