DEC 28, 2022

OPEN ACCESS

Protocol Citation: Paul Lorenzo A Gaite, Dr Ritchie Mae T Gamot, Prof Lyre Anni E Murao 2022. Establishing processes to capture standardized contextual data. **protocols.io** https://protocols.io/view/establishing-processes-to-capture-standardized-con-cghztt76

License: This is an open access protocol distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Protocol status: Working
We use this protocol and it's working

Created: Sep 13, 2022

Last Modified: Dec 28, 2022

PROTOCOL integer ID: 69913

Establishing processes to capture standardized contextual data

Paul Lorenzo A Gaite[1], Dr Ritchie Mae T Gamot[1], Prof Lyre Anni E Murao[1,2]

[1]PGC Mindanao; [2]UP Mindanao

phagesubgrantph

ABSTRACT

One of the obstacles of biosurveillance is the non-standard recording and storage of contextual data. This is a problem especially when establishing a national biosurveillance program made up of different public health laboratories having differing data collection standards (e.g. each laboratory requiring or recording a different set of contextual data categories). As a result, timely submission to databases is hampered due to additional time to request additional pieces of metadata from sampling laboratories. To circumvent this problem, a need for standard contextual data collection and storage system should be in place. For this reason PHA4GE has developed a contextual data standard package.

1 Abstract/Introduction

One of the obstacles of biosurveillance is the non-standard recording and storage of contextual data. This is a problem especially when establishing a national biosurveillance program made up of different public health laboratories having differing data collection standards (e.g. each laboratory requiring or recording a different set of contextual data categories). As a result, timely submission to databases is hampered due to additional time to request additional pieces of metadata from sampling laboratories. To circumvent this problem, a need for standard contextual

data collection and storage system should be in place. For this reason PHA4GE has developed a contextual data standard package.

The sections below show the process of establishing a contextual data standard for PGC Mindanao. The original PGC Mindanao workflow for the handling of contextual data is outlined (Section 2), particularly the collection of contextual data (Section 2.1), storage of contextual data (Section 2.2), stewardship of contextual data (Section 2.3), and preparation of contextual data for submission to GISAID (Section 2.4). The PHA4GE workflow for handling contextual data, as applied by PGC Mindanao is also outlined (Section 3).

2 PGC Mindanao workflow

PGC Mindanao did not have a defined workflow for handling contextual data. The subsections below outline the activities done in connection to the collection, storage, and stewardship of contextual data, as well as preparation of contextual data for submission to GISAID.

2.1 Collection of Contextual Data:

The contextual data associated with the patient samples (e.g. anonymised patient data) were collected by the Sub-National Laboratories (SNLs). The collected contextual data were then collated in template spreadsheets (these contain both required and optional metadata fields for a GISAID submission) by the SNLs and sent back through e-mail to PGC Mindanao (Figures 1 and 2).

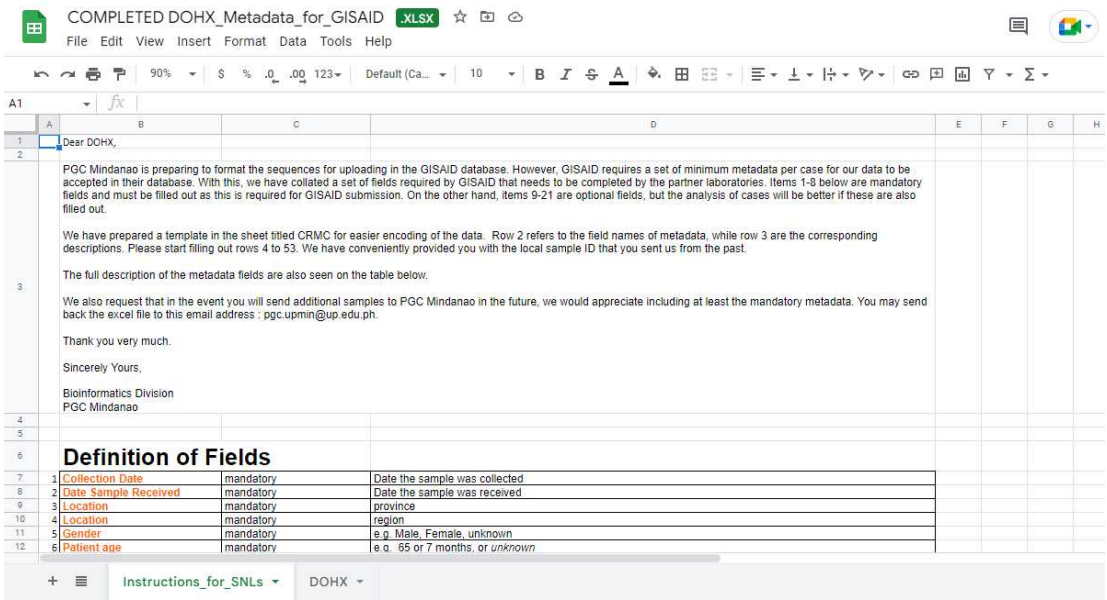
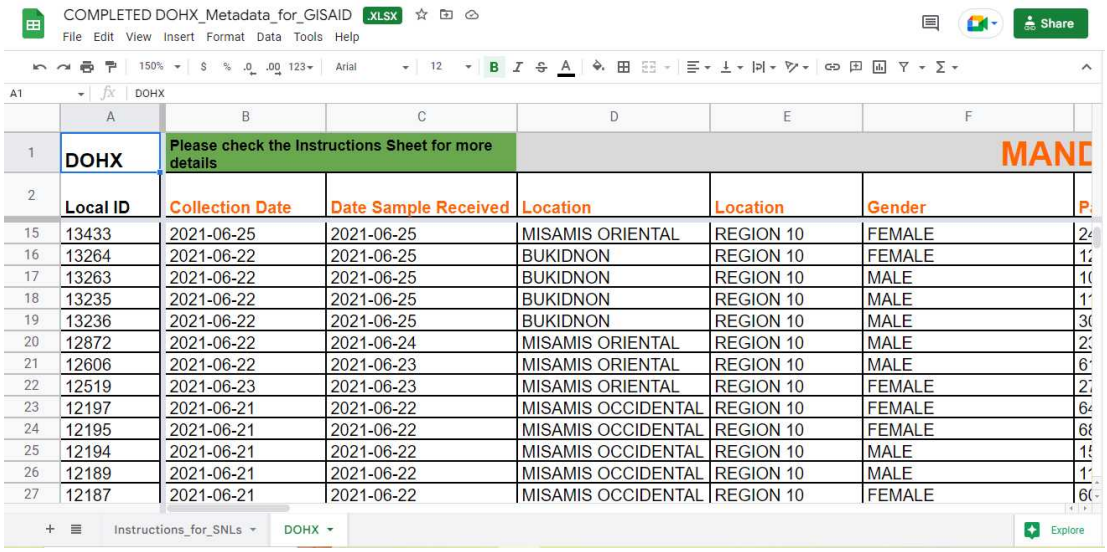


Figure 1. Instructions to SNL for inputting contextual data from sequenced samples



COMPLETED DOHX_Metadata_for_GISAID .XLSX

File Edit View Insert Format Data Tools Help

Share

	A	B	C	D	E	F
1	DOHX	Please check the Instructions Sheet for more details				MAND
2	Local ID	Collection Date	Date Sample Received	Location	Location	Gender
15	13433	2021-06-25	2021-06-25	MISAMIS ORIENTAL	REGION 10	FEMALE
16	13264	2021-06-22	2021-06-25	BUKIDNON	REGION 10	FEMALE
17	13263	2021-06-22	2021-06-25	BUKIDNON	REGION 10	MALE
18	13235	2021-06-22	2021-06-25	BUKIDNON	REGION 10	MALE
19	13236	2021-06-22	2021-06-25	BUKIDNON	REGION 10	MALE
20	12872	2021-06-22	2021-06-24	MISAMIS ORIENTAL	REGION 10	MALE
21	12606	2021-06-22	2021-06-23	MISAMIS ORIENTAL	REGION 10	MALE
22	12519	2021-06-23	2021-06-23	MISAMIS ORIENTAL	REGION 10	FEMALE
23	12197	2021-06-21	2021-06-22	MISAMIS OCCIDENTAL	REGION 10	FEMALE
24	12195	2021-06-21	2021-06-22	MISAMIS OCCIDENTAL	REGION 10	FEMALE
25	12194	2021-06-21	2021-06-22	MISAMIS OCCIDENTAL	REGION 10	MALE
26	12189	2021-06-21	2021-06-22	MISAMIS OCCIDENTAL	REGION 10	MALE
27	12187	2021-06-21	2021-06-22	MISAMIS OCCIDENTAL	REGION 10	FEMALE

Instructions_for_SNLs DOHX Explore

Figure 2. Spreadsheet containing actual contextual data from sequenced samples as inputted by SNL

2.2 Storage of Contextual Data:

Contextual data are stored within workstations at PGC Mindanao. Part of the protocol from the previous project entailed upload of contextual data to the REDCap database.

2.3 Stewardship of Contextual Data:

As contextual data is sensitive information, at the beginning of the collaboration between the previous project and PGC Mindanao an agreement was created and a specific clause was created to outline how data will be shared. Specifically, the source SNL and the project should be informed and permissions should be acquired when sharing data to third-party projects including PHA4GE.

2.4 Preparation of Contextual Data for Submission to GISAID:

The previous project involved collection of all contextual data in a REDCap database. Various scripts were created and provided by the previous project to transfer data from the database to spreadsheets. The GISAID format was followed for uploading to GISAID database. Please refer to protocol "Submission of sequence and contextual data to GISAID, INSDC repositories, or other databases" for details on this process.

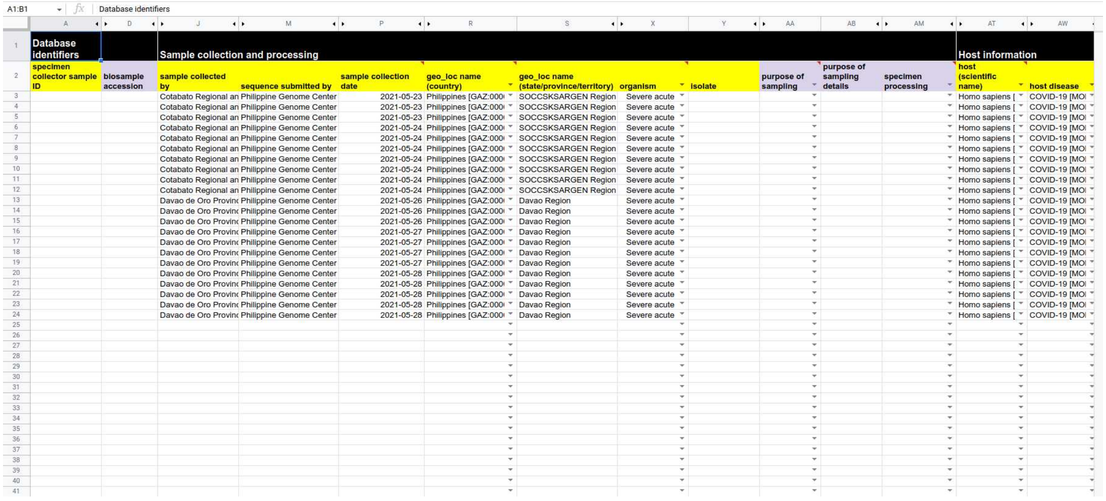
3 PHA4GE workflow

PHA4GE has developed and provided a contextual data specification package (e.g. submission template and scripts) that can facilitate collation of contextual data and eventual upload of consensus sequences and corresponding contextual data to databases.

This section describes the input of contextual data to the PHA4GE contextual data template spreadsheet (Section 3.1), and usage and some feedback on the PHA4GE contextual data template spreadsheet (Section 3.2).

3.1 Input of contextual data to the PHA4GE contextual data template spreadsheet:

Contextual data collected from the samples (please refer to Section 2.1 - "Collection of Contextual Data" of this protocol for details) were inputted into the PHA4GE contextual data template spreadsheet (Figure 3). Required fields were inputted.



A1:B1 | Database identifiers

	A	D	J	M	P	R	S	X	Y	AA	AB	AM	AT	AW
1	Database identifiers		Sample collection and processing										Host information	
2	specimen collector sample ID	biosample accession	sample collected by	sequence submitted by	sample collection date	geo_loc name (country)	geo_loc name (state/province/territory)	organism	isolate	purpose of sampling	purpose of sampling details	specimen processing	host (scientific name)	host disease
3			Cotabato Regional an	Philippine Genome Center	2021-05-23	Philippines [GAZ:000	SOCCSKSARGEN Region	Severe acute					Homo sapiens [	COVID-19 [MOI
4			Cotabato Regional an	Philippine Genome Center	2021-05-23	Philippines [GAZ:000	SOCCSKSARGEN Region	Severe acute					Homo sapiens [	COVID-19 [MOI
5			Cotabato Regional an	Philippine Genome Center	2021-05-23	Philippines [GAZ:000	SOCCSKSARGEN Region	Severe acute					Homo sapiens [	COVID-19 [MOI
6			Cotabato Regional an	Philippine Genome Center	2021-05-24	Philippines [GAZ:000	SOCCSKSARGEN Region	Severe acute					Homo sapiens [	COVID-19 [MOI
7			Cotabato Regional an	Philippine Genome Center	2021-05-24	Philippines [GAZ:000	SOCCSKSARGEN Region	Severe acute					Homo sapiens [	COVID-19 [MOI
8			Cotabato Regional an	Philippine Genome Center	2021-05-24	Philippines [GAZ:000	SOCCSKSARGEN Region	Severe acute					Homo sapiens [	COVID-19 [MOI
9			Cotabato Regional an	Philippine Genome Center	2021-05-24	Philippines [GAZ:000	SOCCSKSARGEN Region	Severe acute					Homo sapiens [	COVID-19 [MOI
10			Cotabato Regional an	Philippine Genome Center	2021-05-24	Philippines [GAZ:000	SOCCSKSARGEN Region	Severe acute					Homo sapiens [	COVID-19 [MOI
11			Cotabato Regional an	Philippine Genome Center	2021-05-24	Philippines [GAZ:000	SOCCSKSARGEN Region	Severe acute					Homo sapiens [	COVID-19 [MOI
12			Cotabato Regional an	Philippine Genome Center	2021-05-24	Philippines [GAZ:000	SOCCSKSARGEN Region	Severe acute					Homo sapiens [	COVID-19 [MOI
13			Davao de Oro Provinc	Philippine Genome Center	2021-05-26	Philippines [GAZ:000	Davao Region	Severe acute					Homo sapiens [	COVID-19 [MOI
14			Davao de Oro Provinc	Philippine Genome Center	2021-05-26	Philippines [GAZ:000	Davao Region	Severe acute					Homo sapiens [	COVID-19 [MOI
15			Davao de Oro Provinc	Philippine Genome Center	2021-05-26	Philippines [GAZ:000	Davao Region	Severe acute					Homo sapiens [	COVID-19 [MOI
16			Davao de Oro Provinc	Philippine Genome Center	2021-05-27	Philippines [GAZ:000	Davao Region	Severe acute					Homo sapiens [	COVID-19 [MOI
17			Davao de Oro Provinc	Philippine Genome Center	2021-05-27	Philippines [GAZ:000	Davao Region	Severe acute					Homo sapiens [	COVID-19 [MOI
18			Davao de Oro Provinc	Philippine Genome Center	2021-05-27	Philippines [GAZ:000	Davao Region	Severe acute					Homo sapiens [	COVID-19 [MOI
19			Davao de Oro Provinc	Philippine Genome Center	2021-05-27	Philippines [GAZ:000	Davao Region	Severe acute					Homo sapiens [	COVID-19 [MOI
20			Davao de Oro Provinc	Philippine Genome Center	2021-05-28	Philippines [GAZ:000	Davao Region	Severe acute					Homo sapiens [	COVID-19 [MOI
21			Davao de Oro Provinc	Philippine Genome Center	2021-05-28	Philippines [GAZ:000	Davao Region	Severe acute					Homo sapiens [	COVID-19 [MOI
22			Davao de Oro Provinc	Philippine Genome Center	2021-05-28	Philippines [GAZ:000	Davao Region	Severe acute					Homo sapiens [	COVID-19 [MOI
23			Davao de Oro Provinc	Philippine Genome Center	2021-05-28	Philippines [GAZ:000	Davao Region	Severe acute					Homo sapiens [	COVID-19 [MOI
24			Davao de Oro Provinc	Philippine Genome Center	2021-05-28	Philippines [GAZ:000	Davao Region	Severe acute					Homo sapiens [	COVID-19 [MOI

Figure 3. Contextual data from a sequencing batch inputted into PHA4GE contextual data template spreadsheet

3.2 Use of PHA4GE contextual data template spreadsheet

Usage of the contextual data template has proven to be useful for our purposes since this can standardize data requirements, submission, and storage within the center, as well as possible sharing of data with other sequencing laboratories hence facilitating ease of collaboration. For example, the naming convention provided by the package is now being used by the center to name sequenced samples. The package is also easy to use since the instructions are clear and intuitive. The fields present in the package are also the same as those required in submission to databases such as GISAID and NCBI GenBank hence is useful for facilitating submission to these databases. It is noted that the fields contained in the template sufficiently cover information that are needed for public health activities.