



University
of Glasgow

Metadata

Essel Bautista

*RAGE workshop,
NCDC, Abuja*

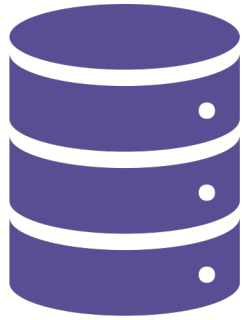
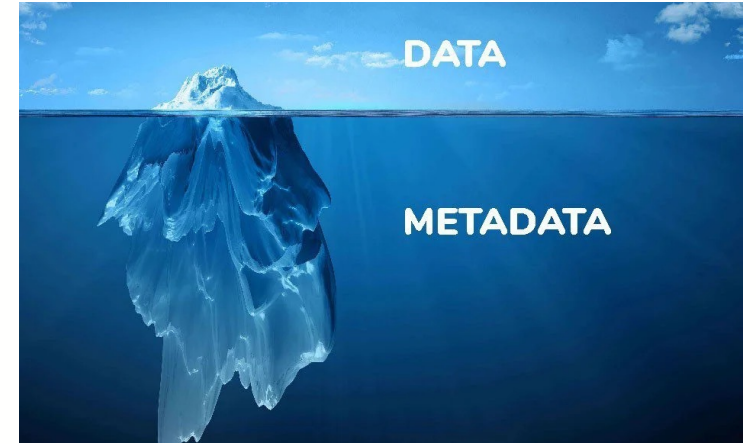
12-16 February 2024



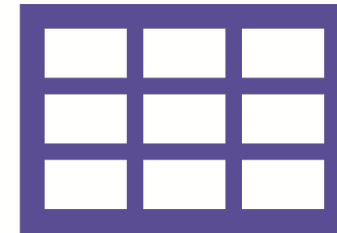


University
of Glasgow

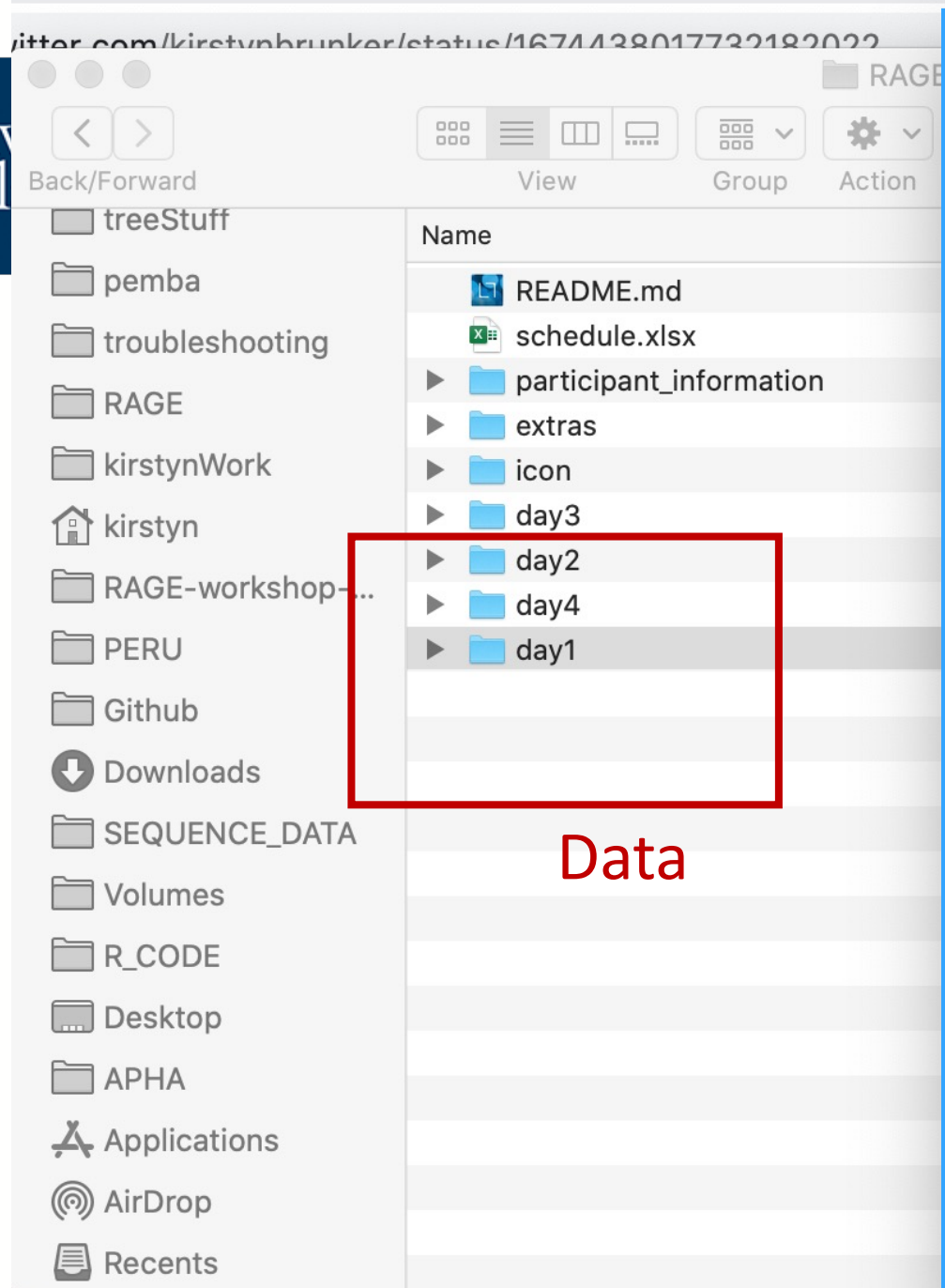
'Data about data'



Description and context of the
data.

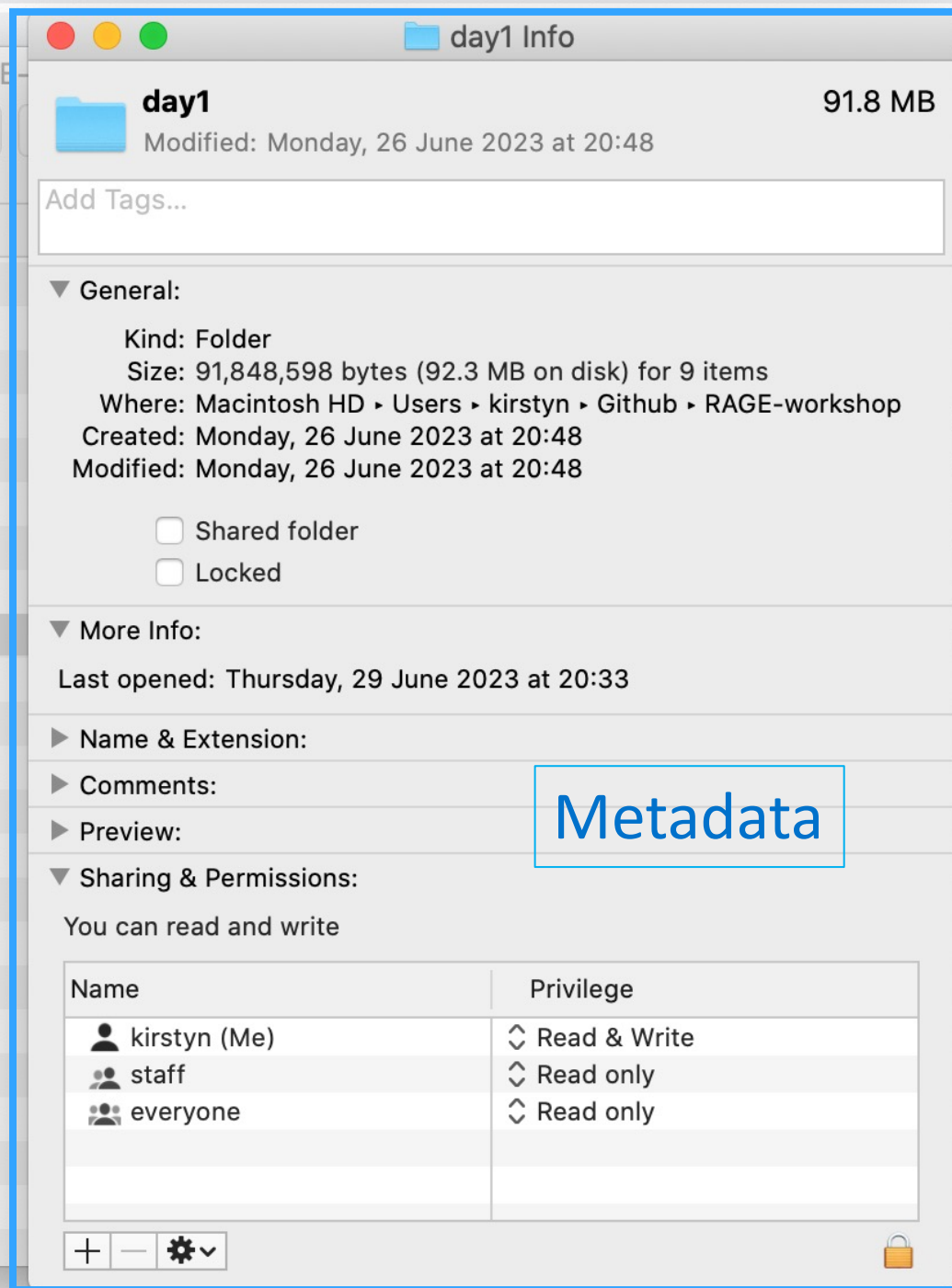


It helps to organize, find and
understand data.



More

Desktop



Conda environment

- 1 Conda environment
- 2 Clone artic-rabv repository
- 3 Create environment
- 4 Activate the environment

Metadata

1 Conda environment

Today we will be using a rabies virus version of the [ARTIC Network's](#) bioinformatics pipeline. This is a pipeline and set of accompanying tools for working with viral nanopore sequence data, generated from tiling amplicon schemes.

This is nicely bundled up for us in a conda environment!

2 Clone artic-rabv repository

First, we need to "clone" the artic-rabv repository from GitHub. Clone essentially means we copy the repository from GitHub.com to your local machine (or in this





Kirstyn Brunker
@kirstynbrunker

Data

We did it! Despite some flowcell problems the RAGE workshop group managed to load 2 @nanopore libraries today. Congrats to @Yusufhusain1989 who won the pot luck to prime and load the flowcell! Nailed it! 👍



Metadata

4:21 PM · Jun 29, 2023 · **1,333** Views

📊 View Tweet analytics

3 Retweets **26** Likes



Why is it important?

Valuable context and information about the samples used in a study to draw meaningful conclusions and share data effectively

1. Experimental Design and Context
2. Data Interpretation and Analysis
3. Quality Control and Data Validation
4. Reproducibility and Data Sharing
5. Ethical Considerations and Compliance



Best practices

- It is absolutely vital that you have good data recording, organisation and validation practices for any metadata
 - Case metadata
 - Sample metadata
 - Sequence metadata



Case metadata

Index Case

	Case #1	Case #2	Case #3	Case #9
Municipality	Santa Maria	San Agustin	Santa Maria	Alcantara
Confirmed case	No (probable)	Yes	Yes	Yes
Biting animal	Dog	Dog	Dog	Dog
Date of biting incident	Nov. 12, 2022	Nov. 14, 2022	Nov. 18, 2022	Sept. 30, 2022
No. of bite victims	1	2	3	1
Status of biting animal	Killed and consumed (Nov. 12)	Expired (Nov. 17)	Killed (Nov. 18)	Expired (Oct. 1)
Sample collected	No	Yes	Yes	Yes
Confirmation test	--	DFAT (Nov. 23)	RDT (Nov. 21) DFAT (Nov. 22)	DFAT (Dec. 7)



Sample metadata

A	AM	AT	AV	AW	AZ	BA	BB	BD
Case number we are using for ordering and reporting	Date the sample is being sent to the laboratories, lab submission form filled by PVO			free text	Testing results from the RITM - qPCR method	Testing results from the RITM - dFAT method		free text
CASE_NO	DATE_LAB_S	TESTS_RADI	DATE_TESTS	ACCESSION_NUMBER_RADDI	TESTS_RITM_qPCR	TESTS_RITM_dFAT	DATE_TESTS_RITM	ACCESSION_NUMBER_RIT
R-9	2022-10-01	RDT+	2022-10-17	00037-22	qPCR+	dFAT+	2022-12-07	4B-22-37_01
R-11		RDT-		00038-22	NA	NA	NA	NA
R-4	2022-10-21	NA		None	qPCR+	dFAT+	2022-12-07	4B-22-39_01
R-2		NA		None	qPCR+	dFAT+	2022-11-23	Z_22-104_01
R-3		NA		None	qPCR+	dFAT+	2022-11-22	Z-22-103_01
R-5	2022-12-05	RDT+	2022-12-05	00041-22	qPCR+	dFAT+	2022-12-07	4B-22-41_01
R-6	2022-11-26	RDT+	2022-12-05	00042-22	qPCR+	dFAT+	2022-12-07	4B-22-42_01
R-7	2022-12-10	RDT-	2022-12-15	00043-22	qPCR+	dFAT+	2023-01-06	4B-43-22
R-15	2022-12-11	RDT+	2022-12-15	00044-22	NA	NA	PENDING	
R-8		Unkno...			qPCR-	dFAT-	2023-01-04	4B-22-46
R-10	2022-12-28	NA		None	qPCR+	dFAT+	2023-01-04	Z-22-121_01
R-12	2022-12-21	NA		None	qPCR-	dFAT-	2023-01-04	Z-22-120_01
R-14	2022-12-28	NA		None	qPCR+	dFAT+	2023-01-04	Z-22-119_01
R-13		NA			qPCR+	dFAT+	2023-01-13	Z-23-005_01



Sample metadata

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	ID	country	year	lineage	brgy	place	province	region	latitude	longitude	species	date	
28	4B-22-45	Philippines	2022	Asian SEA4_	Concepcion Sur	Santa Maria	Romblon	Southwester	12.4031022	122.083657	Canis familia	12/12/2022	
29	4B-23-01	Philippines	2023	Asian SEA4_	Concepcion Norte	Santa Maria	Romblon	Southwester	12.4031022	122.083657	Canis familia	20/01/2023	
31	4B-23-05	Philippines	2023	Asian SEA4_	Sugod	San Agustin	Romblon	Southwester	12.5540355	122.112405	Canis familia	27/01/2023	
32	4B-23-03	Philippines	2023	Asian SEA4_	Sugod	San Agustin	Romblon	Southwester	12.5540355	122.112405	Canis familia	22/01/2023	
33	4B-23-06	Philippines	2023	Asian SEA4_	Concepcion Norte	Santa Maria	Romblon	Southwester	12.4031022	122.083657	Canis familia	31/01/2023	
34	H-23-011Sa11	Philippines	2023	Asian SEA4_	Concepcion Sur	Santa Maria	Romblon	Southwester	12.4031022	122.083657	Homo sapier	02/02/2023	
35	H-23-011Sk12	Philippines	2023	Asian SEA4_	Concepcion Sur	Santa Maria	Romblon	Southwester	12.4031022	122.083657	Homo sapier	02/02/2023	
36	4B-23-07	Philippines	2023	Asian SEA4_	Concepcion Sur	Santa Maria	Romblon	Southwester	12.4031022	122.083657	Canis familia	07/02/2023	
37	4B-23-11	Philippines	2023	Asian SEA4_	Jun Carlo	San Andres	Romblon	Southwester	12.536421	122.046948	Canis familia	09/02/2023	
38	4B-23-12		2023		Concepcion Norte	Santa Maria	Romblon				Canis familia	14/02/2023	
39	4B-23-15	Philippines	2023	Asian SEA4_	Anahao	Odiongan	Romblon	Southwester	12.39866	122.01777	Canis familia	16/02/2023	
40	4B-22-37	Philippines	2022	Asian SEA4_	San Isidro	Alcantara	Romblon	Southwester	12.2978471	122.054934	Canis familia	01/10/2022	
41	4B-22-39	Philippines	2022	Asian SEA4_	San Isidro	Alcantara	Romblon	Southwester	12.2978471	122.054934	Canis familia	21/10/2022	
42	Z-22-103	Philippines	2022	Asian SEA4_	Paroyhog	Santa Maria	Romblon	Southwester	12.4031022	122.083657	Canis familia	21/11/2022	
43	4B-22-41	Philippines	2022	Asian SEA4_	Dapawan	Odiongan	Romblon	Southwester	12.39866	122.01777	Canis familia	24/11/2022	
44	4B-22-42	Philippines	2022	Asian SEA4_	Bonlao	Alcantara	Romblon	Southwester	12.2978471	122.054934	Canis familia	26/11/2022	
45	4B-22-44	Philippines	2022	Asian SEA4_	Dapawan	Odiongan	Romblon	Southwester	12.39866	122.01777	Canis familia	11/12/2022	
46	Z-22-121	Philippines	2022	Asian SEA4_	Carmen	San Agustin	Romblon	Southwester	12.5540355	122.112405	Canis familia	19/12/2022	
47	Z-22-119		2022		Tulay	Odiongan	Romblon				Canis familia	28/12/2022	
48	4B-23-02	Philippines	2023	Asian SEA4_	Dapawan	Odiongan	Romblon	Southwester	12.39866	122.01777	Canis familia	13/01/2023	
52	4B-23-04	Philippines	2023	Asian SEA4_	Carmen	San Agustin	Romblon	Southwester	12.5540355	122.112405	Canis familia	31/01/2023	
53	4B-23-13	Philippines	2023	Asian SEA4_	Rizal	Odiongan	Romblon	Southwester	12.39866	122.01777	Canis familia	09/02/2023	
91	4B-23-16	Philippines	2023	Asian SEA4_	Dapawan	Odiongan	Romblon	Southwester	12.39866	122.01777	Canis familia	28/02/2023	
92	4B-23-17	Philippines	2023	Asian SEA4_	Poctoy	Odiongan	Romblon	Southwester	12.39866	122.01777	Canis familia	01/03/2023	



Sequence metadata

ID	n_N	n_gap	length	year	lineage	previous	
Z-16-151	137	3	11795	2016	Asian SEA4_B1.4.1	Asian SEA4	
Z-16-209	155	3	11777	2016	Asian SEA4_A1.3	Asian SEA4	
Z-16-210	126	3	11806	2016	Asian SEA4_B1.2	Asian SEA4	
Z-16-211	125	3	11807	2016	Asian SEA4_B1.1.1	Asian SEA4	
4A-22-171	326	3	11606	2022	Asian SEA4_B1.2	Asian SEA4	
4A-22-203	662	3	11270	2022	Asian SEA4_A1.1.2	Asian SEA4	
4B-22-37	331	3	11601	2022	Asian SEA4_C1.1	Asian SEA4	
4B-22-39	328	3	11604	2022	Asian SEA4_C1.1	Asian SEA4	
4B-22-44	328	3	11604	2022	Asian SEA4_C1.1	Asian SEA4	
4B-22-45	332	3	11600	2022	Asian SEA4_C1.1	Asian SEA4	
4B-23-01	324	3	11608	2023	Asian SEA4_C1.1	Asian SEA4	
4B-23-02	326	3	11606	2023	Asian SEA4_C1.1	Asian SEA4	
4B-23-03	116	3	11816	2023	Asian SEA4_A1.1.2	Asian SEA4	
4B-23-04	119	3	11813	2023	Asian SEA4_A1.1.2	Asian SEA4	
4B-23-05	335	3	11597	2023	Asian SEA4_C1	Asian SEA4	
4B-23-06	327	3	11605	2023	Asian SEA4_C1.1	Asian SEA4	
4B-23-07	327	3	11605	2023	Asian SEA4_C1.1	Asian SEA4	
4B-23-11	327	3	11605	2023	Asian SEA4_C1	Asian SEA4	
4B-23-12	330	3	11602	2023	Asian SEA4_C1.1	Asian SEA4	
4B-23-13	325	3	11607	2023	Asian SEA4_C1.1	Asian SEA4	



University
of Glasgow

**Consider sharing metadata
with collaborators or
depositing it in public
repositories.**





Validate metadata and perform regular quality control checks.



	A	B	C	D	E
1	Group	Sample	Treatment	Barcode*	Library
2	1	sub6988	No normalisation	70	LIB1
3	1	sub6994	No normalisation	69	LIB1
4	1	sub7011	No normalisation	68	LIB1
5	1	sub7105	No normalisation	67	LIB1
6	1	sub7062	No normalisation	66	LIB1
7	1	negative	No normalisation	65	LIB1
8	2	sub6988	Normalise	71	LIB2
9	2	sub6994	Normalise	72	LIB2
10	2	sub7011	Normalise	73	LIB2
11	2	sub7105	Normalise	74	LIB2
12	2	sub7062	Normalise	75	LIB2
13	2	negative	Normalise	76	LIB2
14	3	sub7064	No normalisation	77	LIB1
15	3	sub7087	No normalisation	78	LIB1
16	3	sub7113	No normalisation	79	LIB1
17	3	sub7151	No normalisation	80	LIB1
18	3	sub7153	No normalisation	81	LIB1
19	3	negative	No normalisation	82	LIB2
20	4	sub7064	Normalise	88	LIB2
21	4	sub7087	Normalise	87	LIB2
22	4	sub7113	Normalise	86	LIB2
23	4	sub7151	Normalise	85	LIB2
24	4	sub7153	Normalise	84	LIB2
25	4	negative	Normalise	83	LIB2
26					



University
of Glasgow

**Include relevant information
such as sample identifiers,
characteristics, and
experimental conditions.**

Sample 1

Blood

Collected: Gartnavel hospital

Received: 12-Jun-2022



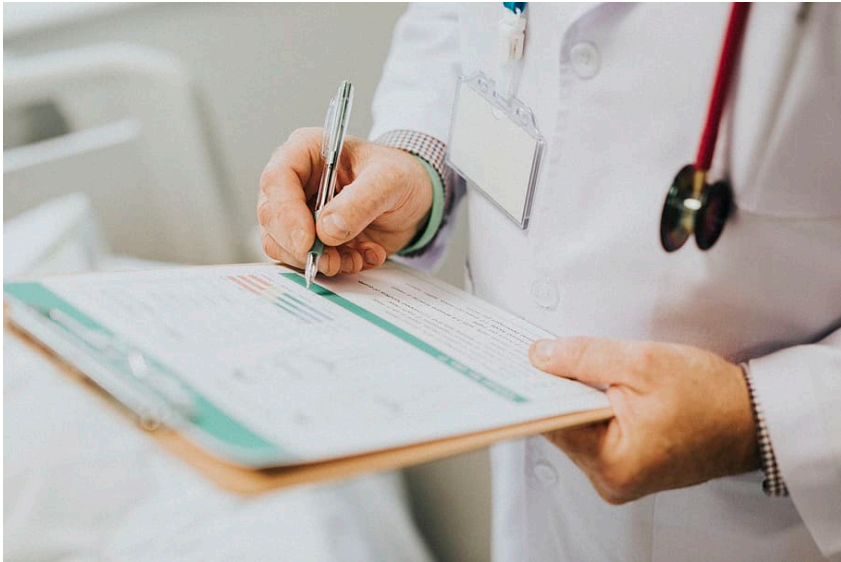


University
of Glasgow

Establish linkages between metadata and associated data files

- Common identifier

Patient A



Sample 1





Utilize standardized terminology and controlled vocabularies.

- Data dictionary

Sample_id	Tissue_type	Host	Date
RV001	Brain	Dog	12/03/21
RV123	Saliva	Canid	01.05.2020
RV432	Brain	Canis familiaris	October 19



University
of Glasgow

**Capture metadata
provenance and maintain a
record of data entry and
updates.**

Source, authorship and
modification history





University
of Glasgow

Backup metadata and use version control.

Can recover data if needed

Trace edits, whether
accidental or not





University
of Glasgow

Implement data security and privacy measures.

- Data protection
- Some data may be sensitive
- Set view, edit and admin privileges





Data formats

- Common source of problems
- Given a choice users will use abbreviations/colloquialisms
- **Standardisation** is so important
- Provide a template and a definition of fields (data dictionary)

	A	B	C	D	
1	Variable / Field Name	Form Name	Section Header	Field Type	Field Label
2	sample_id	diagnostic		text	Field sample id
3	sample_labid	diagnostic		text	Lab sample id
4	sample_sequenceid	diagnostic		text	Sequence sample id
5	sample_batchid	diagnostic		text	Batch id
6	apha_sub	diagnostic		text	APHA submission number
7	apha_rv	diagnostic		text	APHA rabies virus id
8	sample_tissuetype	diagnostic		dropdown	Tissue type



Example: data dictionary

	A	B	C	D	
1	Variable / Field Name	Form Name	Section Header	Field Type	Field Label
2	sample_id	diagnostic		text	Field sample id
3	sample_labid	diagnostic		text	Lab sample id
4	sample_batchid	diagnostic		text	Batch id
5	apha_sub	diagnostic		text	APHA submission number
6	apha_rv	diagnostic		text	APHA rabies virus id
7	sample_tissuetype	diagnostic		dropdown	Tissue type
-

Example: sample metadata

1	sample_id	sample_labid	sample_batchid	apha_sub	apha_rv	sample_tissuetype
34	RB001	Rungubure (RB001)				Brain
35	SD751	SD751				Brain
36	MR009	Ulanga (MR009)				Brain
37	Rab102(D10-1429)1649	RV2776	11-May	SUB1649	RV2776	Brain
38	Rab3/11(D11-145)1656	RV2780	11-May	SUB1656	RV2780	Brain
39	Canine3/10/10(1704)	RV2807	11-May	SUB1704	RV2807	Brain
40	SD185/MBKSDC0611	RV2871	11-Dec	SUB2480	RV2871	Brain
41	SD203/MBRSDC0711	RV2886	11-Dec	SUB2498	RV2886	Brain



Minimal metadata

- Consider your research questions and the type of analysis you want to do
- What data do you need?
- Make sure that data is collected!



Minimal metadata: phylodynamics

- Phylodynamics = analysis that integrates spatio-temporal-genetic data
- Require:
 - Sequences
 - Date associated with case e.g. symptoms started, sample collected
 - Case location e.g. GPS coordinates, village, district

The more resolved this data is the more informative your analysis will be



Summary

- Metadata is extremely important
- Collect the right information for your desired analysis/outputs
- Make your life easier by having good practices in place from the offset