# CBW2023 Contextual Data Curation Scenario Answer Keys

The following mock scenarios describe different types of genomics contextual data for SARS-CoV-2 surveillance. The scenarios in this exercise are fictional and do not represent real data or programs at any public health organization in Canada, and are intended for training purposes only.

## Scenario 1: Clinical SARS-CoV-2 Data

The **BCCDC Public Health Laboratory** obtained a **nasopharyngeal swab** for **diagnostic testing (sample ID Bc-12345-ab)** on **March 1 2023** from a **symptomatic, 44 year old female** that had been **hospitalized** in the **ICU**. The individual had been exhibiting a **cough, fever, muscle weakness**, as well as other symptoms of **Acute Respiratory Distress Syndrome**.

The individual recently **travelled** to the **United States** on holiday and returned on **Feb 19 2023**. The sample was flagged for sequencing as part of the lab’s **International travel surveillance** program. The sample was sequenced on **March 7 2023** using an **Illumina MiSeq** instrument. The raw data was processed using **ncov-tools 2.3.1** as part of their **bioinformatics protocol** (<https://github.com/jts/ncov2019-artic-nf/blob/master/README.md>) and dehosted using **BWA (version 0.7.17)**. The consensus sequence was generated using **iVar 2.3.1**. The sequence was uploaded to GISAID and assigned the accession number **EPI\_ISL\_436489**. Drs **Tejinder Singh, Fei Hu and Joe Blogs** helped to generate the sequence.

***Note: GISAID “isolate” identifiers are generated by data providers in the following format***

hCoV-19/COUNTRY/ISO regional code-Identifier/year

e.g. hCoV-19/CANADA/BC-provlab1234/2020

“hCoV-19” is always the same

The country is capitalized e.g. CANADA

ISO regional codes are 2 letter codes for the province or state e.g. BC, ON, QC, SK, AB

The identifier corresponds to the sample ID.

The “year” corresponds to the year of sample collection.

Template: CanCOGeN

**specimen collector sample ID:** Bc-12345-ab

GISAID accession: EPI\_ISL\_436489

**sample collected by:** BCCDC Public Health Laboratory

**sequence submitted by**: BCCDC Public Health Laboratory

**sample collection date:** 2023-03-01

**sample collection date precision:** day

**geo\_loc name (country):** Canada

**geo\_loc name (state/province/territory):** British Columbia

**organism:** Severe acute respiratory syndrome coronavirus 2

**isolate:** hCoV-19/CANADA/BC-Bc-12345-ab/2023

**purpose of sampling:** Diagnostic testing

**purpose of sampling details:** Not Provided

**NML submitted specimen type:** Not Applicable

**anatomical material:** Not Applicable

**anatomical part:** Nasopharynx (NP)

**body product:** Not Applicable

**environmental material:** Not Applicable

**environmental site:** Not Applicable

**collection device:** Swab

**collection method:** Not Applicable

**host (scientific name):** Homo sapiens

**host health state:** Symptomatic

**host health status details:** Hospitalized (ICU)

**host disease:** COVID-19

**host age:** 44

**host age unit:** year

**host age bin:** 40 - 49

**host gender:** Female

**signs and symptoms:** Cough; Fever; Muscle weakness; Acute Respiratory Distress Syndrome

**destination of most recent travel (country):** United States of America

**most recent travel return date:** 2023-02-19

**purpose of sequencing:** International travel surveillance

**purpose of sequencing details:** Not Provided

**sequencing date:** 2023-03-07

**sequencing instrument:** Illumina MiSeq

**raw sequence data processing method:** ncov-tools 2.3.1

**dehosting method:** BWA 0.7.17

**consensus sequence software name:** iVar

**consensus sequence software version:** 2.3.1

**bioinformatics protocol:** https://github.com/jts/ncov2019-artic-nf/blob/master/README.md

**authors:** Tejinder Singh; Fei Hu; Joe Blogs

## Scenario 2: Wastewater SARS-CoV-2 Data

**Untreated, fast moving, wastewater** is continuously collected in a **municipal sewer system** starting on **Nov 1 2023 for 72hrs**. The sewer system, which collects **rainwater as well as household and institutional waste**, is part of a **routine surveillance** program for tracking **community-level SARS-CoV-2** variants (**sewer site ID WWSC2-ABC-b**) in order to establish **baseline norms**. The sewer is located near a hospital and the **hospital’s effluent is piped into the sewer system**. **Five Moore swabs** from the site of collection are **pooled** (**sample ID BW-WW-12345**). It **rained the day before sample collection (5cm of rain)**. The **wastewater catchment area serves approx 800 000 people** in a **suburban** area (**Mississauga, Ontario, Canada**). The **ambient air temperature** at the time of collection was **15 degrees Celsius**. The water was **8 degrees Celsius at the time of collection**, and **3 degrees Celsius when it was received by the sequencing lab**. The **instantaneous flow rate** is **3 cubic meter per second (m^3/s), with 8% total suspended solids**. The sample was **collected by** the **Region of Peel regional authority**, and **sequenced by** the **Public Health Ontario provincial health laboratory** (**contact: Johnny Bloggs; jbloggs@provlab.ca**). A **watershed shapefile** delineating the geographical coordinates covered by the sewer system is **available**. The **presence** of SARS-CoV-2 was first detected using **qPCR (N1 gene, Ct value of 22)**. The **amplicon-based** sample was sequenced on an **Illumina MiSeq** on **Jan 18 2024** using the **ARTIC V5 400bp primer scheme (artic-v5.3.2\_400)**, and **consensus sequences** were generated using **ViralRecon software v1.23** and **lineage assignments** were performed using **pUShER (v1.2.6)**. The rich contextual data record for the sequence is provided below. This record is for the public health laboratory’s use only, and many details were removed when sharing data according to organization-specific data sharing policies. The associated contextual data record is provided below. This record highlights how rich contextual data can be captured using the specification - including catchment details such as geographical coordinates and population ranges, activity upstream of sampling that may affect results, how to record longitudinal sampling events, capture of environmental conditions and measurements, associated laboratory testing results (Ct values), and lineage designations.

Template: WWSARS-CoV-2

**specimen collector sample ID:** BW-WW-12345

**sampling site ID:** WWSC2-ABC-b

**sample collected by:** Region of Peel Regional Authority

**geo\_loc\_name (country):** Canada

**geo\_loc\_name (state/province/territory):** Ontario

**geo\_loc\_name (city):** Mississauga

**watershed shapefile availability:** Watershed shapefile available

**organism:** Severe acute respiratory syndrome coronavirus 2

**purpose of sampling:** Wastewater pathogen surveillance

**presamping activity:** Healthcare activity

**presampling activity details:** hospital effluent piped into sewer system

**sample collection date:** 2023-11-01

**sample collection end date:** 2023-11-04

**sample collection time duration value:** 72

**sample collection time duration unit:** Hour

**scale of sampling:** Community-level surveillance

**specimen processing:** Pooling specimens

**specimen processing details:** 5 Moore swabs pooled from same sewer

**environmental site:** Sewer

**environmental material:** Wastewater

**environmental material properties:** Untreated; Fluid (fast)

**wastewater system type:** Combined sewer system

**collection device:** Moore swab

**collection method:** Passive composite sampling

**populated area type:** Suburban

**water catchment area human population measurement value**: 800000

**water catchment area human population bin:** 100,000 - 1,000,000 people

**presampling weather conditions:** Rain

**precipitation measurement value:** 5

**precipitation measurement unit:** centimeter (cm)

**ambient temperature measurement value:** 15

**ambient temperature measurement unit:** degree Celsius (C)

**instantaneous flow rate measurement value:** 3

**instantaneous flow rate measurement unit:** cubic meter per second (m^3/s)

**total suspended solids (TSS) measurement value:** 8

**total suspended solids (TSS) measurement unit:** Percent (%)

**sample temperature value (at collection):** 8

**sample temperature unit (at collection):** degree Celsius (C)

**sample temperature value (when received):** 3

**sample temperature unit (when received):** degree Celsius (C)

**purpose of sequencing:** Baseline surveillance (random sampling)

**sequencing assay type:** Amplicon sequencing assay

**sequencing date:** 2024-01-18

**sequenced\_by:** Public Health Ontario

**sequenced\_by\_contact\_name:** Johnny Bloggs

**sequenced\_by\_contact\_email:** jbloggs@provlab.ca

**sequencing instrument:** Illumina MiSeq

**amplicon pcr primer scheme:** artic-v5.3.2\_400

**amplicon size:** 400

**consensus sequence software name:** ViralRecon

**consensus sequence software version:** 1.23

**diagnostic measurement method:** Quantitative real time polymerase chain reaction (qPCR)

**gene name:** N gene (orf9)

**diagnostic target presence:** Diagnostic target present

**diagnostic measurement value:** 22

**diagnostic measurement unit:** cycle threshold (Ct)

**lineage/clade name:** JN.1

**lineage/clade analysis software name:** pUShER

**lineage/clade analysis software version:** 1.2.6