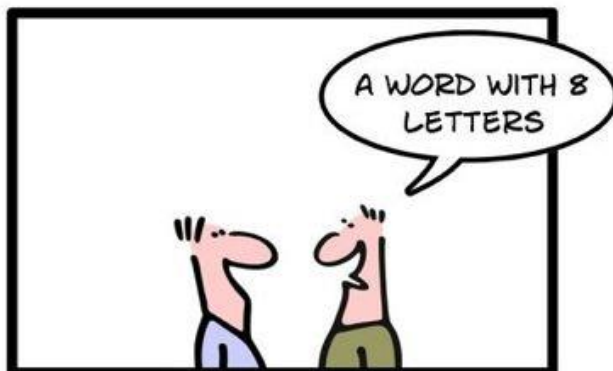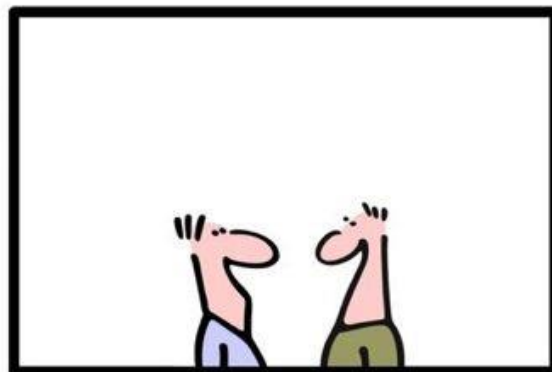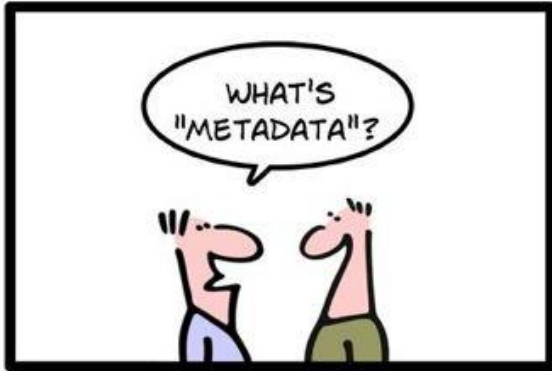# Evolution of Baobab LIMS: coding in metadata standards for infectious disease

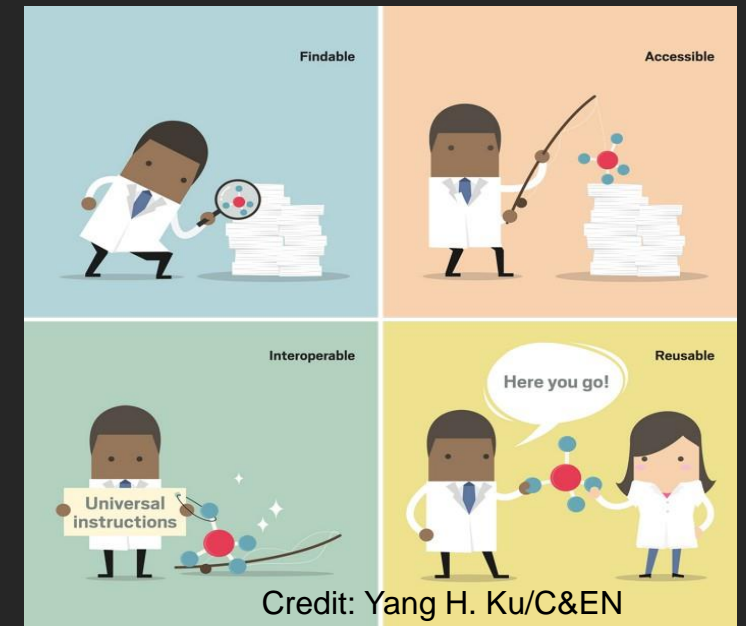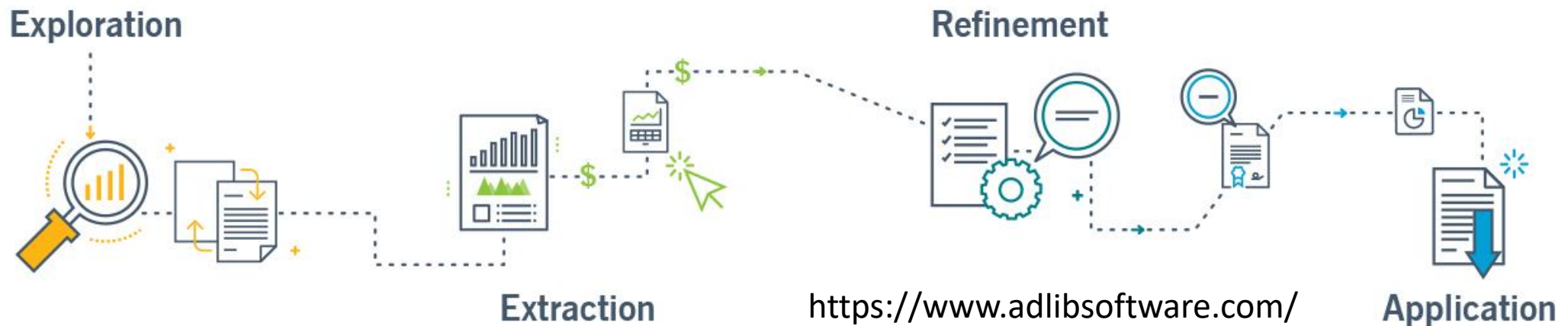Dominique Anderson, PhD

What is metadata and why should we have standards?

- Data about data

- The information needed to find, understand and effectively use the data.

- Key to 'unlock' an information resource

- A metadata standard establishes a commonality
  - Vocabularies
  - Structure
  - Use and interpretation

- Increases the utility of data
  - Interoperability between systems
  - Makes it easier to find relevant data
  - Provides access to data
  - Enhances discovery from data

Credit: Yang H. Ku/C&EN

Metadata management

- The admin of data
  - Integration, access, sharing, linkage, analysis and maintenance
  - Physical or virtual
  - Maximize the potential in an organizations data

Exploration

Extraction

Refinement

Application
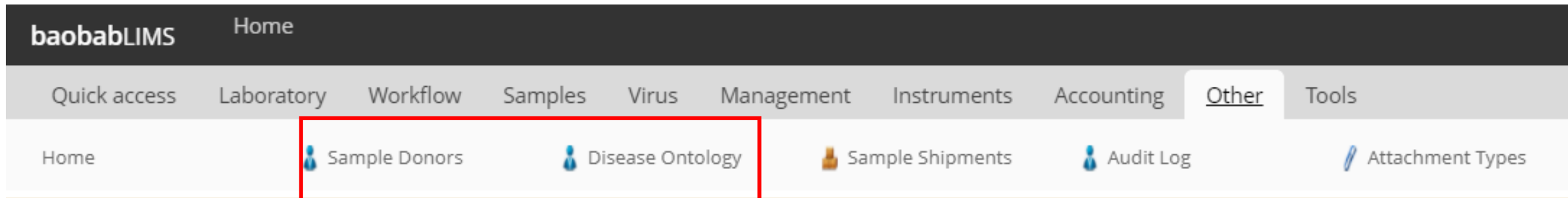
https://www.adlibsoftware.com/
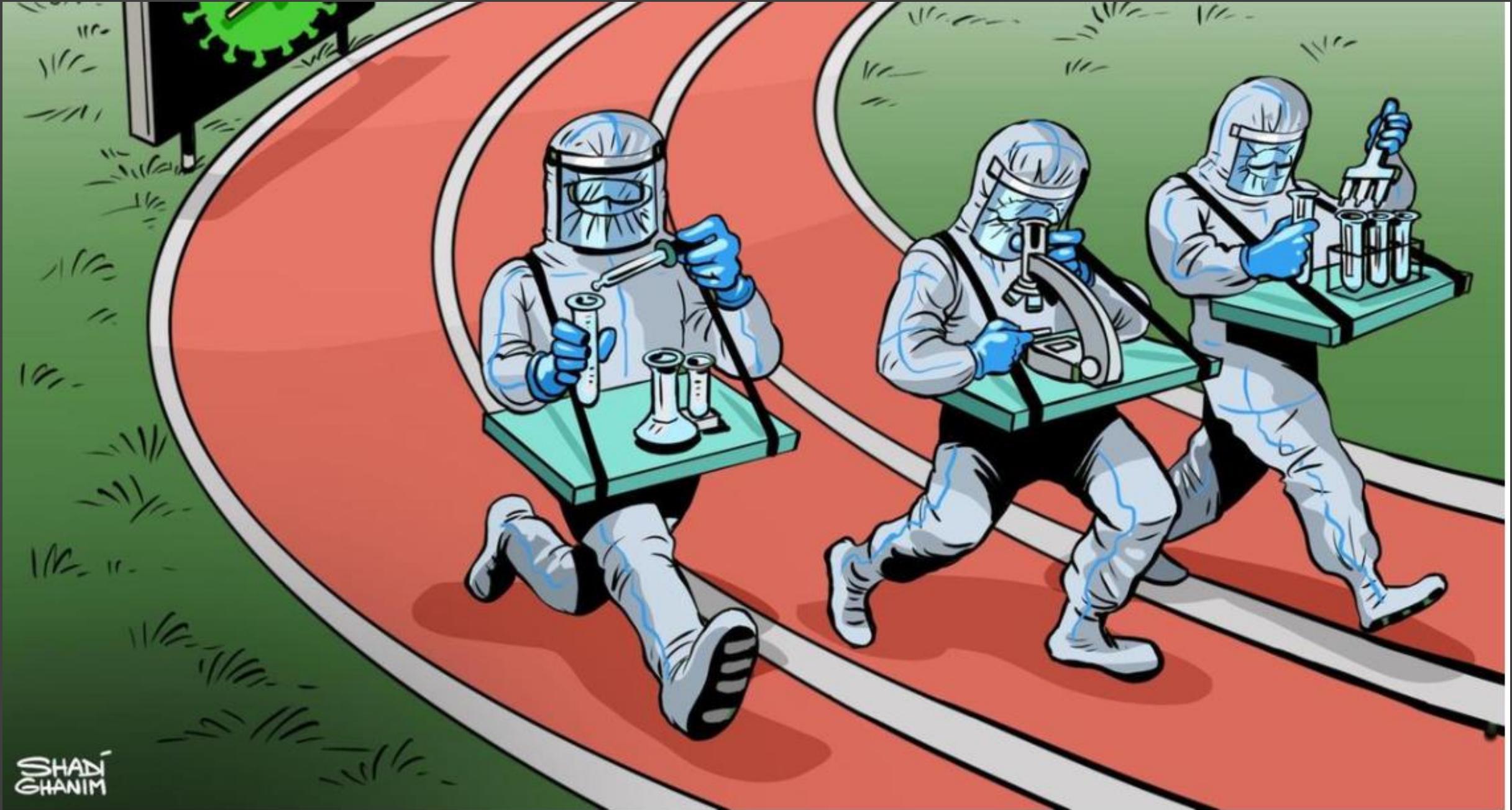
LIMS data management

- Improve data management in lab to increase lab potential

- Sample enters the Laboratory

- Data enters the data management tool
  - Roadmap of the sample
  - Creating valuable linkage
  - Metadata standards make the roadmap readable and universal

Baobab LIMS metadata awareness

- Workflow of a biobank in SA

- Designed with human samples in mind

- Sample lifecycle tracking

MIABIS – Minimum Information About BIobank data Sharing
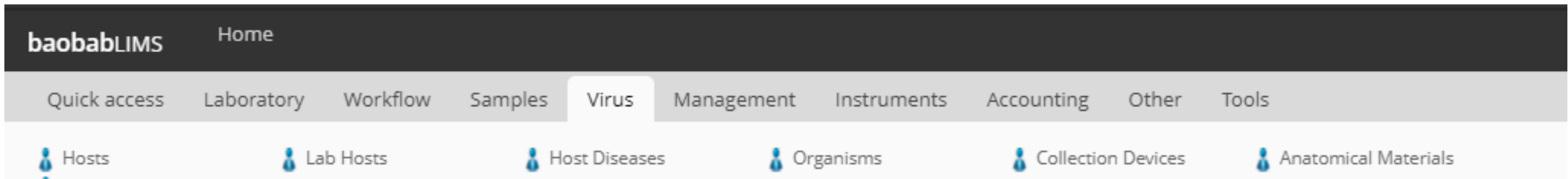
SHADI GHANIM

https://www.thenationalnews.com/

Incorporating the data spec into Baobab LIMS

- Evaluate fields
  - Mandatory, strongly recommended, optional
  - Applicable to Sample LIMS
  - Database/ application interoperability
- Design
  - Keep in mind existing LIMS architecture
  - Aim for simplicity
  - High sample numbers
  - Infrastructure challenges
- Code and test
- Future-proof

Baobab LIMS Virus module is borne

- Configurable content vs hard-coded options

- Import to configure

Left spreadsheet — AnatomicalMaterial

Cell reference: A10 — Ammniotic Fluid

	A	B
1	title	description
2	Anatomical Material	
3	title	description
4	Blood	
5	Fluid	
6	CSF Fluid	
7	Pericardial Fluid	
8	Pleural Fluid	
9	Vaginal Fluid	
10	Ammniotic Fluid	
11	Saliva	
12	Tissue	
13	Not Applicable	
14	Not Collected	
15	Not Provided	
16	Missing	
17	Restricted Access	

Sheet tabs: AnatomicalMaterial | CollectionDevice | Host | HostDisease | LabHost | Organism

Right spreadsheet — Host

Cell reference: G15

	A	B
1	title	description
2	Host	
3	title	description
4	Bos taurus	
5	Canis lupus familiaris	
6	Chiroptera	
7	Columbidae	
8	Felis catus	
9	Gallus gallus	
10	Homo sapiens	
11	Manis	
12	Manis javanica	
13	Panthera leo	
14	Panthera tigris	
15	Rhinolophidae	
16	Rhinolophus affinis	
17	Sus scrofa domesticus	
18	Viverridae	
19	Not Applicable	
20	Not Collected	
21	Not Provided	
22	Missing	
23	Restricted Access	

Sheet tabs: AnatomicalMaterial | CollectionDevice | Host | HostDisease | LabHost | Organism

GT ON OFF   Search Site   Search

Quick access   Laboratory   Workflow   Samples   Virus   Management   Instruments   Accounting   Other   Tools

(i) 1 Instrument's calibration certificate expired:

Cell Counter

You are here: Home

💡 Info   Changes saved.

System Dashboard

Switch to from

Daily   Weekly   Monthly   Quarterly   Biannual   Yearly

Period: Mar 28, 2021 - Apr 04, 2021

Analyses

0   0

of 0 (0.0%)   of 0 (0.0%)
Analyses pending   To be verified

Analysis Requests

0   0   0   0

of 0 (0.0%)   of 0 (0.0%)   of 0 (0.0%)   of 0 (0.0%)
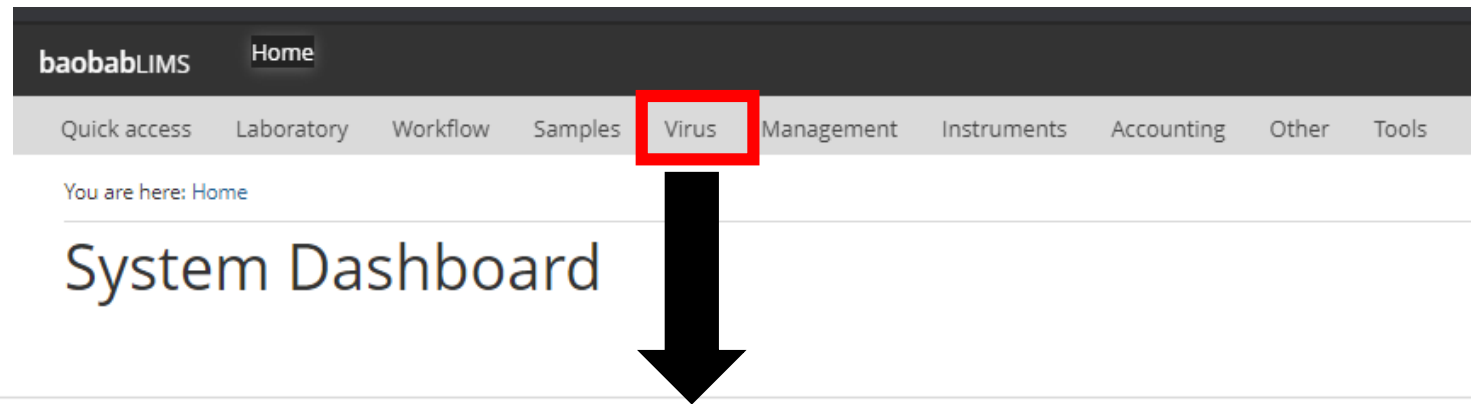Reception pending   Results pending   To be verified   To be published

Worksheets

Adding a sample

- Default fields

- Template fields – related to sample

- Add system interoperability

GT **ON** OFF

Quick access   Laboratory   Workflow   Samples   Virus   Management   Instruments   Accounting   Other   Tools

ⓘ 1 Instrument's calibration certificate expired:

Cell Counter

You are here: Home › Virus Samples

👤 Virus Samples   Add

ⓘ No items found

File | Home | Insert | Page Layout | Formulas | Data | Review | View | Help

A5 | E4002

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	barcode	Project		Sample Type	Storage L	Volume	Unit	Anatomic:	Allow Sha	Will Retur	Bio Samp	Specimen	Sample C	Sample C	Sample R	Geo Loc (	Geo Loc !	Organism	Isolate	Purpose (
2	Virus Sample																			
3	title	Project		SampleType	StorageL	Volume	Unit	Anatomic:	AllowShai	WillReturi	BioSampl	Specimen	SampleCc	SampleCc	SampleRe	GeoLocC	GeoLocS	Organism	Isolate	PurposeC
4	E4001	Green Goblin Project		Whole Blood	Room-2.Fr	12	ml	Arm			A1001	C1001	James Car	########	########	Australia	Canberra	Rocket	Tresae	Research
5	E4002	Green Goblin Project		Whole Blood	Room-2.Fr	23	g	Arm			A1002	C1002	Spies Jan	########	########	South Afri	Western C	Rocket	Wormwoo	Restricted
6																				

AnatomicalMaterial | CollectionDevice | Host | HostDisease | LabHost | Organism | **VirusSample**

Quick access   Laboratory   Workflow   Samples   Virus   Management   Instruments   Accounting   Other   Tools

ⓘ 1 Instrument's calibration certificate expired:
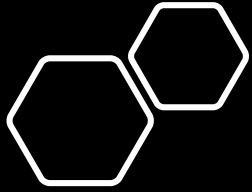
Cell Counter

You are here: Home › Viral Genomic Analyses

👤 Viral Genomic Analyses   [Add]

ⓘ No items found

Where to next?

- Need user input and external testing
- Metadata for bacteria
- Needs to be measured against system complexity
  - System speed vs infrastructure availability
- Modular design?
- Sustainability issues

https://xkcd.com/927/

Resources

- Follow us on twitter
  - @BaobabLIMS

- *Website*
  - www.baobablims.org

- Get the code (and more)
  - https://github.com/BaobabLims

- Send us an email
  - Training – dominique@sanbi.ac.za
  - Helpdesk – help@baobablims.org

- Hands on training event – April 2021

Acknowledgements

- Baobab LIMS development team
- South African Medical Research Council
- Africa CDC
- PHA4GE
- South African Department of Science and Innovation
- NRF
- South African Bioinformatics Institute
- B3 Africa and the EU
- Thank you to all participants