
Administration, Coordination and Management 3

SOP: Reimbursements from the University of Cape Town (SADaCC)

PURPOSE:

To describe the procedure for obtaining reimbursements from the University of Cape Town (SADaCC).

SCOPE

This procedure applies to members of SPARCO and SADaCC who might incur costs on behalf of SADaCC.

GLOSSARY/DEFINITIONS

SADaCC Sickle Africa Data Coordinating Centre at the University of Cape Town

SOP Standard Operating Procedure: Detailed written instructions to achieve uniformity of the performance of a specific function.

RESPONSIBILITIES & PROCEDURES

1. Assess whether service provider cannot be paid directly from SADaCC

Before incurring costs, liaise with SADaCC to check if the service provider cannot be paid directly from SADaCC, personal reimbursements will only be made where service providers cannot be paid directly.

2. Approval from SADaCC

Before you incur costs, liaise with the SADaCC coordinating center to get the go ahead.

3. Submitting Documents

Keep all receipts for costs incurred. Where possible please make payment using your card so that the transaction reflects in your bank account.

Submit:

- a) Receipt
- b) Your bank statement showing a transaction that corresponds to the amount on the receipt
- c) A bank certified account details which the transfer will be deposited to (if this is applicable).
- d) For your own information – SADaCC will submit an fm049 form to the University of Cape Town together with your supporting documents.

Version No.	Date	Authorizer	External Reviewer	Internal Reviewer(s)	Author	Details of changes
					Vicky Nembaware	n/a (first version)



Administration, Coordination and Management 1

SOP: Writing & Versioning of Documents

PURPOSE

To describe the procedure for authorship, review, authorisation, issue and control of documents.

SCOPE

This procedure applies to the development of documents for public access, including standard operating procedures (SOPs) by the Sickle Africa Data Coordinating Centre (SADaCC). It does not apply to contact lists, agendas, minutes and 'living' documents that are used only for internal purposes.

GLOSSARY/DEFINITIONS

SADaCC	Sickle Africa Data Coordinating Centre at the University of Cape Town
SOP	Standard Operating Procedure: Detailed written instructions to achieve uniformity of the performance of a specific function.

RESPONSIBILITIES & PROCEDURES

- The SADaCC leadership will delegate authorship and review of documents including SOPs to appropriately qualified and experienced personnel.
- Each document has 'SADaCC', 'SPARCO' and 'SPAN' logos its header.
- Each document explains acronyms/abbreviations in full the first time they are used and/or includes a glossary.
- Processes that are mandatory are described using 'should', 'must' etc., while processes that are recommended are described using 'may', 'could' etc.
- The first draft version of a document is labelled 0.1 and the first signed-off version of a document is labelled published version 1.0. Minor changes are indicated by .1 increment in the version numbers, until the next signed off version is indicated by a 1-point increment.
- The name and version (including the date) of each document should be clearly indicated in the document footer. The page number should also be clearly indicated in the footer.
- Each document should have a named author, internal reviewer, external reviewer and authorizer. Sometimes a group can act as author or reviewer. The author develops a draft/amended document and forwards to the internal reviewer. The reviewer proof-reads the document, checks that the format and version are appropriate and provides suggestions for improvements to the author. The

author then forwards the document to the external reviewer. The external reviewer also proof-reads the document, checks that the format and version are appropriate and provides further suggestions for improvements to the author. Sometimes a second level of external review may be necessary, following the same process. Once finalized, the author forwards the document to the authorizer. The authorizer signs off the document and the document status is changed to 'published'.

E.g.:

Document Name_Draft version 0.1 (28 Jan 2017) – by author

Document Name_Draft version 0.2 (14 Feb 2017) – by internal reviewer

Document Name_Draft version 0.3 (21 Feb 2017) – reworked by author

Document Name_Draft version 0.4 (28 Feb 2017) – by external reviewer

Document Name_Published version 1.0 (03 Mar 2017) – by authorizer

- The document's published version history should be summarized at the end of each document.

E.g.:

Version No.	Date	Authorizer	External Reviewer	Internal Reviewer	Author	Details of changes
1.0	03 Mar 2017	Dr James Smith	Mrs Norma Jones	Dr John Black	Dr Jack Adams	n/a (first version)
2.0	09 Mar 2018	Dr James Smith	Mrs Norma Jones	Dr Jenna Johnason	Dr Jack Adams	Updated section X to reflect Y.

Version No.	Date	Authorizer	External Reviewer	Internal Reviewer	Author	Details of changes
				Ambroise Wonkam	First Data Management Workshop participants	n/a (first version)

Administration, Coordination and Management 2

SOP: Site Visits

PURPOSE

To describe the procedure for site visits.

SCOPE

This procedure applies to research site visits conducted by the Sickle Africa Data Coordinating Centre (SADaCC).

GLOSSARY/DEFINITIONS

SADaCC	Sickle Africa Data Coordinating Centre at the University of Cape Town
SOP	Standard Operating Procedure: Detailed written instructions to achieve uniformity of the performance of a specific function.

RESPONSIBILITIES & PROCEDURES

1. Designate a site visit leader

Before the visit, the visitors will designate among themselves a site visit leader that will be responsible for making final arrangements with the sites and lead the following procedures at each site visit.

2. Set up the site visit

Send SOPs and information about the purpose of the particular visit to the site Principal Investigator (PI). Answer any questions and address any concerns the site may have about the impending visit. Arrange where and when to meet.

3. Meeting with key personnel

There will be a brief introductory meeting (± 20 min) with key personnel at the site, e.g. the site PI, the site research coordinator, the data manager and the Information Technology (IT) representative. The meeting will be used to make introductions and discuss the purpose of the particular visit.

4. Walkabout of facilities

The visitors will then take a brief walk around the site to see where key areas are located in relation to the data offices/spaces and to observe some of the activities taking place in the site. Take photos of relevant areas and persons. (Get photo-release forms signed when photos are taken.)

5. Completion of site assessment checklist

The visitors will then need the data team to help them complete a site assessment checklist with questions particular to data management procedures and IT setups at the site. Discuss aspects of site assessment that will be done 'online' with the site personnel.

6. Training of personnel

The visitors will review training records at the site and assess further training needs. Conduct training with site personnel, if possible.

7. Compile report on visit

The site visit leader will ensure that a report on the entire visit is compiled within 7 days of returning.

Version No.	Date	Authorizer	External Reviewer	Internal Reviewer(s)	Author	Details of changes
				Vicky Nembaware, Mario Jonas, Furahini Tluway, Gaston Mazunda	Annemie Stewart	n/a (first version)

Version No.	Date	Authorizer	External Reviewer	Internal Reviewer	Author	Details of changes
1.0				Mario Jonas	Annemie Stewart	n/a (first version)

Data and Database Management 1

SOP: Username & Password Management

PURPOSE

To explain the procedures of proper username and password management.

SCOPE

This procedure applies to any research studies conducted by collaborators in the Sickle Africa Data Coordinating Centre (SADaCC) project.

GLOSSARY/DEFINITIONS

Electronic Platform	database, website or other computer-based portal or system
Password	unique code created by the person for their particular username, in order to access an electronic platform securely
SADaCC	Sickle Africa Data Coordinating Centre at the University of Cape Town
SOP	Standard Operating Procedure: Detailed written instructions to achieve uniformity of the performance of a specific function.
User	natural, named, individual person
Username	unique name or number issued to a user who needs to access an electronic platform

RESPONSIBILITIES & PROCEDURES

Each user's username and password combination must be unique and stands in the place of their handwritten signature. Therefore:

- Usernames must never be used by more than one person. ie users must not share their username and password.
- Passwords must not be revealed to any other persons.
- Users must never write down their username and password in an unsecure place.
- Users must never save their passwords to the internet browsers of unsecured computers, i.e. computers that do not require a unique login on start-up.
- If a user suspects that their password has been compromised in anyway, they must immediately reset their password.
- If a user forgets their password, they must immediately reset their password.

- Users must log out whenever they have finished their session.
- Passwords should be strong (alphanumeric), but easy to remember.

Version No.	Date	Authorizer	External Reviewer	Internal Reviewer(s)	Author	Details of changes
				Daniel Kadonga, Olivia Safari	Annemie Stewart	n/a (first version)

Data and Database Management 2

SOP: Set up of User Rights & Data Access Groups

PURPOSE

To provide instruction to those data management team members delegated responsibility for assigning user rights to research staff and collaborators.

SCOPE

This procedure applies to any research studies conducted by collaborators in the Sickle Africa Data Coordinating Centre (SADaCC) project.

GLOSSARY/DEFINITIONS

REDCap	Research Electronic Data Capture (REDCap) is a secure web application for building and managing online surveys and databases. While REDCap can be used to collect virtually any type of data (including 21 CFR Part 11, FISMA, and HIPAA-compliant environments), it is specifically geared to support online or offline data capture for research studies and operations.
SADaCC	Sickle Africa Data Coordinating Centre at the University of Cape Town
SOP	Standard Operating Procedure: Detailed written instructions to achieve uniformity of the performance of a specific function.
SPARCO	Sickle Pan African Research Collaboration (SPARCO) is a sister collaboration of SADaCC, existing of multiple sites across three countries, including one site per country that acts as a country-level data centre or 'hub'.

RESPONSIBILITIES & PROCEDURES

1. The Principal Investigator (PI) will delegate responsibility for assigning user rights to research staff and collaborators of the project to an appropriately experienced data manager.
2. The data manager will create appropriate User Roles that specify the level of permissions for project setup, data exports, data resolution workflow, creating/renaming/deleting records, locking/unlocking records, and data collection instruments. The User Roles will be documented in the database specifications document.
3. The data manager will also create Data Access Groups (DAGs) to represent each participating local site and use consistent DAG naming across all projects.

4. The data manager will then add users to the project in by assigning them to the appropriate role and DAG according to the delegation log.

Version No.	Date	Authorizer	External Reviewer	Internal Reviewer(s)	Author	Details of changes
				Raphael Sangeda, Mario Jonas	Annemie Stewart	n/a (first version)

Data and Database Management 3

SOP: Data Recording

PURPOSE

To describe the procedure for recording source and Case Record Form (CRF) data for research studies (and associated activities).

SCOPE

This procedure applies to any research studies conducted by collaborators in the Sickle Africa Data Coordinating Centre (SADaCC) project.

GLOSSARY/DEFINITIONS

CRF	Case Record/Report Form: A printed, optical, or electronic document designed to record data on each trial participant during the course of the trial as defined by the protocol. The data should be collected by procedures, which guarantee preservation, retention and retrieval of information and allow easy access for verification, audit and inspection.
Essential documents	Documents which individually and collectively permit evaluation of the conduct of a clinical research study and the quality of the data produced.
ISF	Investigator Site File: Files of Essential Documents held by the Investigator.
SADaCC	Sickle Africa Data Coordinating Centre at the University of Cape Town
SOP	Standard Operating Procedure: Detailed written instructions to achieve uniformity of the performance of a specific function.
Source data	Source documents/data are original records and certified copies of original records of clinical findings, observations, or other activities in a clinical trial necessary for the reconstruction and evaluation of the trial. Source data are contained in source documents which are original documents, data, and records (e.g., hospital records, clinical and office charts, laboratory notes, memoranda, subjects' diaries or evaluation checklists, pharmacy dispensing records, recorded data from automated instruments, copies or transcriptions certified after verification as being accurate copies, microfiches, photographic negatives, microfilm or magnetic media, x-rays, subject files, and records kept at the pharmacy,

at the laboratories and at medico-technical departments involved in the clinical research study).

RESPONSIBILITIES & PROCEDURES

GENERAL

- It should be clear as regards to who completes original information, who transcribes data to a case record/report form (CRF) who performs quality control checks (if relevant) and who enters data into a database.
- Signatures and dates should be in wet ink or through a validated electronic method. Pre-typed dates or stamps are not acceptable. Dates should have a consistent format, preferably dd-mm-yyyy. For time, use a 24 hour clock e.g. 09:31 for 31 minutes past 9 in the morning, 15:05 for 5 minutes past 3 in the afternoon. NB: Midnight is 00:00.
- All documentation relating to data recording and quality control will be kept in the Investigator Site File (ISF).

SOURCE DATA

Study Specific Source documents

- Study specific source documents are designed to reflect data that is protocol specific. These documents must be clearly identified, i.e., Header and Footers of documents should contain the name of the project, name of the document, protocol number, filename (including version control), references to relevant Standard Operating Procedures and page numbers. It should also contain the study visit number, space to add participant initials, participant number and the visit date. An Investigator will sign and date the source document once all the information needed is recorded.
- Source data should preferably be in black ink. Entries should be signed/dated at least on a visit basis and incorrect entries deleted with a single line. The correction should have an initial/date (and reason if not obvious). Alternatively each correction can be numbered with the initial and date entered in a footnote.

Furthermore:

- Unless otherwise specified, participant initials should be recorded with three letters e.g. 3 parts to name = RDW, 2 parts = R-W.
- Results of assessments, including clinical observations, laboratory assays, ECG traces etc., are reviewed by an investigator who documents clinical relevance with his/her signature and date. The process and review, however, may be study-specific.

- As soon as possible after source data are recorded a designee will review them to identify obvious errors and omissions, internal consistency and relation to exclusions/withdrawal criteria.
- Errors detected should be discussed with an investigator as soon as possible, and corrective actions documented.
- If participants are contacted to clarify issues relating to data, the date and content of the conversation should be recorded in his/her file.
- Source documents stored between visits in a safe and confidential manner, and made available to team members or the external monitor as required.
- A log of protocol deviations relating to data recording should be maintained throughout the study.

PAPER CRF DATA

- Paper CRFs will only be completed for enrolled subjects unless otherwise decided by the PI or Sponsor. Original and corrected entries into the CRF are maintained as above. Clinical data should not be entered into a CRF before it has been reviewed by an investigator. CRF entries will be logged on a log.
- Prior to the investigator signing declarations of completeness in a CRF, a designated team member may oversee a review of all, or a pre-specified selection of, CRF data for accurate transcription. This check will be recorded on a data entry quality check log. Corrections made to CRF data after the investigator has signed as above should be counter-signed and dated by an investigator.

ELECTRONIC CRF (ECRF) DATA

- In the absence of any Sponsor-specific procedures the following procedure may be used or adapted as per trial-specific requirements:
- eCRFs will only be completed for enrolled subjects unless otherwise decided by the PI or Sponsor.
- A member of the team designated to perform data entry direct from source documents will ensure that the relevant data have been declared complete by the investigator.
- Data entry should be documented in a log.
- Prior to the investigator (electronically or otherwise) signing declarations of completeness in an eCRF, a designated team member may oversee a review of all, or a pre-specified selection of, CRF data for accurate transcription. This check will be recorded on a data entry quality check log.

Version No.	Date	Internal Reviewer(s)	Author	Details of changes
		Gaston Mazunda	Annemie Stewart	n/a (first version)

Assessment of Haematological Parameters 1

SOP: Measure - Complete Blood Count

DEFINITION

A Complete Blood Count (CBC) is a standardized test to assess components of the blood (e.g., red blood cells, white blood cells, hematocrit, hemoglobin, platelets).

PURPOSE

Complete Blood Count (CBC) tests are performed to determine the overall health of an individual and may help identify underlying conditions. CBC test results may identify a broad range of common and rare diseases.

SCOPE

This procedure applies to any research studies conducted by collaborators in the Sickle Pan African Research Consortium and the Sickle Africa Data Coordinating Centre (SADaCC) project.

DESCRIPTION OF PROTOCOL

The Complete Blood Count (CBC) protocol is a standard blood panel from the National Health and Nutrition Examination Survey (NHANES) and performed on participants aged 1 and older. Blood is collected from participants via standard venipuncture in a 3 or 4 ml K₃ EDTA tube. The ambient blood sample is placed into a hematological analyzer for CBC and differential analyses. The blood should be tested as soon as possible after venipuncture and no later than 24 hours.

SPECIFIC INSTRUCTION

Blood should be drawn directly into an EDTA anticoagulated tube. Venipuncture is preferred rather than capillary blood draw although arterial blood may be used. The tube should be filled to the maximum volume stated.

Although this assay specifies the Coulter® HMX Hematology Analyzer, there are a number of different assays and instruments that are appropriate to perform the CBC analyses. Commercial laboratories perform CBC tests routinely, and they are inexpensive. Blood should be analyzed according to the instrument manufacturer's instructions and according to the laboratory's standard operating procedure and interpreted in the context of age specific normal values. Once an assay is chosen for a particular study, the Working Group recommends that no changes in the protocol be made over the course of the study. To aid comparability, the Working Group

recommends that the investigator record the make and manufacturer of equipment used and the repeatability and coefficients of variation for the assay.

Investigators who want to include participants that have hemophilia or have received cancer chemotherapy in the last 4 weeks will need to implement special venipuncture procedures.

PROTOCOL

Exclusion Criteria

Persons will be excluded from this component if they:

- Report that they have hemophilia; or
- Report that they have received cancer chemotherapy in the last 4 weeks

SP = Sample Person.

1. Do you have hemophilia?

Yes

No

Don't know

If the SP answers, "Yes," the SP is excluded from the blood draw.

If SP answer "No" or "Don't Know," blood is drawn from the SP.

2. Have you received cancer chemotherapy in the past four weeks or do you anticipate such therapy in the next four weeks?

Yes

No

Don't know

If the SP answers, "Yes," the SP is excluded from the blood draw.

If SP answer "No" or "Don't Know," blood is drawn from the SP.

Venipuncture Procedures

Venipuncture must be done by a qualified phlebotomist with the knowledge and skill required for this procedure. A full description of phlebotomy procedure is found on chapter 4 of the Laboratory Procedures Manual. Fill a 3 or 4 ml K₃ EDTA tube with blood.

Recording the Results of the Venipuncture Procedure

Immediately after completing the venipuncture, record the results of the blood draw, the reasons for a tube not being drawn according to the protocol, and any comments about the venipuncture.

Perform CBC Analyses

Note: a full description of this procedure is found in the 2009-2010 NHANES Lab Procedures Manual.

The following parameters and units are measured in the CBC panel test:

White blood cell count (1000 cells/uL)	Basophils	number	(1000
Lymphocyte (%)	cells/uL)		

Monocyte (%) Segmented neutrophils (%) Eosinophils (%) Basophils (%) Lymphocyte number (1000 cells/uL) Monocyte number (1000 cells/uL) Segmented neutrophils number (1000 cells/uL) Eosinophils number (1000 cells/uL)	Red cell count (million cells/uL) Hemoglobin (g/dL) Hematocrit (%) Mean cell volume (fL) Mean cell hemoglobin (pg) MCHC (g/dL) Red cell distribution width (%) Platelet count (1000 cells/uL) Mean platelet volume (fL)
---	---

Interpretation of the CBC Results

The CBC results will be interpreted based on the machine being used in the specific laboratories.

Selection Rationale

A complete blood count is done by a machine and the numbers and indices may suggest a range of underlying diagnoses. Examination of a blood smear or additional testing is likely to be necessary to confirm a specific diagnosis.

Personnel and Training Required

Phlebotomist

Laboratory that can perform Complete Blood Count (CBC) test

Equipment Needs

Phlebotomy supplies

Version No.	Date	Internal Reviewer(s)	Author	Details of changes

Assessment of Haematological Parameters 2

SOP: Measure - Reticulocyte Count

DEFINITION

A bioassay to measure the number of reticulocytes (immature red blood cells).

PURPOSE

Reticulocyte count measures how fast red blood cells are made. Elevated reticulocyte counts are associated with blood loss, hemolysis and hemolytic anemia, and in response to the anemia of clinical conditions such as sickle cell disease. Reticulocyte counts can also be used to measure efficacy of treatments (e.g., iron replacement therapy).

DESCRIPTION OF PROTOCOL

This protocol provides instructions for drawing, processing and storing blood according to the National Health and Nutrition Examination Survey (NHANES) methods. Because there are many comparable assays and instruments for measuring reticulocyte count, the protocol also provides basic guidelines to aid comparability among different studies.

SPECIFIC INSTRUCTIONS

The National Health and Nutrition Examination Survey (NHANES) instructions for drawing, processing, and storing blood provide a standard methodology used successfully for many years to ensure comparable results across study sites. However, the Sickle Cell Disease Cardiovascular, Pulmonary, and Renal Working Group notes that certain aspects (e.g., exclusion criteria) of the NHANES protocol are study specific and might not be applicable to all types of studies (e.g., sickle cell disease). Investigators who want to include participants that have hemophilia or have received cancer chemotherapy in the last 4 weeks will need to implement special venipuncture procedures.

Reticulocyte count can be combined with other indirect markers of hemolysis ([Aspartate Aminotransferase Level](#), [Haptoglobin Level](#), [Lactate Dehydrogenase Level](#), and [Bilirubin Level](#)) to derive a hemolytic component for sickle cell disease patients. Reticulocyte count analysis is performed on anticoagulated blood, collected in an EDTA tube.

PROTOCOL

The following is a summary version of the full National Health and Nutrition Examination Survey 2011-2012 protocol.

Exclusion Criteria

Persons will be excluded from this component if they:

- Report that they have hemophilia; or
- Report that they have received cancer chemotherapy in the last 4 weeks

SP = Sample Person.

1. Do you have hemophilia?

Yes

No

Don't know

If the SP answers, "Yes," the SP is excluded from the blood draw.

If SP answer "No" or "Don't Know," blood is drawn from the SP.

2. Have you received cancer chemotherapy in the past four weeks or do you anticipate such therapy in the next four weeks?

Yes

No

Don't know

If the SP answers, "Yes," the SP is excluded from the blood draw.

If SP answer "No" or "Don't Know," blood is drawn from the SP.

Venipuncture Procedures

See Protocol for CBC above

Recording the Results of the Venipuncture Procedure

See Protocol for CBC above

Laboratory Assay for Reticulocyte Count

The Sickle Cell Disease Cardiovascular, Pulmonary, and Renal Working Group notes that there are a number of different assays and instruments that are appropriate to measure the reticulocyte count from serum. Once an assay is chosen for a particular study, the Working Group recommends that no changes in the protocol be made over the course of the study. To aid comparability, the Working Group recommends that the investigator record the make and manufacturer of equipment used and the repeatability and coefficients of variation for the assay.

Reference Ranges for Reticulocyte Count:

Children: 3 - 6%

Adults: 0.5 - 1.5%

Percentage of reticulocytes and absolute reticulocyte counts (ARC) are given by the following formulas:

Reticulocyte Percentage = [Number of Reticulocytes (thousands/ μ L) x RBC count (millions/ μ L)] X 100

Absolute reticulocyte count (thousands/ μL) = reticulocyte % x RBC count (millions/ μL) x 10

Version No.	Date	Internal Reviewer(s)	Author	Details of changes

SOP: Data Migration from SPARCO site data to new version of standard SADaCC dictionary

PURPOSE

To describe the procedure for migration of existing data onto a new data dictionary in REDCap.

SCOPE

This procedure applies to the members of SADaCC and SPARCO (see below).

GLOSSARY/DEFINITIONS

REDCap	Research Electronic Data Capture (REDCap) is a secure web application for building and managing online surveys and databases. While REDCap can be used to collect virtually any type of data (including 21 CFR Part 11, FISMA, and HIPAA-compliant environments), it is specifically geared to support online or offline data capture for research studies and operations.
SADaCC	Sickle Africa Data Coordinating Centre at the University of Cape Town
SOP	Standard Operating Procedure: Detailed written instructions to achieve uniformity of the performance of a specific function.
SPARCO	Sickle Pan African Research Collaboration (SPARCO) is a sister collaboration of SADaCC, existing of multiple sites across three countries, including one site per country that acts as a country-level data centre or 'hub'.

RESPONSIBILITIES & PROCEDURES

The Principal Investigator (PI) will delegate responsibility for data migration to an appropriately experienced data manager.

1. The data manager exports the current Data Dictionary from REDCap, through the following steps:
 - Inside the project, under Project Setup tab, click on 'Download the current Data Dictionary' link:

[Project Home](#) | [Project Setup](#) | [Other Functionality](#) | [Project Revision History](#) | [Edit proj.](#)

Project status: ✔ Production Completed steps 4 of 7

Main project settings

✔ Complete!

Enable Use surveys in this project? [?](#) [VIDEO: How to create and manage a survey](#)

Enable Use longitudinal data collection with defined events? [?](#)

[Not complete?](#)

Design your data collection instruments

↻ In progress

Add or edit fields on your data collection instruments. This may be done by either using the Online Designer (online method) or by uploading a Data Dictionary (offline method). Quick links: [Download PDF of all instruments](#) OR [Download the current Data Dictionary](#)

Go to or

Learn how to use [Smart Variables](#) [Piping](#) [Action Tags](#)

- This action will download a .csv file containing the Data Dictionary.
 - Save the .csv file to computer/network storage.
2. Select an analysis software, e.g. Stata, R, SPSS, or another application in which subsequent renaming and recoding of variables can be **scripted/saved**. This will now be called the 'chosen application' for the purpose of the rest of this document. It is absolutely crucial that the data manipulation is documented. In Stata, this is done using a 'do-file', in R or in SPSS this is done using a 'script'.
 3. The data manager exports the current data set from REDCap, through the following steps:
 - Inside the project, under the 'Applications' section in the left-hand bar, click on 'Data Exports, Reports & Stats'.
 - Use the default 'All data' report, and click on 'Export Data':

My Reports & Exports		
	Report name	View/Export Options
A	All data (all records and fields)	<input type="button" value="View Report"/> <input checked="" type="button" value="Export Data"/> <input type="button" value="Stats & Charts"/>

- Then select the 'CSV / Microsoft Excel (raw data)' format or a format associated with the chosen application (e.g. R) and ensure that the 'Export Data Access Group' name is selected (checked) under 'Additional export options', then click 'Export Data'.
- This action will download a .csv file containing the data set, and in the case of a format associated with the chosen application, also a script file.
- Save the file(s) to computer/network storage.

4. The data manager then prepares the existing data for import onto the new data dictionary, through the following steps, ensuring to **record all actions in the script file**:
- Start / Open the script file in the chosen application.
 - Import the .csv file of current data set into the chosen application.
 - Begin by renaming variables/items to match the new data dictionary as necessary (Use python script as advised below)
 - Format all date fields into one of the following formats accepted by REDCap for imports:
 - dd/mm/yyyy
 - yyyy-mm-dd
 - Save the script file as well as the data file in the chosen application's native format.
- Export the data file to .csv format and save to computer/network storage.

Data Mapping between the SPARCO site data dictionary and the standard SaDaCC dictionary

Use the provided Excel template for mapping (currently Saddac_sparco_tz_mapping_ver7.xls)

The sheet consists of 2 sheets **standard_sadacc_dictionary** and **site_vs_sadacc_mapping**. The former contains the the new standard SADaCC dictionary, do not change any of this. in the **site_vs_sadacc_mapping** is where you need to link SPARCO site variables to the **standard variables from SADaCC**.

Copy the first 6 columns data from the site dictionary made in step xx above and paste them in the first column (A-F) corresponding to "Variable / Field Name, Form Name, Field Type, Field Label and Choices, Calculations, OR Slider Labels respectively into the mapping sheet (site_vs_sadacc_mapping)

Note values resulting in columns (G-K) corresponding to Variable / Field Name standard_Form Name standard_Field Type standard_Field Label standard_Choices and Calculations, OR Slider Labels from "standard_sadacc_dictionary"

Note in column L the and comparison of whether site_codes are similar to standard codes

Select columns G2-L2, double click on the black cross to copy all the formulas across your selected variables

Variable / Field Name										
A	B	C	D	E	F	G	H	I	J	K
Variable / Field Name	Form Name	Section H	Field Type	Field Label	Choices, Calculations, OR Slider Labels	Variable / Field Name	standard_Form Name	standard_Field Type	standard_Field Label	standard_Choices, Calculations, OR Slider Labels
record_id						record_id	sparco_ba_text		Research Subject ID <a	
date_today						date_today	sparco_ba_text		Visit Date (dd/mm/yyyy)	
registry_number						#N/A	#N/A	#N/A	#N/A	#N/A
scd_number						scd_number	sparco_ba_text		SCD Number	
medical_record_number						#N/A	#N/A	#N/A	#N/A	#N/A
hospital_name						#N/A	#N/A	#N/A	#N/A	#N/A
first_name						#N/A	#N/A	#N/A	#N/A	#N/A
middle_name						#N/A	#N/A	#N/A	#N/A	#N/A
last_name						#N/A	#N/A	#N/A	#N/A	#N/A
date_of_birth						date_of_birth	sparco_ba_text		Date of Birth (dd/mm/	

Note : where the variables match the name of variable and related information is displayed ; #N/A implies no match and 0 correspond to blank/empty value

Please match all the fields by manually finding the new variable in the standard SADDaCC dictionary that correspond to the variable name in your site data dictionary

Example “registry_number” in the example above correspond to “participant_number”

Paste the new variable name (participant_number) in the corresponding row under column G (standard_Variable/Field Name).

Repeat this until all the **#N/A variable names** have the right much in the standard dictionary.

In case no match is found please make notes in the “Notes” column N, example “No match in standard”, in this case consult the SADaCC/SPARCO database team to make sure these variables are included and then to receive the latest dictionary and the Excel mapping file (Saddac_sparco_tz_mapping_ver7.xls)

Mapping the codes for “radio” type of data

In the mapping sheet activate filter under the DATA tab in Excel

Once the filter list is active, click on the field standard_Field Type (column I)

Deselect all option and then select “radio”

To map options

Focus on the rows where site_codes_similar_to_standard (column L) are not similar (FALSE)

For these you must map according to the codebooks (site code vs standard code) or column K vs F

- Therefore recode categorical variables' response options to match the new data dictionary. Sometimes the variable/item names may match, but the response codes/labels are not aligned. This needs to be done with extreme caution, as mistakes will invalidate the data set. (The python script and a mapping file can help to simplify the process as shown below)

- - Consider this example for which there are two recoding strategies:

Old data set	
Item	Response options
sex	1, female 2, male

New data dictionary	
Item	Response options
sex	0, female 1, male

- Strategy 1 – recode 1 to 0, then recode 2 to 1.
 - Strategy 2 – recode 1 to 'f' and 2 to 'm', then recode 'f' to 0 and 'm' to 1 (preferred strategy)
- The codebook provided by REDCap can be useful for coding purposes, it is found at "Project Home" tab in REDCap

🏠 Project Home

☰ Project Setup

Quick Tasks

📖 Codebook

The Codebook is a human-readable, read-only version of the project's Data Dictionary and serves as a quick reference for viewing field attributes.

📄 Export data

Export your data from REDCap to open or view in Excel or various stats packages.

📊 Create a report

Build custom reports for quick views of your data, and export reports to Excel/CSV.

✅ Check data quality

Build or execute data quality rules to find discrepancies and errors in your project data.

- Use the coding in the new codebook to code/change your data

4	address	Street, City, State, ZIP	notes, Identifier														
5	telephone	Phone number <i>Include Area Code</i>	text (phone), Identifier														
6	email	E-mail	text (email), Identifier														
7	dob	Date of birth	text (date_ymd), Identifier														
8	age	Age (years)	calc Calculation: rounddown(datediff([dob], 'today', 'y'))														
9	ethnicity	Ethnicity	radio <table border="1"> <tr><td>0</td><td>Hispanic or Latino</td></tr> <tr><td>1</td><td>NOT Hispanic or Latino</td></tr> <tr><td>2</td><td>Unknown / Not Reported</td></tr> </table> Custom alignment: LH	0	Hispanic or Latino	1	NOT Hispanic or Latino	2	Unknown / Not Reported								
0	Hispanic or Latino																
1	NOT Hispanic or Latino																
2	Unknown / Not Reported																
10	race	Race	dropdown <table border="1"> <tr><td>0</td><td>American Indian/Alaska Native</td></tr> <tr><td>1</td><td>Asian</td></tr> <tr><td>2</td><td>Native Hawaiian or Other Pacific Islander</td></tr> <tr><td>3</td><td>Black or African American</td></tr> <tr><td>4</td><td>White</td></tr> <tr><td>5</td><td>More Than One Race</td></tr> <tr><td>6</td><td>Unknown / Not Reported</td></tr> </table>	0	American Indian/Alaska Native	1	Asian	2	Native Hawaiian or Other Pacific Islander	3	Black or African American	4	White	5	More Than One Race	6	Unknown / Not Reported
0	American Indian/Alaska Native																
1	Asian																
2	Native Hawaiian or Other Pacific Islander																
3	Black or African American																
4	White																
5	More Than One Race																
6	Unknown / Not Reported																
11	sex	Sex	radio <table border="1"> <tr><td>0</td><td>Female</td></tr> <tr><td>1</td><td>Male</td></tr> </table>	0	Female	1	Male										
0	Female																
1	Male																

The syntax for mapping to use in Python Script

- Consider this example for which there are two recoding strategies:

Old Site data set	
Item	Response options
sex	1, female 2, male

New standard data dictionary	
Item	Response options
sex	0, female 1, male

- Strategy 1 – recode 1 to 0, then recode 2 to 1.
- Strategy 2 – recode 1 to 'f' and 2 to 'm', then recode 'f' to 0 and 'm' to 1

Now in the mapping Excel file go to column coding_map_for_script (column M)

Here you must use the following annotation strictly

For the sex example above the mapping is noted as **1:0,1:2**

There are two codes on both site dataset and standard data dictionary

code 1 will be changed to 0 and code 2 will be changed to 1

Note how the 2 codes in the annotation are separated using the symbols (: and ,)

Save your mapping

Separately save the `site_vs_sadacc_mapping` sheet as `site_vs_sadacc_mapping.csv`
File--> Save As[(CSV (comma delimited))]

This file `site_vs_sadacc_mapping.csv` will be used to make the new data migration template

Use Python Script to make the data migration template

The input to this file are

1. The mapping file `site_vs_sadacc_mapping`
2. The data file Exported from site's REDCap instance in step xxx.3 above.

Please provide these two files as input to the python script


The python script will

1. rename unmatched variables that do not match between the site data dictionary vs standard dictionary.
2. Recode the data elements of the type "radio" to match the current standard code
3. Delete local variables that do not find match to standard data dictionary.

The output of the script is the data migration template (csv) that can be used to migrate data to the new site instance of the standardized REDCap instance.


5. The data manager then imports the new data dictionary into REDCap, through the following steps:

- Inside the project, under the 'Project Setup' tab, click on the 'Data Dictionary' button:


In progress

Design your data collection instruments

Add or edit fields on your data collection instruments. This may be done by either using the Online Designer (online method) or by uploading a Data Dictionary (offline method). Quick links: [Download PDF of all instruments](#) OR [Download the current Data Dictionary](#)

Go to or 


Learn how to use [\[+\] Smart Variables](#) [Piping](#) [@ Action Tags](#)

- When prompted, click on ‘Enter Draft’ Mode’:

Project Home
Project Setup
Online Designer
Data Dictionary

NOTE: The project is currently in PRODUCTION status, and thus changes cannot be made in real time to the project as when in Development status. However, changes to the project may be drafted in DRAFT MODE, after which such changes will be reviewed and approved by a REDCap administrator. Once those changes are approved, you will then receive an email confirmation informing you that those changes have taken effect on your production project.

Would you like to enter DRAFT MODE to begin drafting changes to the project?



- Scroll down to the section ‘Upload your Data Dictionary file’ and select the new data dictionary before clicking on ‘Upload File’:

Upload your Data Dictionary file (CSV file format only)

Format for min/max validation values for date and datetime fields:

No file selected.

- After uploading the file, click on ‘Submit changes for review’ in order for changes to be effected in REDCap.
- If there are any data items that the site will not use at all (e.g. in the case of a test that is not available/being done at the site), these items can be hidden in REDCap. This is done from the Online Designer. Click on ‘edit’ item icon, then insert ‘@HIDDEN’ into the ‘Action Tags / Field Annotation’ space and click ‘Save’:

Edit Field ✕

You may add a new project field to this data collection instrument by completing the fields below and clicking the Save button at the bottom. When you add a new field, it will be added to the form on this page. For an overview of the different field types available, you may view the [Field Types video \(4 min\)](#).

Field Type: Yes - No ▼

Field Label
Hep A tested

Variable Name (utilized in logic, calcs, and exports)
hep_a_done Enable auto naming of variable based upon its Field Label?

ONLY letters, numbers, and underscores

How to use: [Smart Variables](#) [Piping](#)

Required?* No Yes
* Prompt if field is blank

Identifier? No Yes
Does the field contain identifying information (e.g., name, SSN, address)?

Custom Alignment Right / Horizontal (RH) ▼
Align the position of the field on the page

Action Tags / Field Annotation (optional)
@HIDDEN

Learn about [@ Action Tags](#) or [using Field Annotation](#)

Field Note (optional)
Small reminder text displayed underneath field

- 6. The data manager then imports the transformed data set back into REDCap, through the following steps:
 - Inside the project, under the 'Applications' section in the left-hand bar, click on 'Data Import Tool'.
 - Scroll down to select and upload the file:

Record format: The file to be uploaded has its records stored as separate Rows ▼

Format for date and datetime values: DD/MM/YYYY or YYYY-MM-DD ▼

Allow blank values to overwrite existing saved values? No, ignore blank values in the file (default) ▼

Upload your CSV file:

No file selected.

- Please note: Once the file has been uploaded, the data will NOT be immediately imported but will be displayed and checked for errors to ensure that all the data is in correct format before it is finally imported into the project by explicitly clicking 'Import Data' button.

Version No.	Date	Authorizer	External Reviewer	Internal Reviewer	Author	Details of changes
0.2	10 Oct 2018			Raphael Sangeda	Annemie Stewart	n/a first version

Layperson Ontology 1

SOP: Writing Layperson Annotations

PURPOSE

To describe the procedure for writing layperson annotations (labels and descriptions) for an ontology's terms.

SCOPE

This procedure is used by curators to produce layperson annotations for an ontology.

GLOSSARY/DEFINITIONS

SOP Standard Operating Procedure: Detailed written instructions to achieve uniformity of the performance of a specific function.

PROCEDURE

1. Prepare Working File:

- a. Extract term labels and descriptions from the ontology file by performing the procedure in **Ontology Development 1 SOP** and open the output file in Excel.
- b. Create either a blank Excel file or a Google Spreadsheet (if the process will involve people collaborating on the layperson annotations, a Google Spreadsheet is useful).
- c. Copy the following to the header row of the blank file:

current ID label	layperson label	current description	layperson description	hasExactSynonym	newExactSynonym
---------------------	--------------------	------------------------	--------------------------	-----------------	-----------------

- d. Copy from the file generated in *step 1*:
 - a) term IDs into the "ID" column of the blank file.
 - b) term labels into the "current label" column of the blank file.
 - c) term descriptions into the "current description" column of the blank file.
 - d) term exact synonyms into the "hasExactSynonym" column of the blank file.

2. Work on layperson annotations:

- a. Provide a layperson label for each term (row) in the "layperson label" column.
 - a) If the "current label"/column B is appropriate, copy it across to "layperson label"/column C.
 - b) If the current label is too technical, see if there's an appropriate synonym in the "hasExactSynonym" column.

- i. If there is one, copy it across to column C.
- ii. If there isn't one, provide an appropriate label in column C and copy it to the "newExactSynonym" column.

[Note: You can try search for the term in OLS as it sometimes provides layperson terms/synonyms in the "Term info" section for a term (see example below). Otherwise, a simple Google search can often provide potential options.]

EMBL-EBI Services

Ontology Lookup Service

Home **Ontologies** Documentation About

OLS > Human Phenotype Ontology **HP** > **HP:0000217**

Xerostomia

http://purl.obolibrary.org/obo/HP_0000217

Dryness of the mouth due to salivary gland dysfunction. [HPO:probinson]

Synonyms: Dry mouth syndrome, Dry mouth, Reduced salivation

Tree view **Term history**

Graph view
Reset tree
Show all siblings

Term info

database cross reference

- SNOMEDCT_US:56893005
- MSH:D014987
- SNOMEDCT_US:300268000
- SNOMEDCT_US:87715008
- UMLS:C0043352

Subsets

hposlim_core

layperson term
Dry mouth

layperson term
Reduced salivation

layperson term
Dry mouth syndrome

layperson term
Decreased salivary flow [ORCID:0000-0001-5889-4463]

has alternative id
HP:0002709, HP:0100756

has broad synonym
Decreased salivary flow

has obo namespace
human_phenotype

id
HP:0000217

- iii. You can add additional new exact synonyms to the "newExactSynonym" column, but ensure you separate terms in this column with a "|".
 - c) Ensure that labels are capitalised correctly according to the ontology's protocol.
- b. Provide a layperson description for each term (row) in the "layperson description" column.
- a) If the "current description"/column D is appropriate, copy it across to "layperson description"/column E.
 - b) If not, provide a layperson-appropriate description in column E.
[Note: A simple Google search is a good way to find simpler descriptions of terms.]
 - c) Ensure that descriptions are written according to the ontology's protocol (e.g. full sentences; not including the label, etc.)

USING LAYPERSON ANNOTATIONS

The layperson annotations can be incorporated into the ontology file by performing the procedure in either **Layperson Ontology 2.1 SOP** (if the ontology file does NOT include language tags for term annotations) or **Layperson Ontology 2.2 SOP** (if the ontology file DOES already include language tags for term annotations).

Version No.	Date	Authorizer	External Reviewer	Internal Reviewer	Author	Details of changes
					Jade Hotchkiss	n/a (first version)

Layperson Ontology 2.2

SOP: Incorporating Layperson Annotations

Note:

*This SOP is specifically for when the input ontology file DOES already have language tags for labels and descriptions of ontology terms. Use **Layperson Ontology 2.1 SOP** if the ontology input file does NOT already contain language tags for labels and descriptions of ontology terms.*

PURPOSE

To describe the procedure for incorporating layperson annotations (labels and descriptions) of ontology terms into the ontology file, using the *creating_layperson_ontology_files_existing_lang_tags.py* python program.

SCOPE

This procedure is used by curators when producing ontology files that include layperson annotations.

GLOSSARY/DEFINITIONS

- Python** The programming language in which the program used in this SOP was written.
- SOP** Standard Operating Procedure: Detailed written instructions to achieve uniformity of the performance of a specific function.

SYSTEM REQUIREMENTS TO RUN PROGRAM

Instructions below are for Windows 10.

For the python program to run successfully, the following need to be installed:

1. Python.
Ensure that Python is setup correctly on your PC (Use **Program Running 1 SOP**)

REQUIRED INPUT FOR PROGRAM

1. Input files:
 - a. A .txt file containing the layperson annotations of the ontology.
 - b. The owl file of the version of the ontology to which layperson annotations must be added (Note: for this SOP, this file will already contain language tags for labels and descriptions of ontology terms)
2. These additional variables are required in the python program code before it is run:
 - a. **layperson_annotations_file_cell1** The content of the first cell in the input file containing layperson annotations.

- b. **iri_path** The specific IRI generic path that is used for terms in the ontology being translated.
- c. **layperson_tag** The tag to be added to layperson annotations.
- d. **outputfile1_name** name of rdf file created with all ontology terms and layperson annotations included where necessary.
- e. **outputfile2_name** name of rdf file created with only terms with layperson annotations

PREPARING INPUT FOR PROGRAM

1. Preparing input files and/or specifying them with relevant variables in the *creating_layperson_ontology_files_existing_lang_tags.py* program code:
 - a. To prepare the .txt file with the layperson annotations of the ontology:
 - a) Open the shared Google spreadsheet containing layperson annotations (product of **Layperson Annotation 1 SOP**).
 - b) Download the file as a Microsoft Excel file (.xlsx).
 - c) Open the downloaded file and save it as a .txt (tab delimited) file.
 - d) Move the .txt file to the same location as the python program.
 - e) Ensure that the **layperson_annotations_file** variable in the python program code (row 121) points to the name (including file extension) of this text file.
 - b. Ensure that the **owl_file_path** variable in the python program code (row 122) points to the name (including file extension) of the file of the version of the ontology to which layperson annotations must be added.
2. Specifying additional required variables in the *creating_layperson_ontology_files_existing_lang_tags.py* program code:
 - a. For the **layperson_annotations_file_cell1** variable in row 125, supply the content of the first cell in the input file containing layperson annotations.
 - b. For the **iri_path** variable in row 126, supply the specific IRI generic path that is used for terms in the ontology being translated.
 - c. For the **layperson_tag** variable in row 127, supply the tag to be added to layperson annotations (e.g. "layperson").
 - d. For the **outputfile1_name** variable in row 128, supply the name of the rdf file that will be generated by the program and will contain all ontology terms, with layperson annotations included where necessary.
 - e. For the **outputfile2_name** variable in row 129, supply the name of the rdf file that will be generated by the program and will contain only terms with layperson annotations.

RUNNING THE PROGRAM

Run the *creating_layperson_ontology_files_existing_lang_tags.py* python program by following the procedure in the **Program Running 2 SOP**.

PROGRAM OUTPUT

Two files are generated:

1. An rdfxml file containing all terms in the ontology, including layperson annotations where necessary.
2. An rdfxml file containing only terms in the ontology that have layperson annotations.

USING THE OUTPUT FILES

The output files will appear in the same location as the python program file and can now be used as necessary, e.g. used to create a new project in Web Protégé or uploaded to a relevant repository.

Version No.	Date	Authorizer	External Reviewer	Internal Reviewer	Author	Details of changes
					Jade Hotchkiss	n/a (first version)

Ontology Development 1

SOP: Extracting Term Annotations from OWL File

PURPOSE

To describe the procedure for extracting annotations of all terms in an ontology's OWL file, using the python program "*create_terms_for_review_file.py*", into a format that can be read and shared in Excel format.

SCOPE

This procedure is used by curators when producing term annotations in a format that can be reviewed by reviewers, or for other processes involved in ontology development, such as ontology translation.

GLOSSARY/DEFINITIONS

Python	The programming language in which the program used in this SOP was written.
Protégé	An ontology development software program.
SOP	Standard Operating Procedure: Detailed written instructions to achieve uniformity of the performance of a specific function.

SYSTEM REQUIREMENTS TO RUN PROGRAM

Instructions below are for Windows 10.

For the python program to run successfully, the following need to be installed:

1. Python.
Ensure that Python is setup correctly on your PC (Use ***Program Running 1 SOP***)
2. Desktop Protégé.
If you do not already have desktop Protege installed, download and install it from: <https://protege.stanford.edu/products.php>

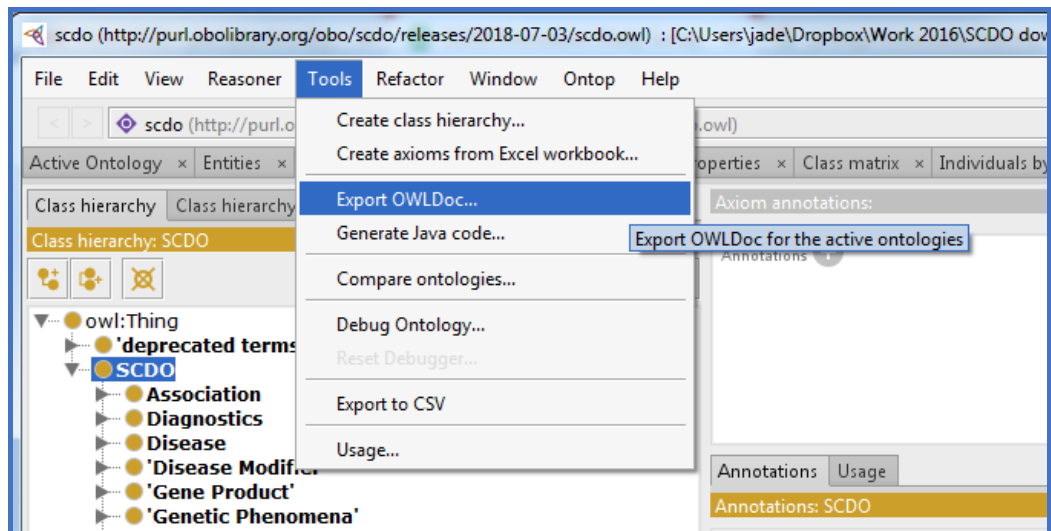
REQUIRED INPUT FOR PROGRAM

Before the program can be run, these required input files and folders need to be prepared and named in the program code:

1. Html files generated by Protégé
 - The locations of the "index.html" file in the "classes" folder and the location of this "classes" folder.
2. A file containing the list of terms to be extracted.
3. A file of important annotation properties, in the order of preference (used to output annotations in a certain order).

PREPARING INPUT FOR PROGRAM

1. To generate html files of an ontology:
 - a) Open the owl file in Protégé Desktop.
 - b) Export the entire ontology from Protégé as html by using the “Export OWL Doc” function from the Tools tab in Protégé [Note: export to a folder in the same location as the python program file.]



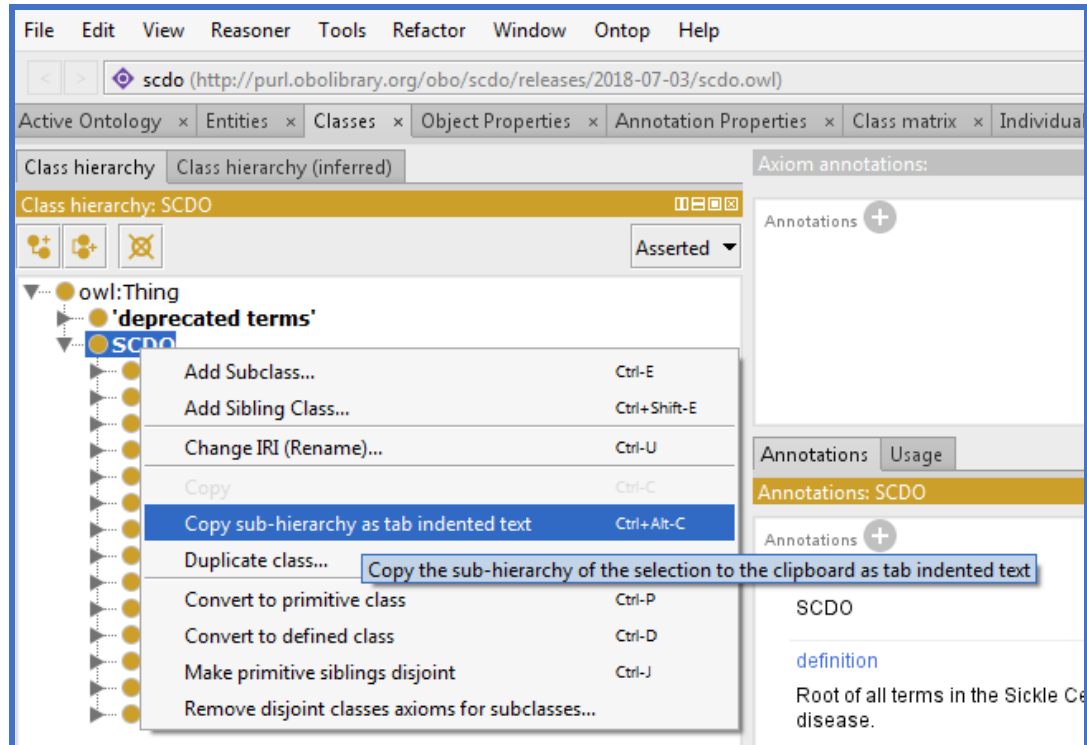
- c) Locate the “classes” folder that was exported and make sure the “classes_location” in the python program code (row 286) points to the location of the “classes” folder that contains the htmls of classes.

[Note: The location can be relative to the python program’s location, e.g. if the folder containing exported htmls from Protégé is called “Exported_from_Protege” and the “classes” folder is within it, then the location provided in row 286 should be “Exported_from_Protege\classes”.]

- d) Locate the “index.html” file (within the “classes” folder”) and make sure “input_file2” in the python program code (row 282) points to the location of the “index.html” file. [In keeping with the example already used in the previous step, the location would be “Exported_from_Protege\classes\index.html”.]

2. To prepare the file with the list of terms to be extracted:
 - a) Open the OWL file in Desktop Protégé

- b) Copy terms from classes to be reviewed (e.g. for all terms within the main "SCDO" term in the SCDO, right click on "SCDO" and select "Copy sub-hierarchy as tab indented text") and paste in a .txt file.

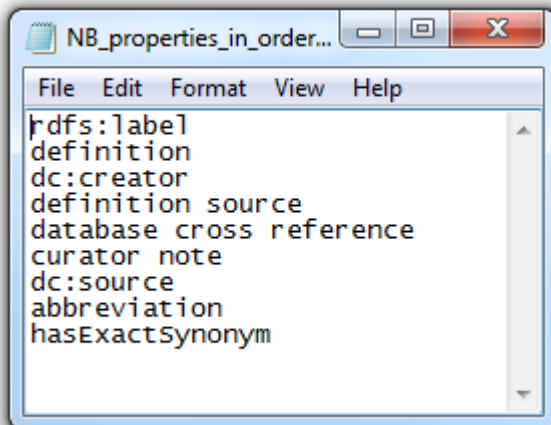


- c) Delete the empty row that automatically appears at the bottom of the copied terms.
d) Save the file in the same location as the python program file.
e) Make sure "input_file1" in the python program code (row 281) points to the name of this text file.

[The template file, "input_file_complete_ontology_template.txt" shows an example of what the content of the file could be.]

3. To create the input file that orders annotations as desired:

- a) Create a .txt file containing a custom list of important properties that are listed, one per row, according to the order in which they should appear in the header row of the exported terms. (This file can be re-used or edited as necessary for future runs of the program).



- b) Save the file in the same location as the python program file.
 - c) Make sure "input_file3" in the python program code (row 283) points to the name of this text file.
4. Use the "outputfile_name" in the python program code (row 287) to specify the name of the output file that will be generated by the program (Do not include a file extension; it will automatically be a .txt file).

RUNNING THE PROGRAM

Run the *create_terms_for_review_file.py* python program by following the procedure in the **Program Running 2 SOP**.

USING THE OUTPUT FILE

The output .txt file will appear in the same location as the python program file and can now be:

1. opened in Excel and saved in that format.
2. uploaded as a Google Spreadsheet which can be shared.
3. used in the process when the ontology is translated (In **Ontology Translation 1 SOP** and **Ontology Translation 5 SOPs**).

Version No.	Date	Authorizer	External Reviewer	Internal Reviewer	Author	Details of changes
					Jade Hotchkiss	n/a (first version)

Ontology Development 2

SOP: Extracting Property Annotations from OWL File

PURPOSE

To describe the procedure for extracting annotations of all properties in an ontology's OWL file into a format that can be read and shared in Excel format, using the python program "*extracting_property_annotations_from_ontology.py*".

SCOPE

This procedure is used by curators when producing property annotations in a format that can be reviewed by reviewers, or for other processes involved in ontology development, such as ontology translation.

GLOSSARY/DEFINITIONS

- SOP** Standard Operating Procedure: Detailed written instructions to achieve uniformity of the performance of a specific function.
- Python** The programming language in which the program used in this SOP was written.
- Protégé** An ontology development software program.
- WebProtege** A free, open source collaborative ontology development environment for the Web (by the same creators of Protégé).

SYSTEM REQUIREMENTS TO RUN PROGRAM

Instructions below are for Windows 10.

For the python program to run successfully, the following need to be installed:

1. Python.
Ensure that Python is setup correctly on your PC (Use **Program Running 1 SOP**)

REQUIRED INPUT FOR PROGRAM

1. Input file:
 - a. Untranslated ontology OWL file.
2. Additional variables:
 - a. The file name of the output file that will contain the annotations of the ontology's annotation properties.
 - b. The file name of the output file that will contain the annotations of the ontology's object properties.

PREPARING INPUT FOR PROGRAM

In the *extracting_property_annotations_from_ontology.py* python program code:

1. ensure that the full path of the untranslated ontology OWL file is correct for the **owl_file_path** variable in row 45.
*[Note: If this is the first time the ontology is being translated, the file must be the same version of the ontology used in **Ontology Development 1 SOP**]*
2. edit row 48 so that the **outputfile_name_1** variable points to the path (including file name and extension) of the output file for the annotation property annotations.
3. edit row 49 so that the **outputfile_name_2** variable points to the path (including file name and extension) of the output file for the object property annotations.

RUNNING THE PROGRAM

Run the *extracting_property_annotations_from_ontology.py* python program by following the procedure in the **Program Running 2 SOP**.

PROGRAM OUTPUT

The program outputs two .txt files:

1. A file containing IDs, labels and descriptions of annotation properties in the ontology.
2. A file containing IDs, labels and descriptions of object properties in the ontology.

USING THE OUTPUT FILES

The output .txt files will appear in the same location as the python program file (unless a different location was included in the filenames provided for them in the code).

1. Post-processing of files:
 - a. Before these files can be used, the "rdfs:"-type annotation properties, such as "rdfs:label", need to be added to these files manually (because of a fault in the owl2ready python module used in the code).
 - b. To do this, simply open the output file containing annotations of annotation properties and add a separate row for each "rdfs:"-type annotation property in the ontology, providing its label (without "rdfs:" component) in the first two columns and its description in the third column (view the ontology's annotation properties in WebProtege or Desktop Protégé to obtain the annotations).
 - c. Resave the file.

2. Files can now be:

- a. opened in Excel and saved in that format.
- b. uploaded as a Google Spreadsheet which can be shared.
- c. used in the process when the ontology is translated (**Ontology Translation 1 SOP** for first translation or **Ontology Translation 5 SOPs** for translation updates).

Version No.	Date	Authorizer	External Reviewer	Internal Reviewer	Author	Details of changes
					Jade Hotchkiss	n/a (first version)

Ontology Development 3

SOP: Producing Index File of Object Properties and their Web Published Association Files

PURPOSE

To publish association sheets (from the Google Spreadsheet used for recording associations between term in the ontology) to the web and to produce an index file that contains the object properties for which associations are published, each alongside the URL of its individually web published association sheet.

SCOPE

This procedure is used by curators to produce the index file that allows for easy download of the association files before they are included in the OWL file (**Ontology Development 4 SOP**).

GLOSSARY/DEFINITIONS

SOP Standard Operating Procedure: Detailed written instructions to achieve uniformity of the performance of a specific function.

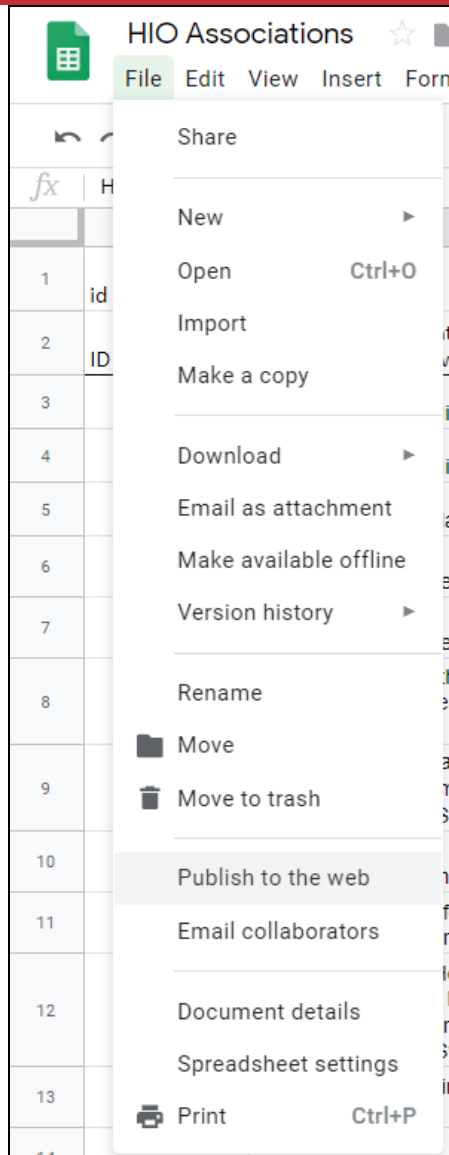
ASSUMPTIONS

This SOP assumes that:

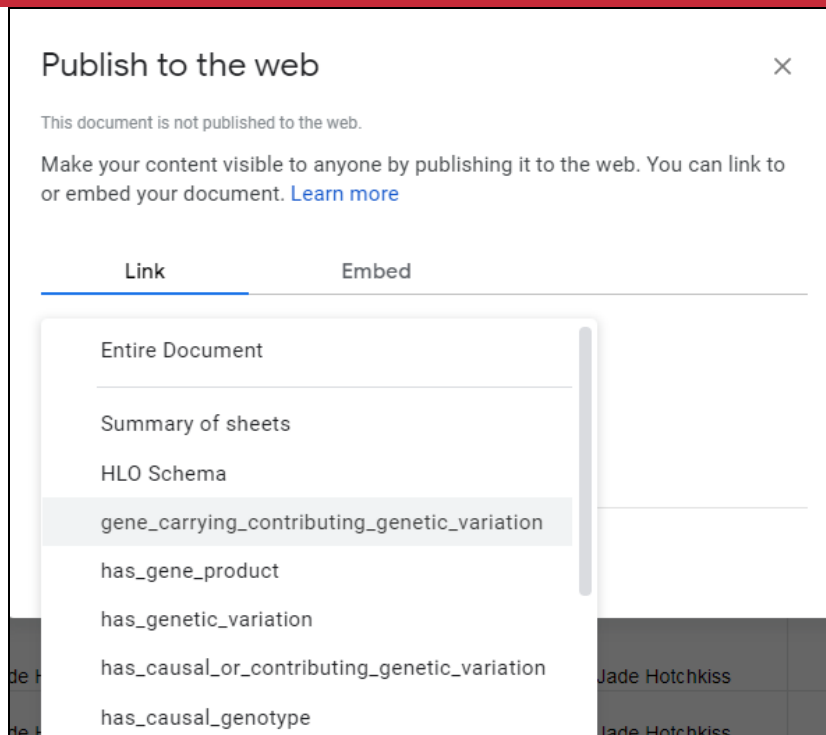
- the ontology of interest has used a Google Spreadsheet to record associations made between terms by object properties, with associations made by different object properties saved in separate sheets.
- the sheets are in the **same format as the template provided**.

PROCEDURE

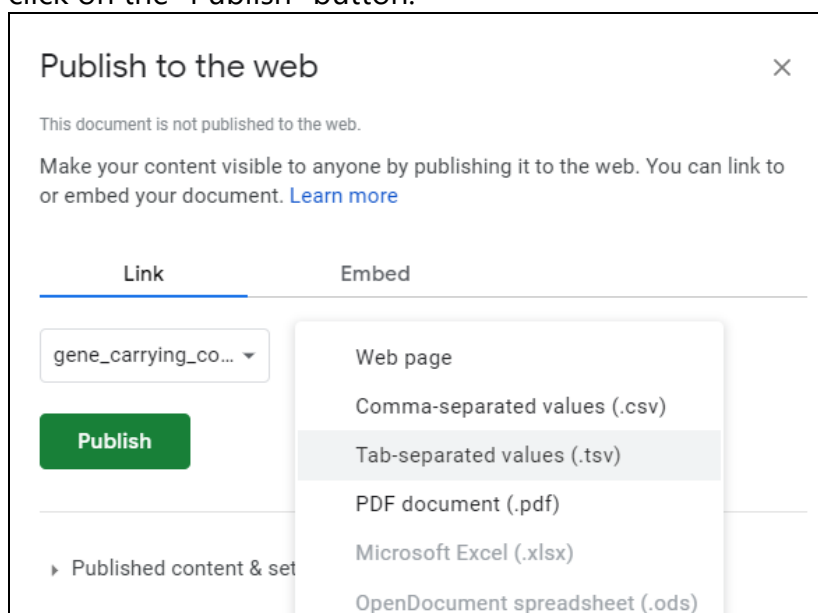
1. Create a new text file in which object property names will be saved alongside the URL of their web published spreadsheet.
2. Repeat these steps for each sheet in the Google Spreadsheet for associations:
 - a. Add the sheet's name (object property name) to the text file created in step 1.
 - b. Click on "Publish to the web" under the "File" tab of that object property's association sheet.



c. Select the relevant sheet below "Link".



- d. Select "Tab-separated values (.tsv)" from the "Embed" drop-down and click on the "Publish" button.



- e. Copy the link that is created to the text file and paste it alongside the sheet's name (added to the file in step 2(a)).

Publish to the web ×




This document is published to the web.

Make your content visible to anyone by publishing it to the web. You can link to or embed your document. [Learn more](#)

Link Embed

gene_carrying_co... Tab-separated values (.tsv)

<https://docs.google.com/spreadsheets/d/e/2PACX-1vQYUAvV31BlaBMPD7kk>

Or share this link using:   

Note: Viewers may be able to access the underlying data for published charts. [Learn more](#)

Published

▶ Published content & settings

- f. Ensure that the 'Automatically republish to the web when changes are made' option is selected (expand the "Publish content & settings" section to see this).

Publish to the web ✕

This document is published to the web.

Make your content visible to anyone by publishing it to the web. You can link to or embed your document. [Learn more](#)

Link
Embed

gene_carrying_co... ▾

Tab-separated values (.tsv) ▾

https://docs.google.com/spreadsheets/d/e/2PACX-1vQYUAvV31BlaBMPD7k

Or share this link using:

Note: Viewers may be able to access the underlying data for published charts. [Learn more](#)

Published

▾ Published content & settings

gene_carrying_co... ▾

Stop publishing

Automatically republish when changes are made

USING THE FILE PRODUCED

The index file produced will be used as input in the **Ontology Development 4 SOP**.

Version No.	Date	Authorizer	External Reviewer	Internal Reviewer	Author	Details of changes
					Jade Hotchkiss	n/a (first version)

Ontology Development 4

SOP: Compiling OWL File

PURPOSE

To describe the procedure for compiling an OWL file using ROBOT. This compiling process involves the inclusion of associations made between ontology terms/classes via object properties (recorded in separate excel sheets) into the ontology OWL file.

SCOPE

This procedure is used by curators when producing a version of the ontology to be published online (updated in GitHub, BioPortal, dedicated ontology website, etc).

GLOSSARY/DEFINITIONS

Python	The programming language in which the program used in this SOP was written.
Protégé	An ontology development software program.
SOP	Standard Operating Procedure: Detailed written instructions to achieve uniformity of the performance of a specific function.

ASSUMPTIONS

1. Step 1 of the procedure in this SOP assumes that the user already has a GitHub account and therefore has a GitHub username.
2. The instructions in this SOP assume that a GitHub repository (with ROBOT setup on it) exists for the ontology for which the OWL file is being compiled.
3. It is assumed that the person performing the procedure in this SOP has permission/access to edit the remote GitHub repository (if not, the creator of the remote repository must be asked to give this person .access, by giving the person's GitHub username relevant rights)

SYSTEM REQUIREMENTS TO CONDUCT SOP PROCESS

Instructions below are for Windows 10.

1. Git.
Ensure that Git is installed on your PC (Use **Program Running 3 SOP**).
2. Python.
Ensure that Python is setup correctly on your PC (Use **Program Running 1 SOP**).

PROCEDURE

1. **Configuring Git and connecting to a remote repository**

- a) Create a directory/folder to which you will clone the remote GitHub repository.
- b) Follow instructions 1 – 7 under the *Configuring and connecting to a remote repository* section on this web page: <https://www.computerhope.com/issues/ch001927.htm> (Note: step 6 is unnecessary as its content is repeated in 7).
- c) In order to push changes to the remote GitHub repository (done later in this SOP), the password for your GitHub username must be set in Windows.
 - a. Type “credential manager” in the Windows search bar and select *Credential Manager* from the search results.
 - b. Select “Windows Credentials”.
 - c. Click on “Add a Windows credential”.
 - d. Type the url of the GitHub repository you will be accessing in the “Internet or network address” field (e.g. for the SCDO it would be “github.com/scdodev/scdo-ontology”).
 - e. Type your GitHub username that has access to edit the repository into the “User name” field.
 - f. Type your corresponding GitHub password into the “Password” field.
 - g. Click on “OK”.

2. Preparing association files

Follow instructions in the ***Ontology Development 5 SOP*** to download and reformat the association files, using the *get_association_files.py* python program.

3. Checking association files for errors

Follow instructions in the ***Ontology Development 6 SOP*** to check the association files for inconsistencies, using the *checking_association_files_for_errors.py* python program.

4. Preparing Local Git Repository

USING THE OUTPUT FILES

Version No.	Date	Authorizer	External Reviewer	Internal Reviewer	Author	Details of changes
					Jade Hotchkiss	n/a (first version)

Ontology Development 5

SOP: Downloading Association Files

PURPOSE

To describe the procedure for downloading association files, using the *get_association_files.py* python program.

SCOPE

This procedure is used by curators to download the association files of an ontology for use in the process of compiling the complete ontology OWL file.

GLOSSARY/DEFINITIONS

- Python** The programming language in which the program used in this SOP was written.
- SOP** Standard Operating Procedure: Detailed written instructions to achieve uniformity of the performance of a specific function.

SYSTEM REQUIREMENTS TO RUN PROGRAM

Instructions below are for Windows 10.

For the python program to run successfully, the following needs to be installed:

1. Python.
Ensure that Python is setup correctly on your PC (Use **Program Running 1 SOP**)

ASSUMPTIONS

1. To run the python program in this SOP, it is assumed that:
 - a. the index file for ontology association files has been created (using the **Ontology Development 3 SOP**).
 - b. there is internet access when running the python program.

REQUIRED INPUT FOR PROGRAM

1. Input file:
A .txt file containing an index with the object property names vs the name of the file containing associations made by that object property.
2. Additional variables:
 - a. The location of the folder to which the downloaded files will be written.
 - b. The location of the folder to which the reformatted files will be written.

PREPARING INPUT FOR PROGRAM

1. In the *get_association_files.py* python program code:
 - ensure that the full path of the index file for ontology association files is correct for the **input_file1** variable in row 54.
2. Ensure that these folders exist in the same location as the python program (create them if necessary):
 - a. output_files
 - b. reformatted_files
3. If the folders in step 2 already contain files, delete these files before continuing.

RUNNING THE PROGRAM

Run the *get_association_files.py* python program by following the procedure in the **Program Running 2 SOP**.

PROGRAM OUTPUT

The program outputs:

1. association files (in .txt format) downloaded from the web in the *output_files* folder.
2. reformatted association files (in .txt format) in the *reformatted_files* folder.

USING THE OUTPUT FILES

The reformatted files (in the *reformatted_files* folder) will be used as input in **Ontology Development 4 SOP** (in which the OWL file is compiled using ROBOT).

Version No.	Date	Authorizer	External Reviewer	Internal Reviewer	Author	Details of changes
					Jade Hotchkiss	n/a (first version)

Ontology Development 6

SOP: Check Association Files for Inconsistencies

PURPOSE

To describe the procedure for checking reformatted association files for inconsistencies between their content and the content of the latest owl file, using the *checking_association_files_for_errors.py* python program.

SCOPE

This procedure is used by curators before association files are used in the process of compiling the complete ontology OWL file.

GLOSSARY/DEFINITIONS

- Python** The programming language in which the program used in this SOP was written.
- SOP** Standard Operating Procedure: Detailed written instructions to achieve uniformity of the performance of a specific function.

SYSTEM REQUIREMENTS TO RUN PROGRAM

Instructions below are for Windows 10.

For the python program to run successfully, the following needs to be installed:

1. Python.
Ensure that Python is setup correctly on your PC (Use **Program Running 1 SOP**)

REQUIRED INPUT FOR PROGRAM

1. Input file:
The latest output file of the *create_terms_for_review_file.py* program (Use **Ontology Development 1 SOP** to obtain this if you do not already have it at hand).
2. Additional variables:
The location of the folder containing downloaded association files (not the folder containing reformatted files).

PREPARING INPUT FOR PROGRAM

In the *checking_association_files_for_errors.py* python program code:

1. ensure that the path of the latest .txt file output by the *create_terms_for_review_file.py* program (see **Ontology Development 1 SOP**) is correct for the **input_file1** variable in row 72.

- edit row 73 so that the **association_files_location** variable points to the name of the output folder containing the latest downloaded association files.

RUNNING THE PROGRAM

Run the *checking_association_files_for_errors.py* python program by following the procedure in the **Program Running 2 SOP**.

PROGRAM OUTPUT

The program outputs on the screen there are inconsistencies in the association files:

- any term IDs or labels that are present in the association files but not in the curator's version of the ontology
- cases where an ID is associated with more than one label.

USING THE OUTPUT

Inconsistencies found in the association files are likely to point to human errors made when recording associations or when updating terms in the ontology's OWL file.

Depending on the inconsistencies identified, the curator might need to make updates to the Google Spreadsheet of Associations and/or to the ontology OWL file. If such updates are made, the association files will need to be redownloaded (rerun **Ontology Development 1 SOP** and **Ontology Development 5 SOP**).

Version No.	Date	Authorizer	External Reviewer	Internal Reviewer	Author	Details of changes
					Jade Hotchkiss	n/a (first version)

Ontology Translation 1

SOP: Preparing Ontology Content for First Round of Ontology Translation

PURPOSE

To describe the procedure for preparing ontology content in appropriate formats for the first time that an ontology is to be translated into a certain language.

SCOPE

This procedure is used by curators to prepare ontology content in appropriate formats to be reviewed by translators for the first time that it is translated into a certain language.

GLOSSARY/DEFINITIONS

- SOP** Standard Operating Procedure: Detailed written instructions to achieve uniformity of the performance of a specific function.
- Babylon** Translation software used to produce automatic translations.

PROCEDURE – TERM ANNOTATIONS

This procedure is for creating the first file of ontology term auto-translations to be reviewed by ontology translators/reviewers.

1. Run the *create_terms_for_review_file.py* program (as per the **Ontology Development 1 SOP**).
2. Copy the output file and rename it as "**ONTABRV_LANGUAGE**_translations_for_review_**DATE**", replacing:
 - a. **ONTABRV** with the relevant ontology abbreviation (e.g. SCDO)
 - b. **LANGUAGE** with the relevant language being translated into (e.g. French)
 - c. **DATE** with the current date (e.g. 11July2019)

For example: SCDO_French_translations_for_review_15Feb2020

3. Open the file in Excel and resave it as an .xlsx file.
4. Rename "Sheet 1" to "Review this".
5. Delete the "parent class" column (column A).
6. Create a text file, named "annotation_properties_not_translated" that lists which columns (their headers) will not need to be translated.

Example of the list used for the SCDO:

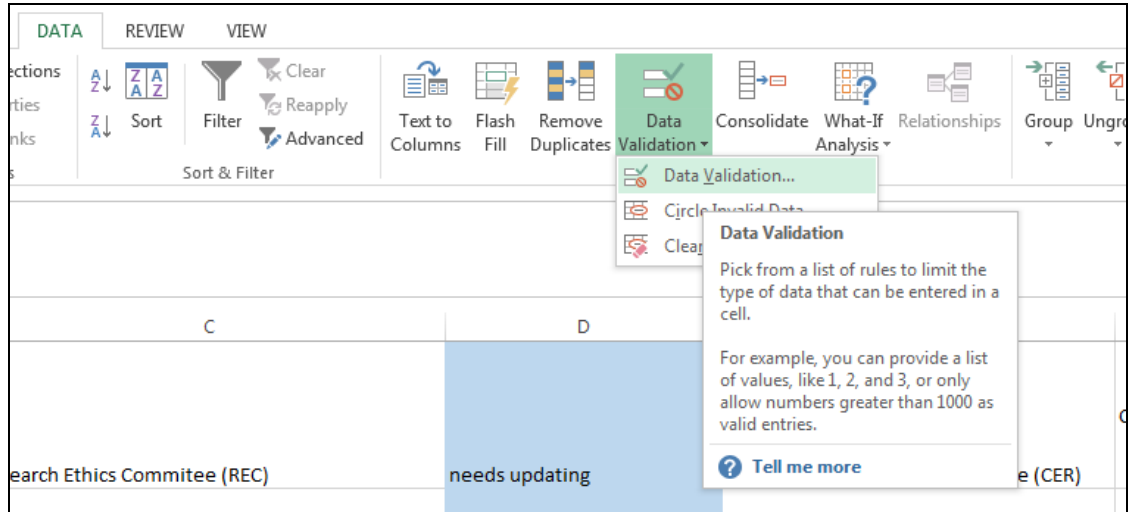
SCDO_ID
dc:creator
definition source
database cross reference
curator note
dc:source
creation_date
rdfs:seeAlso
created_by)

7. Delete all columns of annotation properties included in the above list, except for the first ID column and save the file.
8. Use Babylon to translate the Excel file into the necessary language. A translated Excel file will be generated.
9. Save the generated Excel file appropriately (e.g. "SCDO_French_Terms_Auto_Translation_4June2019").
10. Merge English and auto-translations of terms into one file for the translation reviewers:
 - a. In the "**ONTABRV_LANGUAGE_translations_for_review_DATE**" Excel file, add three empty columns to the right of each column, except for the first ID column.
 - b. Copy the translated columns (from the file output by Babylon) and paste their content in the empty column directly next to the relevant column in the "**ONTABRV_LANGUAGE_translations_for_review_DATE**" file.
 - c. Label the columns from step (b) to indicate that they contain auto-translations for specific annotation properties, e.g. the column with auto-translations for the rdfs:label annotation property would be labelled "rdfs:label|auto-translation".
 - d. Label the next empty columns to indicate that they contain the reviewed and validated translations, e.g. for rdfs:label, this column would be labelled "rdfs:label|reviewed translation".
 - e. Label the last empty column of each annotation property as "validation provided for translation".
11. Add data validation to all "validation provided for translation" columns:
 - a. Add a new sheet, name it "For DV – Ignore this", and copy this content to it:

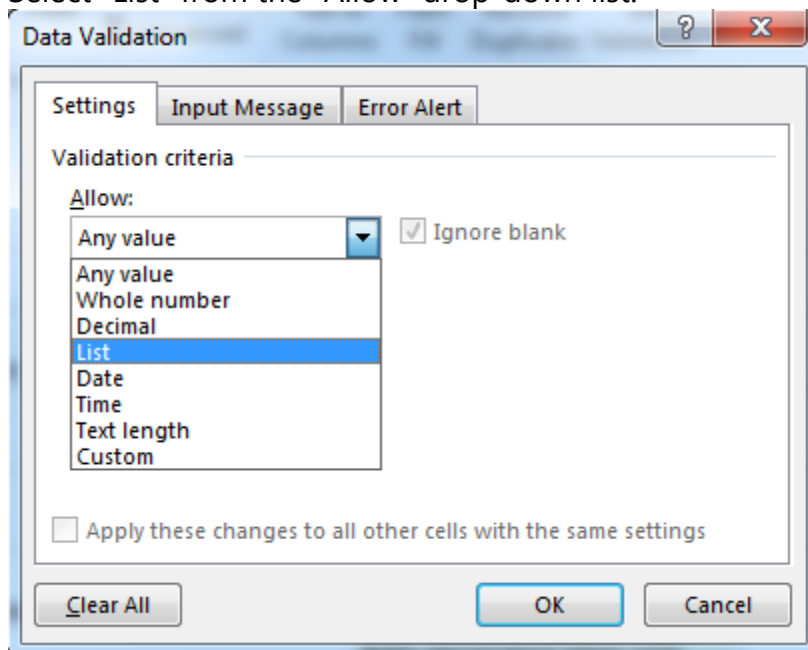
Data validation
checked, edited and validated
checked and validated without any edits
edited but needs re-checking

NA

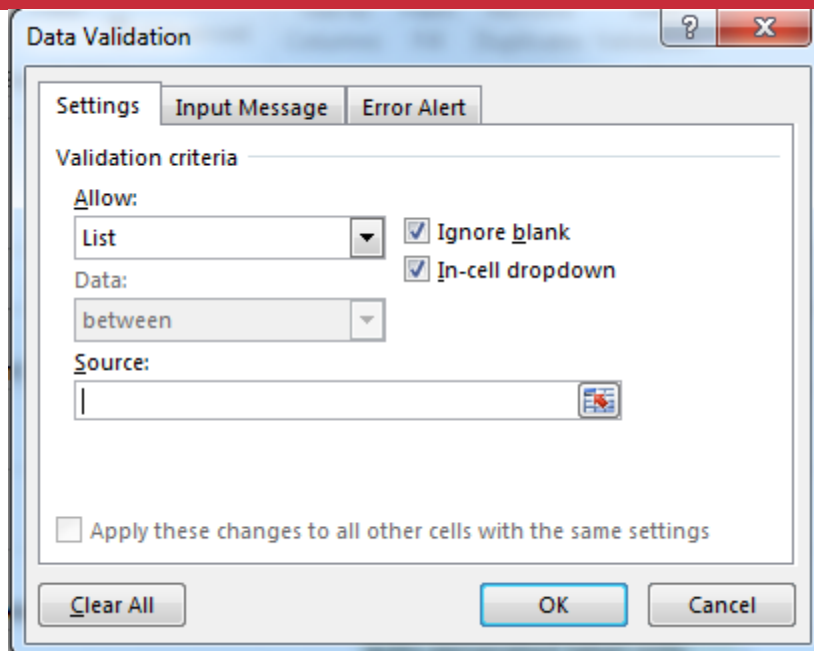
- b. Copy the text "Indicate translation work done here" into all cells (except header cell) of all "validation provided for translation" columns.
- c. Highlight all the columns populated in the last step and open the Data Validation window:



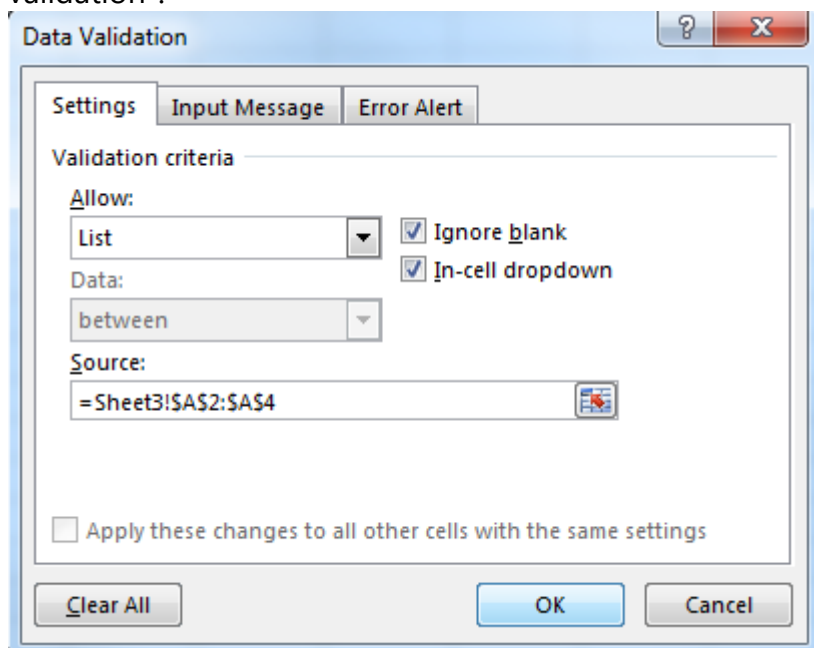
- d. Select "List" from the "Allow" drop-down list:



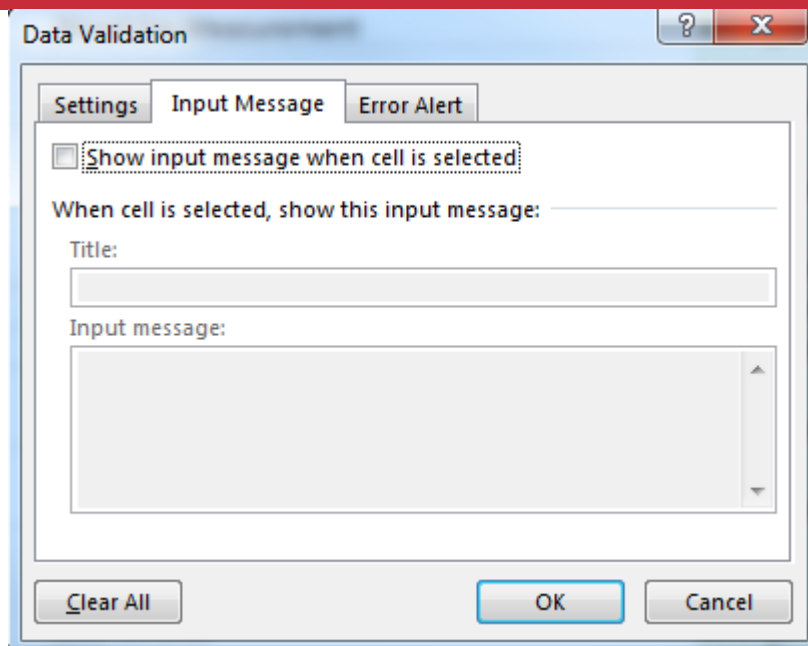
- e. You will need to select a source for the list:



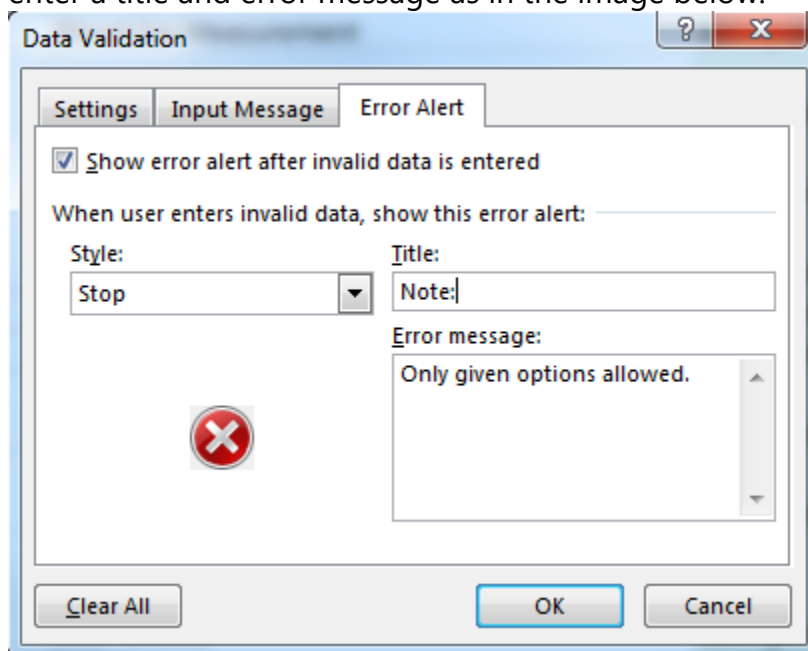
- f. From the additional sheet you added, select the cells below “Data validation”:



- g. In the “Input Message” tab, deselect “Show input message when cell is selected”:



- h. In the "Error Alert" tab, select "Stop" from the "Style" drop-down and enter a title and error message as in the image below:



- i. Click "OK" in the "Data Validation" window. The cells with data validation will now have a drop-down list with the provided data validation options.
- j. Remove data validation from headers by highlighting the top row, going to data validation and clicking "Clear All" and "OK".

12. Perform final edits to file:

- a. Embolden the header row content.

- b. Add a bold border to the header row.
- c. From the "View" tab, freeze the top row.
- d. Add a new "Review stats" sheet.
- e. Copy this table to the "Review stats" sheet:

Number of label translation checks:	
Number of description translation checks:	
Labels:	
checked, edited and validated	
checked and validated without any edits	
edited but needs re-checking	
NA	
Definitions:	
checked, edited and validated	
checked and validated without any edits	
edited but needs re-checking	
NA	

- f. Sort the "Review this" sheet by the "rdfs:label" column and see how many labels are in the file. Add this number to the top right of this table.
- g. Sort the "Review this" sheet by the "definition" column and see how many descriptions need translation reviews. Add this number to the next cell down of this table.

The remainder of the table in the "Review stats" sheet will be completed after review is complete.

PROCEDURE – PROPERTY ANNOTATIONS

This procedure is for creating the first files containing auto-translations of annotation properties and object properties of an ontology, to be reviewed by ontology translators/reviewers.

1. Run the program *extracting_property_annotations_for_translation.py* (as per the **Ontology Development 2 SOP**). This produces two .txt. files, one with labels and descriptions of the ontology's annotation properties and the other with labels and descriptions of the ontology's object properties.
2. Open the output files in Excel and resave them as .xlsx files (named appropriately, e.g.

- "SCDO_ann_property_annotations_for_auto_translation_input_4June2019" and "SCDO_obj_property_annotations_for_auto_translation_input_4June2019".)
3. Use Babylon to translate the Excel files, one at a time, into the necessary language and save the generated Excel files (named appropriately, e.g. "SCDO_French_Ann_Property_Auto_Translation_4June2019" and "SCDO_French_Obj_Property_Auto_Translation_4June2019").
 4. Obtain copies of these Excel template files:
 - a. ann_property_translations_for_reviewers_template.xlsm
 - b. obj_property_translations_for_reviewers_template.xlsm
 5. Open and resave these template files appropriately (e.g. "SCDO_French_ann_property_translations_for_reviewers_4June2019" and "SCDO_French_obj_property_translations_for_reviewers_4June2019")
 6. Copy content to the template files:
 - a. Copy the term IDs and English labels (first 2 columns) from the file with annotation properties (ann_property) (step 2) into the first two columns of the template for annotation properties (obtained in step 5). Do likewise for object properties (obj_property).
 - b. Copy the label translations (column B) from the translated annotation properties file (step 3) to column C of the template for annotation properties. Do likewise for object properties (obj_property).
 - c. Copy the English descriptions (column C) from the file with annotation properties (ann_property) (step 2) into column F of the template for annotation properties. Do likewise for object properties (obj_property).
 - d. Copy the description translations (column C) from the translated annotation properties file (step 3) to column G of the template for annotation properties. Do likewise for object properties (obj_property).
 7. Save the file.

PROCEDURE – SENDING FILES TO REVIEWERS/TRANSLATORS

Email the 3 Excel files generated in this SOP along with the **Ontology Translation SOP 7.1 SOP** to the reviewers/translators.

Version No.	Date	Authorizer	External Reviewer	Internal Reviewer	Author	Details of changes	of
					Jade Hotchkiss	n/a	(first version)

Ontology Translation 2

SOP: Adding Language Tags to OWL File Before Adding Translations

PURPOSE

To describe the procedure for adding language tags of the language of development (for the SCDO it was English) to an OWL file's term and property annotations, using the *adding_lang_tags.py* python program.

SCOPE

This procedure is used by curators to add language tags (i.e. for English, "en" tags are added) to an OWL file's term and property annotations before translations of annotations into a different language are included into the OWL file.

GLOSSARY/DEFINITIONS

- SOP** Standard Operating Procedure: Detailed written instructions to achieve uniformity of the performance of a specific function.
- OWL** The Web Ontology Language (OWL) is a family of knowledge representation languages for authoring ontologies.
- Python** The programming language in which the program used in this SOP was written.
- Protégé** An ontology development software program.
- WebProtege** A free, open source collaborative ontology development environment for the Web (by the same creators of Protégé).

SYSTEM REQUIREMENTS TO RUN PROGRAM

Instructions below are for Windows 10.

For the python program to run successfully, the following need to be installed:

1. Python.
Ensure that Python is setup correctly on your PC (Use **Program Running 1 SOP**)
2. Desktop Protégé.
If you do not already have desktop Protégé installed, download and install it from: <https://protege.stanford.edu/products.php>

REQUIRED INPUT FOR PROGRAM

1. Input files:
 - a. Untranslated ontology OWL file.
 - b. A file containing the list of annotation properties that will have language tags added.
2. Additional variables:

- a. **outputfile_name_1** The path of the file to be produced by the program.
- b. **lang_tag** The language tag (e.g. "en") that will be added to the OWL file.

PREPARING INPUT FOR PROGRAM

1. Preparing input files and specifying them with relevant variables:

- a. Ensure that the path of the untranslated ontology OWL file is correct for **owl_file_path** variable in row 83 of the *adding_lang_tags.py* program code.

[This must be the same version of the ontology used in **Ontology Translation 1 SOP** to generate the files for translation by reviewers.]

- b. To prepare the file with the list of annotation properties that are to have language tags added:

[Note:

- The file must contain 2 columns; column A must contain the annotation property's ID and column B must contain its label.
- "Built-in" properties do not have an ID related to the ontology, but instead their label should be used in column A as well. Examples are "hasExactSynonym", "label", "description".]

- a) Copy and rename the output file produced by the **Ontology Development 2 SOP** that contains annotation property annotations (before auto-translations are done).
- b) Delete all rows of annotation properties that will not require translation (e.g. dc:source annotations should not require translation) and thus don't require language tags to be added to them.
- c) Delete columns B and C.
- d) Save the file.
- e) Ensure that the full path of this file (including file extension) is correct for the **input_file2_path** variable in row 84 of the *adding_lang_tags.py* program code.

2. Specifying additional required variables in the *adding_lang_tags.py* program code:

- a. For the **outputfile_name_1** variable in row 87, supply the desired path (including filename) of the OWL file to be output by the program.
- b. For the **lang_tag** variable in row 88, supply the desired language tag to be added to the OWL file (e.g. "en" tag used for English, but could be different if the original OWL file is in a different language, e.g."fr" for French).

RUNNING THE PROGRAM

Run the *adding_lang_tags.py* python program by following the procedure in the **Program Running 2 SOP**.

EDITING OUTPUT

Because of a bug in the owlready2 python module used in the code, the *rdfs:label* and *rdfs:comment* annotation properties must have language tags added manually in the output OWL file. This can be done either on Desktop Protégé or WebProtege, by opening the ontology file created by the *adding_lang_tags.py* program in Protégé or creating a new project in WebProtege using the file, adding language tags manually to the label and description annotations and resaving the ontology.

USING THE OUTPUT FILE

The output file will be used as input for the program that adds translations to the owl file (**Ontology Translation 4 SOP**).

Version No.	Date	Authorizer	External Reviewer	Internal Reviewer	Author	Details of changes
					Jade Hotchkiss	n/a (first version)

Ontology Translation 3.1

SOP: Extracting Reviewed Translations – First Time

PURPOSE

To describe the procedure for extracting translated term annotations from the first file of terms reviewed by translators and producing a separate file for each annotation property (containing all translated annotations attributed to it), using the *extracting_annotation_translations_before_inclusion_general.py* python program. It is one of the steps performed before term annotation translations are added to the OWL file for the first time.

SCOPE

This procedure is used by curators after receiving reviewed translations from translators/reviewers for the first time in preparation for adding the translated annotations to the ontology's OWL file.

GLOSSARY/DEFINITIONS

- SOP** Standard Operating Procedure: Detailed written instructions to achieve uniformity of the performance of a specific function.
- Python** The programming language in which the program used in this SOP was written.

SYSTEM REQUIREMENTS TO RUN PROGRAM

Instructions below are for Windows 10.

Ensure that Python is setup correctly on your PC (Use **Program Running 1 SOP**).

REQUIRED INPUT FOR PROGRAM

1. Input files:
 - a. A .txt file version of the reviewed term translations received back from reviewers/translators.
 - b. A .txt file with the list of relevant annotation properties to be translated (same file used as input in **Ontology Translation 2 SOP**).
2. Additional variables:
 - a. **input_file_first_cell** The content of the first cell in the input file's header row.
 - b. **index_file_name** The name of the index file to be generated containing annotation property IDs and corresponding names of files containing their translated annotations.
 - c. **output_folder_name** The name of the folder to which the output files will be added.

PREPARING INPUT FOR PROGRAM

1. Preparing input files and specifying them with relevant variables:
 - a. Open the Excel file of reviewed term annotation translations received back from reviewers/translators and save it as a "Text (tab delimited) (*.txt)" file.
 - a) Ensure that the path of this file (including file extension) is correct for the **input_file1** variable in row **88** of the *extracting_annotation_translations_before_inclusion_general.py* program code.
 - b. Ensure that the path of the .txt file with the list of relevant annotation properties to be translated is correct for the **input_file2** variable in row **89** of the *extracting_annotation_translations_before_inclusion_general.py* program code.
 2. Specifying additional required variables in the *extracting_annotation_translations_before_inclusion_general.py* program code:
 - a. For the **input_file_first_cell** variable in row **92**, supply the content of the first cell of the header row of the first input file (specified by the **input_file1** variable).
 - b. For the **index_file_name** variable in row **93**, supply the name of the index file to be generated (don't add file extension to name).
 3. Create a folder to which the output files should be added. This folder should be in the same location as the *extracting_annotation_translations_before_inclusion_general.py* python program.
 - a. Ensure that the name of this folder is correct for the **output_folder_name** variable in row **94** of the *extracting_annotation_translations_before_inclusion_general.py* program code.

RUNNING THE PROGRAM

Run the *extracting_annotation_translations_before_inclusion_general.py* python program by following the procedure in the **Program Running 2 SOP**.

PROGRAM OUTPUT

The output folder specified will be populated with:

1. An index file (column A: annotation property ID; column B: name of corresponding file containing translated annotations for the property).
2. a .txt files for each annotation property listed in the index file:
E.g. **extracted_translated_XXX** file, where XXX is the name of the property

Each file contains:

Header row: "ID", "annotation"

column A: IDs of terms in ontology (1 per row)

column B: translated annotation, if available

USING THE OUTPUT FILES

The output files will be used as input for the program that adds translations to the owl file (***Ontology Translation 4 SOP***).

Version No.	Date	Authorizer	External Reviewer	Internal Reviewer	Author	Details of changes	of changes
					Jade Hotchkiss	n/a	(first version)

Ontology Translation 3.2

SOP: Extracting Reviewed Translations – Updated Labels and Descriptions

PURPOSE

To describe the procedure for extracting translated term annotations from a file of reviewed translation updates and producing a separate file for label and definition translations using the *extracting_updated_annotation_translations_before_inclusion_LandOnly.py* python program. It is one of the steps performed before updates to term annotation translations are added to the OWL file.

SCOPE

This procedure is used by curators after receiving reviewed updated translations of labels and definitions from translators/reviewers in preparation for adding the updated annotation translations to the ontology's OWL file.

GLOSSARY/DEFINITIONS

- SOP** Standard Operating Procedure: Detailed written instructions to achieve uniformity of the performance of a specific function.
- Python** The programming language in which the program used in this SOP was written.

SYSTEM REQUIREMENTS TO RUN PROGRAM

Instructions below are for Windows 10.

Ensure that Python is setup correctly on your PC (Use **Program Running 1 SOP**).

REQUIRED INPUT FOR PROGRAM

1. Input files:
 - a. A .txt file version of the reviewed term translation updates received back from reviewers/translators.
2. Additional variables:
 - a. **input_file_first_cell** The content of the first cell in the input file's header row.
 - b. **output_folder_name** The name of the folder to which the output files will be added.

PREPARING INPUT FOR PROGRAM

1. Preparing input files and specifying them with relevant variables:

- a. Open the Excel file of reviewed term annotation translations received back from reviewers/translators and save it as a "Text (tab delimited) (*.txt)" file.
 - a) Ensure that the path of this file (including file extension) is correct for the **input_file1** variable in row **48** of the *extracting_updated_annotation_translations_before_inclusion_LandDonly.py* program code.

2. Specifying additional required variables in the *extracting_updated_annotation_translations_before_inclusion_LandDonly.py* program code:
 - a. For the **input_file_first_cell** variable in row **51**, supply the content of the first cell of the header row of the input file (specified by the **input_file1** variable).

3. Create a folder to which the output files should be added. This folder should be in the same location as the *extracting_updated_annotation_translations_before_inclusion_LandDonly.py* python program.
 - a. Ensure that the name of this folder is correct for the **output_folder_name** variable in row **52** of the *extracting_updated_annotation_translations_before_inclusion_LandDonly.py* program code.

RUNNING THE PROGRAM

Run the *extracting_updated_annotation_translations_before_inclusion_LandDonly.py* python program by following the procedure in the **Program Running 2 SOP**.

PROGRAM OUTPUT

The output folder specified will be populated one .txt file for each annotation property:

1. **extracted_translated_label.txt**, and
2. **extracted_translated_description.txt**

Both files contain:

- Header row: "ID", "annotation"
- column A: IDs of terms in ontology (1 per row)
- column B: translated annotation, if available

USING THE OUTPUT FILES

The output files will be used as input for the program that adds translations to the owl file (**Ontology Translation 4 SOP**).

Version No.	Date	Authorizer	External Reviewer	Internal Reviewer	Author	Details of changes
					Jade Hotchkiss	n/a (first)

						version)
--	--	--	--	--	--	----------

Ontology Translation 4

SOP: Adding Translations to OWL File

PURPOSE

To describe the procedure for adding reviewed translations of property annotations and term annotations to an OWL file, using the *adding_translations_to_owl_file.py* python program.

SCOPE

This procedure is used by curators after receiving reviewed translations from translators/reviewers in order to incorporate the translations into an ontology's OWL file.

GLOSSARY/DEFINITIONS

SOP Standard Operating Procedure: Detailed written instructions to achieve uniformity of the performance of a specific function.

Protégé An ontology development software program.

WebProtégé An open source, lightweight, web-based ontology editor (developed and maintained by the same team that develops and maintains Protégé).

SYSTEM REQUIREMENTS TO RUN PROGRAM

Instructions below are for Windows 10.

Ensure that Python is setup correctly on your PC (Use **Program Running 1 SOP**).

REQUIRED INPUT FOR PROGRAM

1. Input files:

- a. An rdf/xml file (not owl/xml file) of the version of the ontology that has language tags (e.g. "en") added.
 - a) If it is the first time translations are being included in the ontology, the output file from **Ontology Translation 2 SOP** is to be used.
 - b) If it is not the first time, the curator's version of the rdf/xml that was used to generate the translations for updating should be used.
- b. Files output from **Ontology Translation 3 SOPs**:
 - a) an index file of annotation properties and the locations of their corresponding file containing translated annotations (**Note: this file is not necessary if it is a translation update with only label and description translations**).
 - b) files containing reviewed translated annotations (1 file per translated annotation property).

- c. .txt file versions of the reviewed translations of:
 - a) object property annotations
 - b) annotation property annotations.
2. These additional variables are required in the python program code before it is run:
 - a. **extracted_reviewed_translations_folder_path** The path of the folder containing .txt files with reviewed translations of term annotations.
 - b. **iri_path** The specific IRI generic path that is used for terms in the ontology being translated.
 - c. **lang_tag** The language tag to be added to translated annotations.
 - d. **outputfile_name** The name of the output file to be generated by the program.

PREPARING INPUT FOR PROGRAM

1. Preparing input files and/or specifying them with relevant variables in the *adding_translations_to_owl_file.py* program code:
 - a. Ensure that the path of the .rdf file of the untranslated ontology is correct for the **owl_file_path** variable in row 144.
 - b. For the **index_file_for_annotation_property_vs_filename** variable in row 145:
 - a) if translations are only for labels and definitions of terms, ensure that the variable is empty (i.e. set to "").
 - b) if translations are for more than just labels and definitions of terms, ensure that the name of the index file produced in **Ontology Translation 3.1 (or 3.3 SOP)** if it is a subsequent translation) is correct for the variable (*Note: the file extension in the file name is required but its full path is not required*).
 - c. Using the files identified or generated from **Ontology Translation 8.1 SOP**:
 - a) ensure that the full path of the .txt file of reviewed translations of object property annotations is correct for the **obj_prop_translations_input_file** variable in row 146.
 - b) ensure that the full path of the .txt file of reviewed translations of object property annotations is correct for the **ann_prop_translations_input_file** variable in row 147.
2. Specifying additional required variables in the *adding_translations_to_owl_file.py* program code:

- a. For the **extracted_reviewed_translations_folder_path** variable in row **150**, supply the name of the folder containing files generated in **Ontology Translation 3.1 SOP**.
- b. For the **iri_path** variable in row **151**, supply the specific IRI generic path that is used for terms in the ontology being translated.
- c. For the **lang_tag** variable in row **152**, supply the language tag to be added to translated annotations (e.g. if the ontology has been translated into French, the language tag will be "fr").
- d. For the **outputfile_name** variable in row **153**, supply the name of the output file to be generated by the program (not necessary to provide a file extension here).

RUNNING THE PROGRAM

Run the *adding_translations_to_owl_file.py* python program by following the procedure in the **Program Running 2 SOP**.

PROGRAM OUTPUT

The output file is an rdf file of the ontology where language tags (for the development language) have been added to term annotations (for the specified annotation properties) and translations of those annotations into a different language have been included, accompanied by language tags for that language.

The file also includes translations of the labels and descriptions of all annotation and object properties in the ontology.

USING THE OUTPUT FILE

1. Post-processing of files:

Because of a fault with the owlready2 python module used in the python program in this SOP, the output file does not contain translations for "rdfs"-type annotation properties. **Translations for "rdfs:" -type annotation must thus be added manually to the file.**

- a. To do this, open the output file in WebProtege or Desktop Protégé and manually add the translations of labels and descriptions from the reviewed translations for those annotation properties.
- b. Download/export the updated file and save with the suffix "post_edited" at the end of the file name.

2. The ontology file can now be used as input in the **Ontology Development 4 SOP**, where ROBOT is used to include associations (recorded in separate excel sheets) into the ontology.

Version No.	Date	Authorizer	External Reviewer	Internal Reviewer	Author	Details of changes
					Jade Hotchkiss	n/a (first version)



Ontology Translation 5.1

SOP: Identifying Terms Needing Translation Updates - Labels and Descriptions

Note:

This SOP is specifically for ontologies providing **translations only for labels and descriptions** of terms. Use **Ontology Translation 5.2 SOP** for an ontology that provides translations for additional term annotations.

PURPOSE

To describe the procedure for identifying updates to ontology terms, and optionally object and annotation properties, that require updates to their annotation translations (labels and descriptions), using the *translation_updates_needed.py* python program.

SCOPE

This procedure is used by curators on a regular basis to provide reviewers/translators with annotations that need their translations updated before a new version of the ontology can be released.

GLOSSARY/DEFINITIONS

- SOP** Standard Operating Procedure: Detailed written instructions to achieve uniformity of the performance of a specific function.
- Protégé** An ontology development software program.
- WebProtégé** An open source, lightweight, web-based ontology editor (developed and maintained by the same team that develops and maintains Protégé).

SYSTEM REQUIREMENTS TO RUN PROGRAMS

Instructions below are for Windows 10.

Ensure that Python is setup correctly on your PC (Use **Program Running 1 SOP**).

REQUIRED INPUT FOR PROGRAM

1. Required input files:
 - a. The latest output file of the *create_terms_for_review_file.py* program (Use **Ontology Development 1 SOP** to obtain this if you do not already have it at hand).
 - b. A .txt file of previously reviewed term annotations translations.
2. Optional input files.

If the curator wants to check if translations of property annotations need updating, the following files are required for input:

- a. .txt files of previously extracted labels and descriptions of object and annotation properties (previous output of **Ontology Development 2 SOP**)
 - b. .txt files of current extracted labels and descriptions of object and annotation properties (updated output of **Ontology Development 2 SOP**)
3. These additional variables are required in the python program code before it is run:
- a. **type_of_update** An indication of whether this is the first time a translation update is being performed or whether it is a subsequent translation update.
 - b. **input_file1_term_ID_cell** The content of the second cell in the header row of the first input file.
 - c. **input_file2_term_ID_cell** The content of the first cell in the header row of the second input file.
 - d. **outputfile_main** The name of the main output file generated by the program.
 - e. **outputfile_new_terms** The name of the output file containing new terms that must have translations auto-generated.
 - f. **outputfile_updated_object_props** The name of the output file containing updated object property annotations that need translating
 - g. **outputfile_updated_ann_props** The name of the output file containing updated annotation property annotations that need translating

PREPARING INPUT FOR PROGRAM

1. Preparing required input files and/or specifying them with relevant variables:
 - a. ensure that the path of the latest .txt file output by the *create_terms_for_review_file.py* program (see **Ontology Development 1 SOP**) is correct for the **input_file1** variable in row 216 of the *translation_updates_needed.py* python program code.
 - b. ensure that the path of the latest .txt file of reviewed translations is correct for the **input_file2** variable in row 217 of the *translation_updates_needed.py* python program code.
2. Preparing optional input files and/or specifying them with relevant variables:
 - a. Ensure that the paths of .txt files output *previously* by the **Ontology Development 2 SOP** are provided for the **old_obj_prop_annotations** and **old_ann_prop_annotations** variables.
 - b. Use **Ontology Development 2 SOP** to produce new files for object property and annotation property annotations.

- a) Ensure that the paths of these newly produced files are provided for the **updated_obj_prop_annotations** and **updated_ann_prop_annotations** variables.
3. Specifying additional required variables in the *translation_updates_needed.py* program code:
- a. For the **type_of_update** variable in row 220, provide either "first" or "subsequent" here to indicate if it is the first or a subsequent translation update.
 - b. For the **input_file1_term_ID_cell** variable in row 221, supply the content of the second cell in the header row of the first input file.
 - c. For the **input_file2_term_ID_cell** variable in row 222, supply the content of the first cell in the header row of the second input file.
 - d. For the **outputfile_main** variable in row 223, supply the name of the main output file to be generated by the program.
 - e. For the **outputfile_new_terms** variable in row 224, supply the name of the output file that will contain new terms that must have translations auto-generated.
 - f. For the **outputfile_updated_object_props** and **outputfile_updated_ann_props** variables in row 241 and 242:
 - a) If property annotation files were input in **step 2** above, supply the name of output files to be generated with updated object properties and annotation properties.
 - b) Otherwise, provide empty strings (i.e. "") for these two variables.

RUNNING THE PROGRAM

After preparing the input as described above, run the relevant python program, either *translation_updates_needed.py* by following the procedure in the **Program Running 2 SOP**.

PROGRAM OUTPUT

The program outputs:

1. A main file, which functions as the template to which auto-translations are added and which will be the file sent to reviewers.
2. A file containing IDs, labels and definitions of new terms that have been added to the ontology.
3. If files for comparing property annotations were provided as input for the program:
 - a. 1 or 2 files might be output containing properties that need translation updates.
 - b. 1 or 2 files might be output listing properties that have been removed.

USING THE OUTPUT FILES

The output files are used in the procedure described in ***Ontology Translation 6.1 SOP*** and ***Ontology Translation 8.1 SOP***.

Version No.	Date	Authorizer	External Reviewer	Internal Reviewer	Author	Details of changes
					Jade Hotchkiss	n/a (first version)

Ontology Translation 6.1

SOP: Preparing Files for Translation Updates - Labels and Descriptions

Note:

*This SOP is specifically for ontologies providing **translations only for labels and descriptions** of terms. Use **Ontology Translation 6.2 SOP** for an ontology that provides translations for additional term annotations.*

PURPOSE

To describe the procedure for using output files from **Ontology Translation 5.1 SOP** to prepare a file for reviewers/translators to update translations.

SCOPE

This procedure is used by curators on a regular basis to provide reviewers/translators with annotations that need updating before a new version of the ontology can be released.

GLOSSARY/DEFINITIONS

- SOP** Standard Operating Procedure: Detailed written instructions to achieve uniformity of the performance of a specific function.
- Babylon** Translation software used to produce automatic translations.

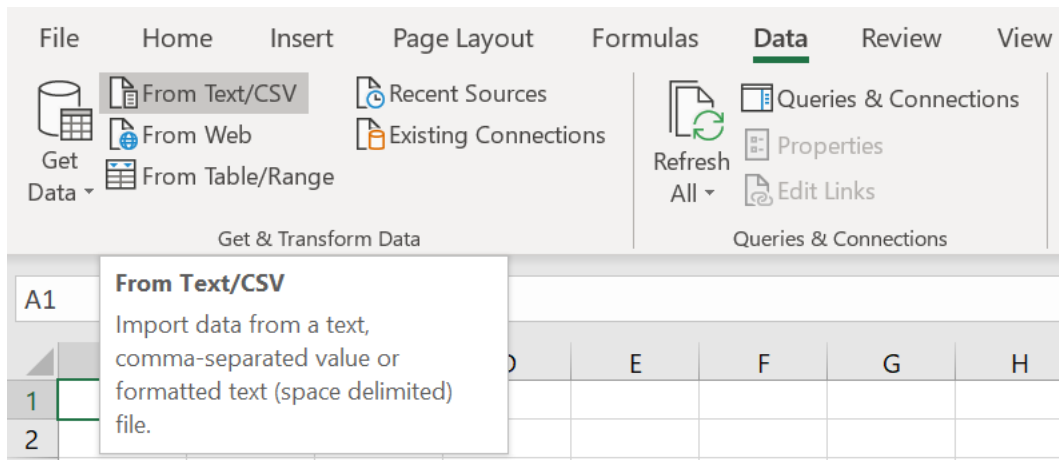
PROCEDURE - USING THE OUTPUT FILES FROM **ONTOLOGY TRANSLATION 5.1 SOP**

1. Auto-generate translations for new terms:
 - a. Open the output file containing labels and descriptions of new ontology terms in Excel and resave it as an .xlsx file.
 - b. Use Babylon to translate the Excel file into the necessary language. A translated Excel file will be generated.
 - c. Save the generated Excel file appropriately (named e.g. *SCDO_New_French_Terms_Auto_Translation_***DATE**).

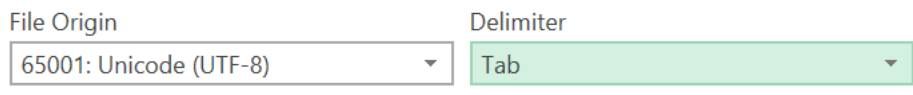
2. Editing the main file:
 - a. Open the main/first output file in Excel and resave it as an Excel file (named e.g. *Translation_Update_Needed_***DATE**):

NB: When opening the main file's data in Excel for the first time:

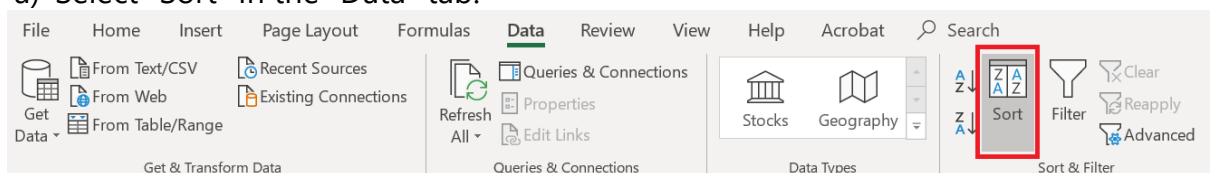
- Open a new blank Excel Workbook
- From the "Data" tab, import the content of the main output file.



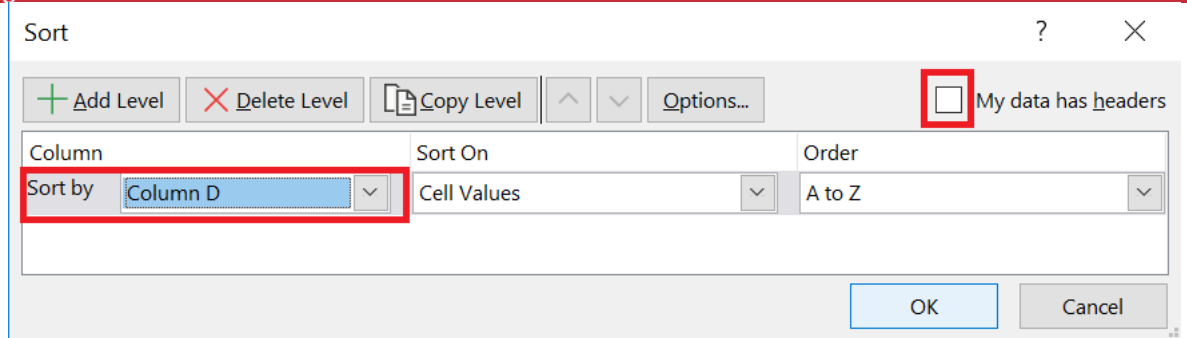
- Make sure to set the "File Origin" to "Unicode (UTF-8)" and the "Delimiter" to "Tab". This will ensure that there aren't strange characters in the output.



- Copy the auto-translated labels (column B) and definitions (column C) from the file generated in *step 1* into column F (current label translation) and column M3 (current definition translation) of the main file, respectively (you will see the new terms are found in the same order at the top of the main file, with "auto-generated only" in column D).
- Copy all content of the sheet to a new sheet named "Review This" and delete the header row in the new "Review This" sheet.
- Sort/order the rows in the new sheet by column D:
 - Select "Sort" In the "Data" tab.



- In the "Sort" window, deselect "My data has headers" and set the "Sort by" field to "Column D".



c) Select "OK".

e. Colour cells in column D accordingly:

a) Colour all column D cells that contain the text "auto-generated" to light green:

D
auto-generated
auto-generated
auto-generated
auto-generated

b) Colour all column D cells that contain the text "edited previously - needs re-checking" to light orange:

D
edited previously - needs re-checking
edited previously - needs re-checking
edited previously - needs re-checking
edited previously - needs re-checking

c) Colour all column D cells that contain the text "needs updating" to blue:

D
needs updating
needs updating

f. Edit the file for terms that have updated labels:

Copy all column E content (current labels) for rows that have "needs updating" in column D to an Excel file (named e.g. *term_labels_needing Updating*) and translate the file with Babylon.

a) Save the translated file from Babylon as *updated_term_labels_auto_translations_from_babylon DATE*

b) Check for unnecessary spaces at the beginning of the translated labels and delete any you find.

c) Copy translations from this file into corresponding column F cells of the sorted sheet.

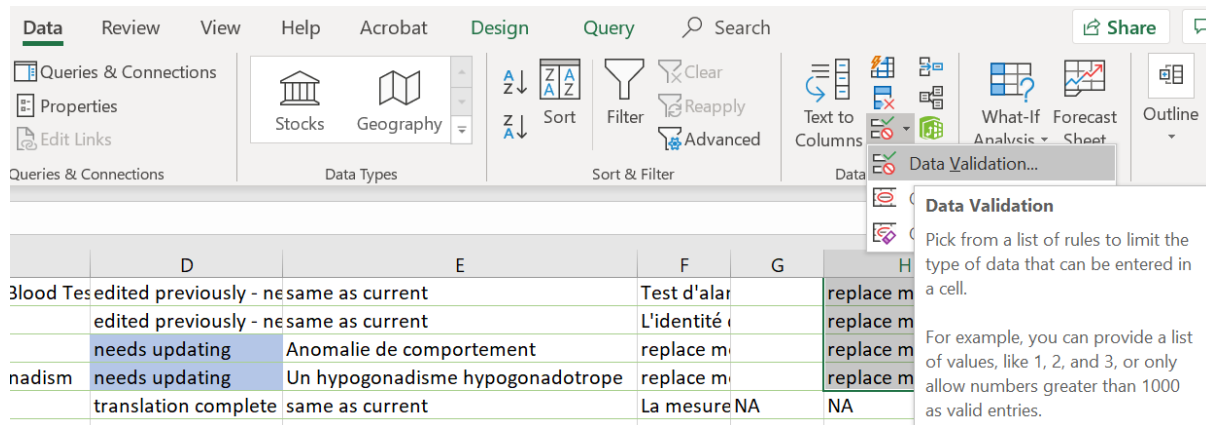
g. Add data validation to column H:

a) Add a new sheet named "For DV – Ignore this" and copy this content to it:

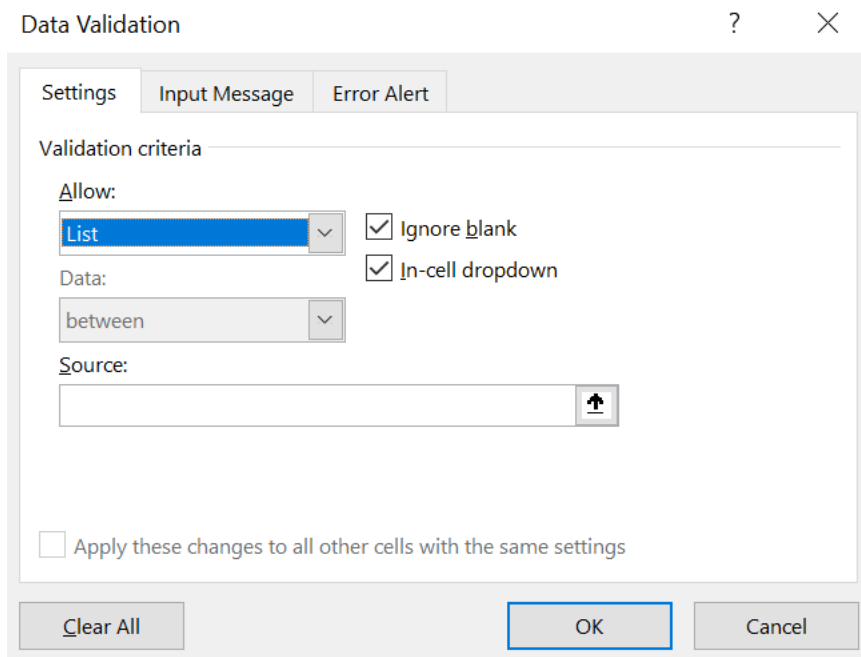
Data validation
checked, edited and validated
checked and validated without any edits
edited but still need to be rechecked again

b) Copy the text "*Indicate translation work done here*" into column H for all rows that have column D coloured (green, orange and blue).

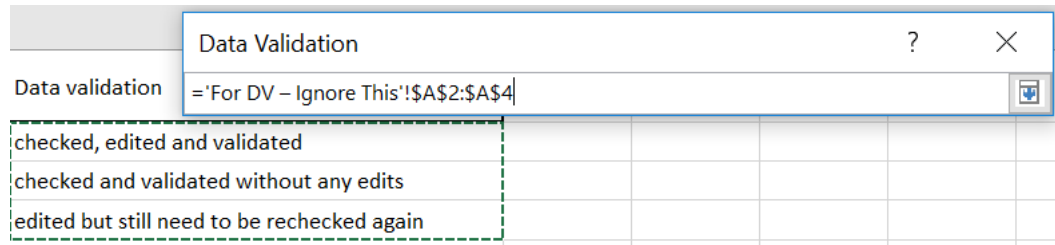
c) Select column H cells that now contain "*Indicate translation work done here*" and open the Data Validation window in the "Data" tab:



d) Select "List" from the "Allow" drop-down list:

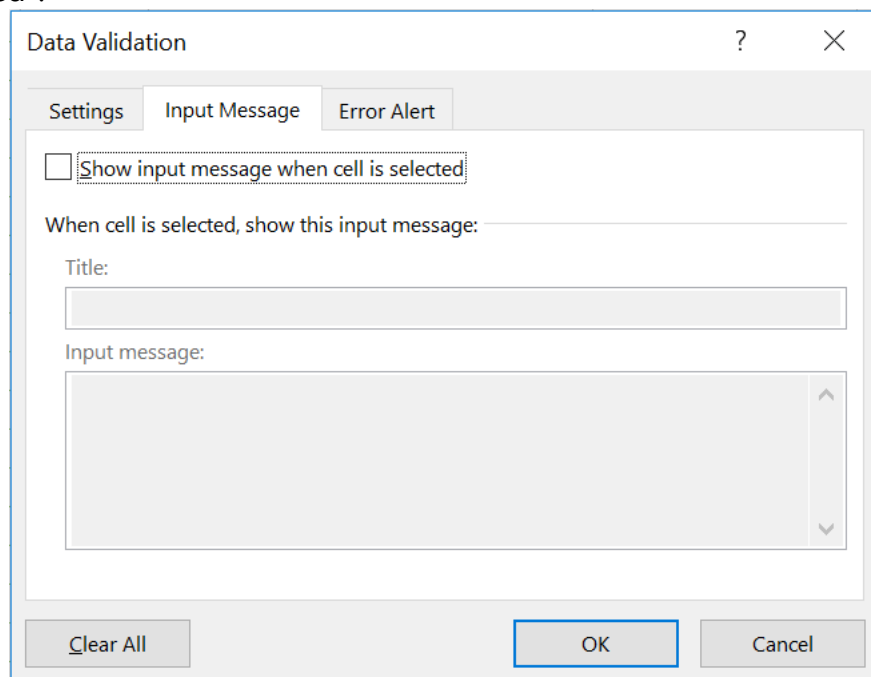


- e) Select a source for the list by clicking on the arrow on the right of the “Source” field and then selecting the data validation options provided in the “For DV – Ignore This” sheet:

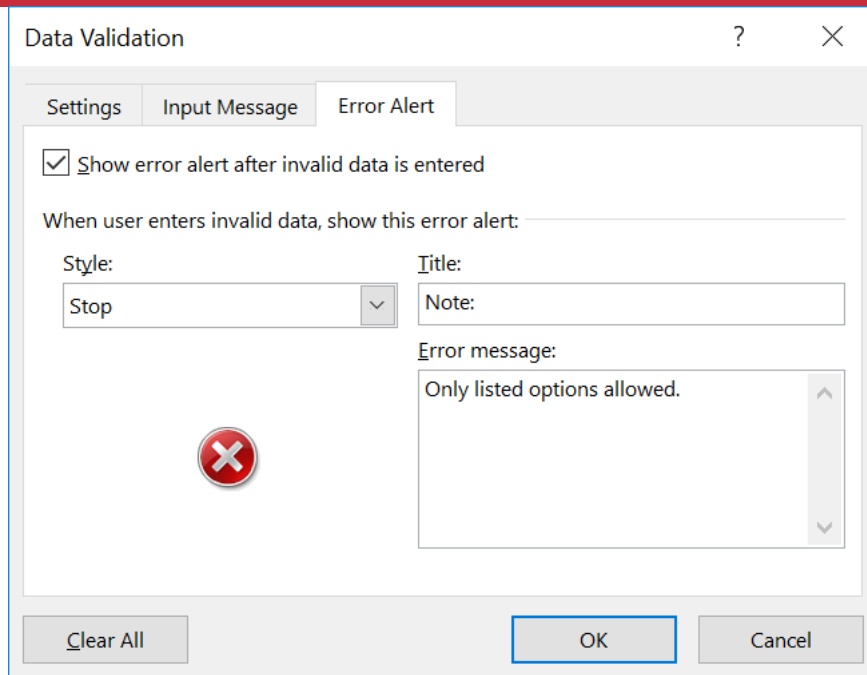


- f) Click the arrow on the right again to go back to the “Data Validation” window.

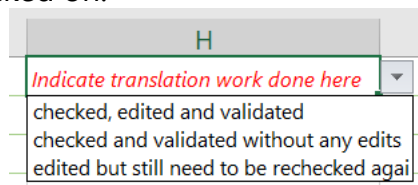
- g) In the “Input Message” tab, deselect “Show input message when cell is selected”:



- h) In the “Error Alert” tab, select “Stop” from the “Style” drop-down and enter a title and error message as in the image below:



- i) Click "OK" in the "Data Validation" window. The column H cells with data validation will now have a drop-down list like this when they are clicked on:



- h. Sort/order the rows in the sheet by column K. (Same steps as used to sort by column D in *step 2d* above).
- i. Colour all column K cells that contain the text "auto-generated", "edited previously - needs re-checking" and "needs updating" to light green, light orange and blue, respectively (as was done in *step 2e*) above).
- j. Edit the file for terms that have updated descriptions:
Copy all column L content (current definitions) for rows that have "needs updating" in column K to an Excel file (named e.g. *term_definitions_needing Updating*) and translate the file with Babylon.
- Save the translated file from Babylon as *updated_term_definitions_auto_translations_from_babylon_DATE*.
 - Check for unnecessary spaces at the beginning of the translations and delete any you find.
 - Copy translations from this file into corresponding column M cells of the sorted sheet.

- k.** Add data validation to column O:
- Copy the text "*Indicate translation work done here*" into column O for all rows that have column K coloured (green, orange and blue).
 - Select column O cells that now contain "*Indicate translation work done here*" and open the Data Validation window (as in *step 2g(c)* above).
 - Follow the steps that were used to add data validation to column H (see *step 2g(d-i)* above) to add data validation to the selected cells in column O.
- l.** Final edits to file:
- Add a new row at the top of the sheet of terms that you have been editing and copy the header row from the original sheet of terms across.
 - Embolden the header row content.
 - Add a bold border to the header row.
 - Colour the headers for columns G and N bright yellow.
 - From the "View" tab, freeze the top row.
 - Delete the original sheet.
 - Add a new "Review stats" sheet.
 - Copy this table to the new sheet:

Terms sent for label translation checks:	
Terms sent for description translation checks:	
Labels:	
checked, edited and validated	
checked and validated without any edits	
edited but still need to be rechecked again	
Definitions:	
checked, edited and validated	
checked and validated without any edits	
edited but still need to be rechecked again	
No definition provided	

- Sort the "Review this" sheet by row D and see how many labels need translation reviews (those containing "auto-generated", "edited previously - needs re-checking" and "needs updating"). Add this number to the top right of this table.
- Sort the "Review this" sheet by row K and see how many descriptions need translation reviews (those containing "auto-

generated”, “edited previously - needs re-checking” and “needs updating”). Add this number to the next cell down of this table.

- h) Save the Excel file as “**ONTABRV_label_and_definition_LANGUAGE_translations_for_review_DATE**”, replacing:
- i. **ONTABRV** with the relevant ontology abbreviation (e.g. SCDO)
 - ii. **LANGUAGE** with the relevant language being translated into (e.g. French)
 - iii. **DATE** with the current date (e.g. 11July2019)

3. If one or two files were output containing property annotations needing translation updates, for each file:
- a. Import the file content into a new Excel file and save the file as either *updated_***object***_property_annotations_for_review_***DATE** or *updated_***annotation***_property_annotations_for_review_***DATE**, depending on the content of the file.
 - b. Translate the file with Babylon.
 - c. Save the translated file generated by Babylon as either *updated_***object***_property_auto_translations_from_babylon_***DATE** or *updated_***annotation***_property_auto_translations_from_babylon_***DATE** depending on the content of the file.
 - d. Check for unnecessary spaces at the beginning of the translated labels and definitions and delete any you find.
 - e. Copy label and description translations from this file into corresponding “Translated...” columns of the original file.
 - f. Save the file.

PROCEDURE – SENDING FILES TO REVIEWERS/TRANSLATORS

Email the main Excel file (step 2) and any property update files (step 3) generated in this SOP along with the **Ontology Translation SOP 7.2** to the reviewers/translators.

Version No.	Date	Authorizer	External Reviewer	Internal Reviewer	Author	Details of changes
					Jade Hotchkiss	n/a (first version)

Ontology Translation 7.1

SOP: Reviewing Translations – First Time

PURPOSE

To describe the procedure for reviewing translations of an ontology for the first time.

SCOPE

This procedure is used by reviewers/translators to review translations of an ontology for the first time.

GLOSSARY/DEFINITIONS

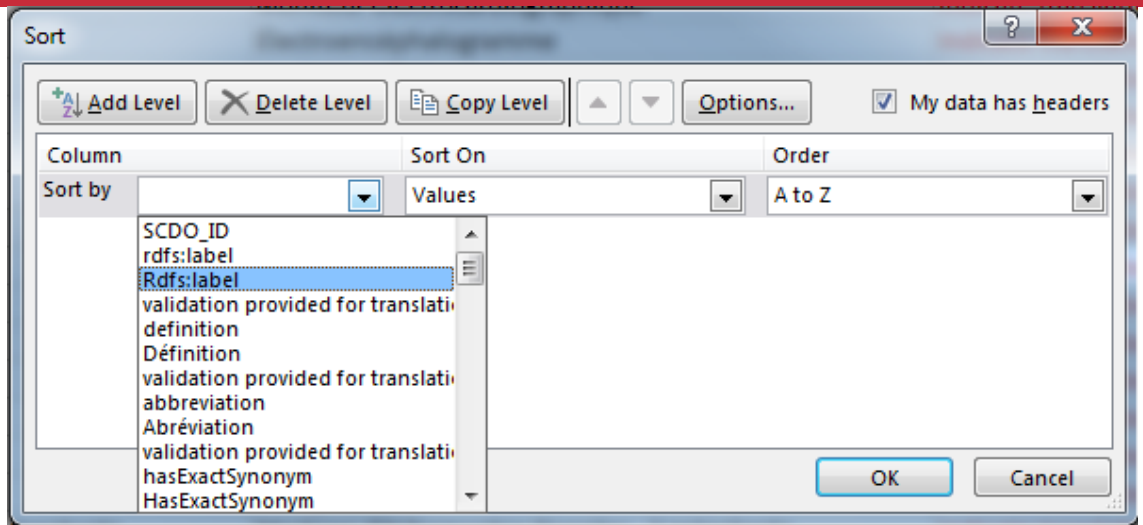
SOP Standard Operating Procedure: Detailed written instructions to achieve uniformity of the performance of a specific function.

PROCEDURE

Send the 3 prepared files to reviewers with these instructions:

1. Work only needs to be done in the “Review this” tab/sheet of all files. (All other tabs/sheets can be ignored.)
2. Copy the auto-translation to the blank column and edit it as appropriate. If no edits are needed, copy it and leave as is.
3. Indicate what you did in the “validation provided for translation” column by selecting the appropriate option from the drop-down list provided:
 - a) Select **“checked, edited and validated”** if the copied auto-translation was edited and is now considered acceptable.
 - b) Select **“checked and validated without any edits”** if the copied auto-translation was not edited because it was considered acceptable.
 - c) Select **“edited but needs re-checking”** if the copied auto-translation was edited but not necessarily acceptable and should be checked by another reviewer/translator.
 - d) Select **“NA”** only if an annotation was not provided.
4. If additional annotations (more than just labels and definitions) are being translated:

For annotation properties where not many annotations are present, sorting the file by that property/column (Data tab → Sort → see figure below) will make it easier to view all terms with that property and to indicate “NA” in the “validation provided for translation” column for all terms with no annotations for a specific property.



Version No.	Date	Authorizer	External Reviewer	Internal Reviewer	Author	Details of changes
					Jade Hotchkiss	n/a (first version)

Ontology Translation 7.2

SOP: Reviewing Translation Updates - Labels and Descriptions

Note:

*This SOP is specifically for ontologies providing **translations only for labels and descriptions** of terms. Use **Ontology Translation 7.3 SOP** for an ontology that provides translations for additional term annotations.*

PURPOSE

To describe the procedure for reviewing translation updates.

SCOPE

This procedure is used by reviewers/translators on a regular basis to review updates to a translation of an ontology before a new version of the ontology can be released.

GLOSSARY/DEFINITIONS

SOP Standard Operating Procedure: Detailed written instructions to achieve uniformity of the performance of a specific function.

PROCEDURE FOR REVIEWING TERM ANNOTATION TRANSLATIONS

Note: *Work only needs to be done in the "Review this" sheet of this file:*

ONTOLOGYABBREVIATION_label_and_definition_LANGUAGE_translations_for_review_DATE"

1. Reviewing label translations [You will focus on columns D – H]:

- a.** Sort the sheet by column D (label translation status).

[This will put the new terms with auto-generated label translations at the top, followed by any term label translations that were edited before but need re-checking and any that need updating. The terms below those will contain "translation complete" in column D and can be ignored.]

- b.** For terms containing "auto-generated", "edited previously - needs re-checking" or "needs updating" in column D, the current translated label provided in column F needs to be reviewed and final translations need to be provided in column G.

- a) Compare the translation provided in column F with the current label in the original language in column E.
- b) If a translation does not need editing, just copy it across from column F to column G, otherwise provide an edited version in column G.

- c) For each term, select an option in column H (“validation provided for current label”) to indicate what validation was provided for the term’s label translation.
- Select “**checked, edited and validated**” if the copied auto-translation was edited and is now considered acceptable.
 - Select “**checked and validated without any edits**” if the copied auto-translation was not edited because it was considered acceptable.
 - Select “**edited but needs re-checking**” if the copied auto-translation was edited but not confirmed and should be checked by another reviewer/translator.
2. Reviewing definition translations [You will focus on columns K – O]:
- a. Sort the sheet by column K (definition translation status).
[This will put the new terms with auto-generated definition translations at the top, followed by any term definition translations that were edited before but need re-checking and any that need updating. The terms below those will contain either “no definition provided” or “translation complete” in column K and can be ignored.]
- b. For terms containing “auto-generated”, “edited previously - needs re-checking” or “needs updating” in column K, the current translated definition provided in column M needs to be reviewed and final translations need to be provided in column N.
- a) Compare the translation provided in column M with the current definition in the original language in column L.
- b) If a translation does not need editing, just copy it across from column M to column N, otherwise provide an edited version in column N.
- c) For each term, select an option in column O (“validation provided for current definition”) to indicate what validation was provided for the term’s definition translation.
- Select “**checked, edited and validated**” if the copied auto-translation was edited and is now considered acceptable.
 - Select “**checked and validated without any edits**” if the copied auto-translation was not edited because it was considered acceptable.
 - Select “**edited but needs re-checking**” if the copied auto-translation was edited but not confirmed and should be checked by another reviewer/translator.

PROCEDURE FOR REVIEWING PROPERTY ANNOTATION TRANSLATIONS

Note: Work in these files:

*updated_object_property_annotations_for_review_***DATE**

*updated_annotation_property_annotations_for_review_***DATE**

If these files have not been sent, this section can be ignored!

1. Compare the translated labels provided in column C with the current label in the original language in column B and compare the translated definitions provided in column F with the current definition in the original language in column E.
2. If a translation does not need editing, just copy it across from the “translated” column to the “reviewed” column, otherwise provide an edited version in the “reviewed” column.

Version No.	Date	Authorizer	External Reviewer	Internal Reviewer	Author	Details of changes
					Jade Hotchkiss	n/a (first version)

Ontology Translation 8.1

SOP: Preparing Reviewed Property Annotations

PURPOSE

To describe the procedure for preparing reviewed translations of property annotations as input for **Ontology Translation 4 SOP**.

SCOPE

This procedure is used by reviewers/translators to prepare files of property translations for inclusion into an updated translation of the ontology.

GLOSSARY/DEFINITIONS

SOP Standard Operating Procedure: Detailed written instructions to achieve uniformity of the performance of a specific function.

PROCEDURE

*Note: The procedure depends on whether files were produced in **Ontology Translation SOP 5.1** that indicate that property annotations need updating.*

1. If no updates need to be made (no files were generated by **Ontology Translation SOP 5.1** that indicate that property updates are necessary), .txt files of reviewed translations of object and annotation properties used in the previous translation will be re-used as input for **Ontology Translation 4 SOP**.
2. If updates need to be made:
 - a. Make a copy of the relevant Excel files of reviewed property annotations (could be file of only object properties, or only annotation properties, or both) used in the last translated version of the ontology.
 - b. Use reviewed updates to translations of property annotations (received back from reviewers as part of **Ontology Translation SOP 6.1**) to update the Excel files from *step 2.a*. Some properties may be new, in which case a new row will need to be created for that property and its annotations.
 - c. Remove from the Excel files the properties indicated in the *removed_object_properties.txt* and *removed_annotation_properties.txt* files (output by **Ontology Translation SOP 5.1**).

Version No.	Date	Authorizer	External Reviewer	Internal Reviewer	Author	Details of changes
-------------	------	------------	-------------------	-------------------	--------	--------------------

					Jade Hotchkiss	n/a (first version)
--	--	--	--	--	----------------	---------------------

Program Running 1

SOP: Python Program System Requirements

PURPOSE

To describe the system requirements for running a python program.

SCOPE

Curators need to have these system requirements met in order to run a python program.

GLOSSARY/DEFINITIONS

Python The programming language in which the program used in this SOP was written.

Protégé An ontology development software program.

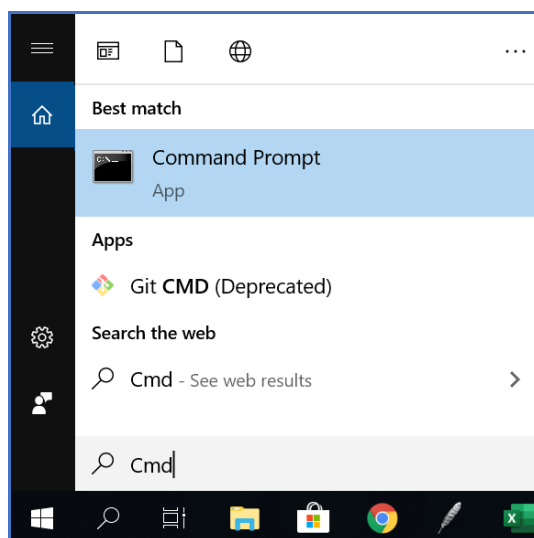
SOP Standard Operating Procedure: Detailed written instructions to achieve uniformity of the performance of a specific function.

SYSTEM REQUIREMENTS TO RUN PYTHON PROGRAM

Instructions below are for Windows 10.

For a python program to run successfully, the following need to be installed:

1. Python.
 - a) To check if python is installed:
 - a. Open a command prompt terminal window (type "cmd" in the search bar and select *Command Prompt* from the search results).



- b. In the Command Prompt window, type "python" and press Enter.

c. If Python is installed and in your path, then this command will run python.exe and show you the version number.

b) If it is not installed, this message will appear:

```
'python' is not recognized as an internal or external command,  
operable program or batch file.
```

You will need to download and install python from:
<https://www.python.org/downloads/windows/>

Note: During the installation process, leave defaults as-is .

Version No.	Date	Authorizer	External Reviewer	Internal Reviewer	Author	Details of changes
					Jade Hotchkiss	n/a (first version)

Program Running 2

SOP: Python Program Execution

PURPOSE

To describe the procedure for executing a python program.

SCOPE

This procedure can be used by anyone wanting to run a python program using Windows 10.

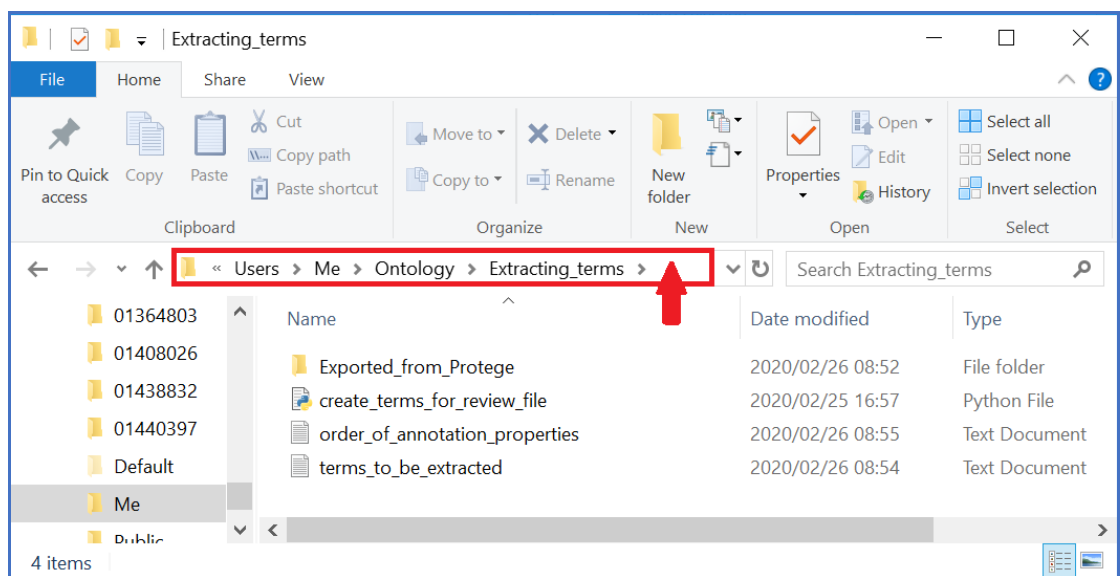
GLOSSARY/DEFINITIONS

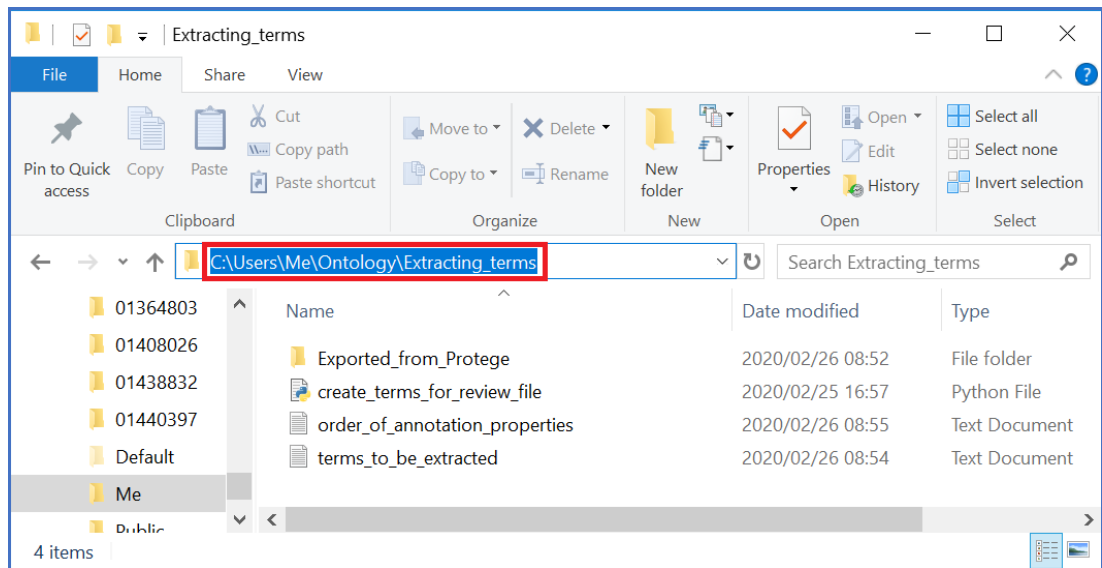
Python The programming language in which the program used in this SOP was written.

SOP Standard Operating Procedure: Detailed written instructions to achieve uniformity of the performance of a specific function.

EXECUTING A PYTHON PROGRAM

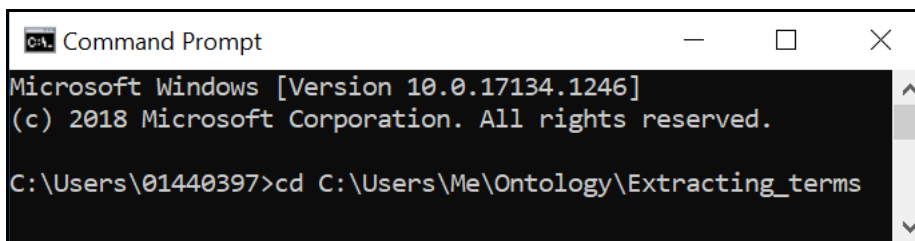
1. Open a command prompt (cmd) window.
2. Copy the location path of the folder where the python program is located. To do this:
 - a. open the folder and click to the right of the content in the address bar. The path of the folder will appear highlighted.





b. Press "ctrl" and "c" to copy the path.

3. In the command prompt window, type "cd" followed by a space and then right click to paste the copied path. Press Enter.

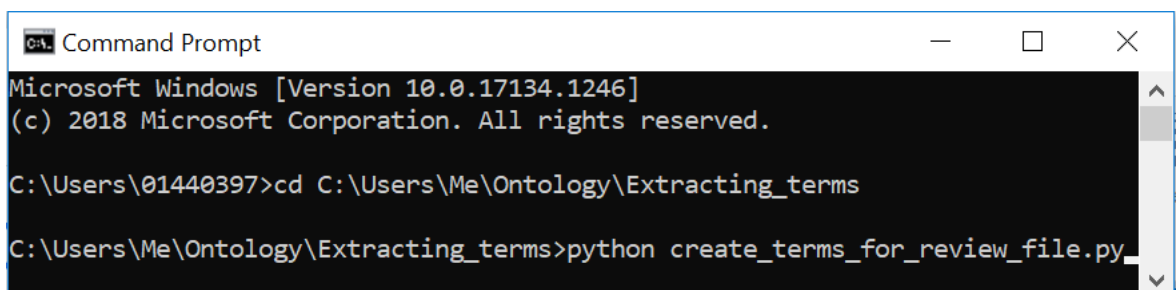


```

Microsoft Windows [Version 10.0.17134.1246]
(c) 2018 Microsoft Corporation. All rights reserved.

C:\Users\01440397>cd C:\Users\Me\Ontology\Extracting_terms
  
```

4. Enter the command "python XXX", where XXX is the name of the python program (including the ".py" file extension), and press Enter.



```

Microsoft Windows [Version 10.0.17134.1246]
(c) 2018 Microsoft Corporation. All rights reserved.

C:\Users\01440397>cd C:\Users\Me\Ontology\Extracting_terms

C:\Users\Me\Ontology\Extracting_terms>python create_terms_for_review_file.py
  
```

PROGRAM ERRORS

If you get an error message similar to this one:

Syntax Error: (unicode error) 'unicodeescape' codec can't decode bytes in position 2-3: truncated \UXXXXXXXX escape: <string>, line 216, pos 20

Ensure that all “\” in input file paths are replaced by “/”.

Version No.	Date	Authorizer	External Reviewer	Internal Reviewer	Author	Details of changes	of (first version)
					Jade Hotchkiss	n/a	(first version)

Program Running 3

SOP: Setting up Git on PC

PURPOSE

To describe the procedure for installing and configuring Git on a Windows installation in order to compile OWL files remotely using ROBOT (**Ontology Development 4 SOP**).

SCOPE

This procedure can be used by anyone wanting to install and configure Git on Windows 10.

GLOSSARY/DEFINITIONS

- Git** Git is a distributed version-control system for tracking changes in source code during software development. It is designed for coordinating work among programmers, but it can be used to track changes in any set of files.
- SOP** Standard Operating Procedure: Detailed written instructions to achieve uniformity of the performance of a specific function.

PROCEDURE

Instructions below are for Windows 10.

1. Check whether Git is already installed on your PC.
 - a) Open a command prompt terminal window (type "cmd" in the search bar and select *Command Prompt* from the search results.).

- b) Type **git --version** into the window and press Enter.
 - a. If Git is installed, a message like this will appear:

```
git version 2.25.0.windows.1
```

In this case you have Git for Windows installed.

→ Continue with **step 2**.

- b. If Git is not installed, you will see this message:

```
'git' is not recognized as an internal or external  
command, operable program or batch file.
```

→ Continue with **step 3**.

2. Check if you need to update Git.

Go to <https://git-scm.com/download/win>

- a. Compare your version of Git to the latest version of Git for Windows (will show at top of page).
 - a) If your version of Git (as seen in 1(b) above) is not up to date, continue with **step 3**.
 - b) If your version of Git is up to date, you do not need to continue with this SOP.

3. Install Git.

Follow instructions 1-12 under the *Installing Git on Windows* section on this web page: <https://www.computerhope.com/issues/ch001927.htm>

Version No.	Date	Authorizer	External Reviewer	Internal Reviewer	Author	Details of changes
					Jade Hotchkiss	n/a (first version)

Administration, Coordination and Management 4

SOP: Managing Online Meetings: SADaCC & SPARCO

PURPOSE:

To describe the procedure for managing and organising online meetings (SADaCC).

SCOPE

This procedure applies to online meetings organised by SPARCO and SADaCC members.

GLOSSARY/DEFINITIONS

SADaCC Sickle Africa Data Coordinating Centre at the University of Cape Town

SOP Standard Operating Procedure: Detailed written instructions to achieve uniformity of the performance of a specific function.

RESPONSIBILITIES & PROCEDURES

When		Action	Templates	Assignee
Before		Asking for confirmation of attendance, sending apologies if not available		
		Logging onto the meeting on time		
		Appointment of acting Chair if chair/vice chair not available		
		Quorum		

	10 days	Ask for agenda items	Dear XXX, We kindly request you to review/edit the draft agenda for the XXX meeting by (3 days) Date: Time: Meeting Room: Chair: Based on our schedule, XXX will be chairing the meeting? Best regards, Proposed Agenda: Attach documents in a SIA template	Vicky & Furahini
	5 days	Circulate agenda items & Action items from previous meeting	Dear XXX members, Find agenda items for the next XXX meeting and action items from the previous meeting. Date: Time: Meeting Room: Proposed Agenda: Action Items from previous meeting Attach documents Best regards,	Vicky & Furahini
	45 mins before meeting	Open the meeting room & recording session	Send a reminder announce that the room has been opened	Vicky Nembaware
	30 mins before	In the event Vicky fails to open the meeting Noluthando or Mario will open	Send a reminder to announce that the room has been opened	Noluthando Manyisa/ Mario Jonas
During	1 hour	Taking minutes	Take minutes and recordings too	
After	Next day	Circulate draft minutes & action items and ask for feedback	Dear XX members, Find attached draft minutes & action items from the meeting. Please review and edit before the XX. Best regards,	

	7 days	Circulate final minutes & action items	Dear XX members, Find attached finalized minutes & action items from the meeting. Best regards,	
--	--------	--	---	--

Administration, Coordination and Management 5

SOP: Foreign students at University of Cape Town

PURPOSE:

To describe the procedure for registering at the University of Cape Town.

SCOPE

This procedure applies to foreign students who might be in the process of registering at the University of Cape Town (UCT).

GLOSSARY/DEFINITIONS

SADaCC Sickle Africa Data Coordinating Centre at the University of Cape Town

SOP Standard Operating Procedure: Detailed written instructions to achieve uniformity of the performance of a specific function.

UCT University of Cape Town

RESPONSIBILITIES & PROCEDURES

1. Letter of award

Once this is received by students, they must formally accept the offer by sending an acceptance letter to the PI and project manager. Please cc Genemapadmin@uct.ac.za in all correspondence.

2. Registration at UCT website

2.1 Please register via the UCT website:

https://applyonline.uct.ac.za/psc/onapplic/EMPLOYEE/HRMS/c/UCT_PUBLIC_MENU.UCT_SS_APPL.GBL?FolderPath=PORTAL_ROOT_OBJECT.UCT_SS_APPL_GBL&IsFolder=false&IgnoreParamTempl=FolderPath%2clsFolder

Students will register here and apply under Health Sciences Faculty.

Once registration is done students will receive an email confirmation.

2.2 Fees

UCT will also email your fees due for the year.

This should be emailed to the student NOT the PI. Your fees will be covered by your scholarship. Your scholarship will be paid into your student account once registration is complete. UCT Fees office will automatically deduct your fees before the balance of your scholarship is paid out. Please note that funds from your student account can only be transferred to a South African bank account (please see below).

Please send a copy of your registration to Genemapadmin@uct.ac.za

3. Medical Aid

A Student Medical Aid is required for any Foreign students to receive a South African study visa. UCT works closely with Momentum Medical aid. We advise that you use the preferred medical aid unless otherwise specified.

Please register with Momentum and make an online Payment with any Master, Visa Credit or Debit Card. You can also do a Bank Transfer directly to Momentum.

<https://www.momentum.co.za/for/you/products/health/mmsa-student-application>

Once your medical aid has been confirmed please send a copy of confirmation email to Genemapadmin@uct.ac.za.

For any Medical aid queries please contact Debbie van Onselen at Debbie.VanOnselen@momentum.co.za

NB: UCT Genemap Admin will not sign or assist with any payments of Medical aid for any Students. If Students choose to use other medical insurance companies, we can't guarantee visa approval by the Consulate.

4. Accommodation

On campus accommodation for UCT is normally fully booked a year in advance. GeneMap has access to a 3 Bedroom on-campus apartment that we rent to students. Kindly note this is only for Single persons and not for students traveling with Spouses and Children. If you are interested, please send an Email to Genemapadmin@uct.ac.za to find out availability of this accommodation. If we cannot assist with accommodation, please seek accommodation with UCT on-campus accommodation department.

Email: nomonde.sikiti@uct.ac.za or melissa.patience@uct.ac.za

If they advise there is no on campus accommodation you can seek UCT official Off Campus accommodation options:

1. Email ocsas@uct.ac.za

2. <http://www.dsa.uct.ac.za/student-housing/off-campus-student-accommodation-services/accommodation>

NB: UCT Genemap Admin will not sign Leases and assist with seeking accommodation for any Students. Kindly note that if you will not be staying in our GeneMap Flat you will need to arrange a payment directly to the project fund. Sub-lease agreements will be sent to potential occupants.

5. IAPO (International Academic Programmes Office)

Letter for Visa

UCT International office will provide you with a letter of support for your Visa application. For you to get this letter you must be registered already.

Please send IAPO the following:

- Registration confirmation

- Medical Aid Confirmation
- Confirmation of accommodation

Please email them on aipo@uct.ac.za and cc Genemapadmin@uct.ac.za

6. Visa requirements

Once all the above is confirmed please apply for a Study Visa at the South African Embassy in your country of stay.

Student visa for South Africa Has the following Requirements:

- a. Student offer letter.
- b. Copy of the PI South African Identity document.
- c. Medical Aid confirmation for the period of stay.
- d. Confirmation of Accommodation.
- e. IAPO letter of confirmation for a study Visa.

For Visa confirmation of information please ask them for any queries to please contact Gabby Emjedi or Mzodidi Tutuka at

Gabby.emjedi@uct.ac.za or mzodidi.tutuka@uct.ac.za Tel: +27 021 650 5632. As soon as the student has submitted their application to the consulate please inform Gabby Emjedi . Please keep the Genemap admin team up to date with any changes or progress in your visa application.

As soon as Visa is approved please inform the PI and cc Genemapadmin@uct.ac.za.

7. Traveling to and arriving in South Africa

1. Traveling to South Africa

Funding for travel to South Africa is up to the discretion of the PI. In general, the project fund will cover one economy class return air ticket from their city of residence. . Once the student has a visa a Flight will be booked with a Travel agent. Please note that no flights will be booked unless the traveling party agrees to the flight schedule.

NB: If the student misses their flight for any reason or cannot come as per the flight booked you will need to contact the following people immediately. Please note that you must have a valid medical reason to miss a flight. This must be communicated to the travel agent in advance.

Leslie Lund on Leslie.lund@flightcentre.co.za /+27 82 348 5098

Gabby Emjedi on Gabby.Emjedi@uct.ac.za /+27 83 487 5305

Any Flights missed with not be covered by Genemap for any reason.

2. Arrival in Cape Town

2.1 Local Transfer

Please inform the admin team in advance if you require assistance with transport. A local transfer will be arranged with local transport company. Your driver will await you at the arrival terminal with a name board. Students will be CC in this confirmation email. Please let the admin team know if there is any changes to your flights. The driver will take you to your accommodation. Please contact kingdom Tours on Cell: +27833060550 or Email: info@kingdomtt.co.za

2.2.1- First day on campus

Please notify Genemapadmin@uct.ac.za when you will come to campus. (date and time). Staff member will need to collect student at security. Please arrange this at least a day in advance.

Campus: Health Science Faculty at Anzio Road, Observatory.

2.2.2 On-campus registration process

Go to post graduate office at the Faculty of Health Sciences and get the final confirmation form.

Once this form is completed the PI will sign.

Please return the signed form to Post graduate office. The student will need this confirmation to obtain a student card.

2.3 Student cards

NO CARD NO ACCESS.

This done on UCT Upper campus please go with the confirmation and your passport to Properties and Services. Please see Map below:

2.4 IAPO Clearance

Once the student has their student card please go to IAO on Middle campus with the following:

- Registration confirmation
- Medical Aid Confirmation
- Confirmation of accommodation
- Passport
- Student card

They will then give post graduate office final clearance for your registration.

This is the last step in your registration process.

2.5 Upload Banking details on the UCT student portal:

IF YOUR BANKING DETAILS IS NOT CAPTURED ON THE UCT STUDENT PORTAL THEN NO FUNDS WILL BE TRANSFERRED FROM YOUR STUDENT ACCOUNT.

Please open a South African Bank account as soon as possible. You will require the following:

1.Passport

2.Registration confirmation

3.Accommodation letter

NB GeneMap Admin Staff will not be available to assist with any bank account related queries.

2.6 Appointment with PI and project Manager/Induction

Please send an email to Genemapadmin@uct.ac.za set up an appointment to meet the team to discuss and sign your Deliverables and 1st installment of the scholarship. Your induction in the unit will be done the same day.

All postgraduate students are required to submit a progress report every six months to the PI and project manager. Your progress report must first be approved before the next installment of the scholarship is processed.

Forms & Documents 1

SOP: Site Assessment Checklist

IDENTIFICATION

Full site name		Date	
Brief site name		DAG name	

DESCRIPTION & LOCATION

Physical address			
Postal address			
Location of main site			
Distance from airport			
What is the road condition? (e.g. dirt, tarmac)			
Location(s) of satellite site(s)			
Web page			
Type of site	Hospital	<input type="checkbox"/> Private	
		<input type="checkbox"/> Provincial	
		<input type="checkbox"/> Referral / Teaching	
		<input type="checkbox"/> Military	
	Other	<input type="checkbox"/> Clinic (not part of hospital or university)	
		<input type="checkbox"/> Clinical Research Unit (CRU) e.g. at university	

CONTACTS

Mark the Primary Contact Person with a *

Include email address and phone numbers where possible.

Site PI	
Other Investigators	
Research Site Coordinator	
Data Manager	
Data Capturer(s)	
Research Assistants	
IT Representative	
Other Key Staff (e.g. Bioinformaticist, Infectious Diseases Physician)	

ON-SITE FACILITIES

Which of the following are available at the site ?	Yes	No	N/A	Comment
Data management office/area				
Designated study document storage area available on site (e.g file room)				
Printer/Photocopy machine				
Scanner				
Long-term document archiving facilities on-site				
Long-term document archiving facilities off-site				
Meeting room				
Teleconference capability				
Administrative support office				

DATA MANAGEMENT-RELATED COMPUTING

Which of the following does your data management staff primarily use?		Yes	No	N/A	Comment
Computer types	Laptops				
	Desktop computers				
	Mobile devices (e.g. tablets or cellphones)				
Operating system	Microsoft Windows				
	Apple Mac				
	Linux operating system				
Internet connection	Institutional WiFi				
	Private WiFi connection				
	LAN connection				
	Mobile internet connection				
Databases	Microsoft Excel				
	Microsoft Access				
	REDCap				
	Other software (specify)				
Data location ('live' data)	Individual computer(s)				
	Network drive				
	Own server				
	Institutional server				
Backup storage	Individual computer(s)				
	Network drive				
	Own server				
	Institutional server				

Which of the following does your data management staff primarily use?		Yes	No	N/A	Comment
	Flash drive / memory stick				
	External hard drive				
	CDs / DVDs				
Security	Computers require password on start-up				
	Username and passwords are unique to each individual person (i.e. not shared)				
	UPS for desktop computers/servers				

OWN SOPS/GUIDELINES

For which of the following does the site have its own SOPs/guidelines?		Yes	No	N/A	Comment
Systems	Setup/Installation				
	System Maintenance				
	Contingency plan (in event of failure of computerised system)				
Data Management	Database design				
	Data collection and handling				
	Data quality control/assurance				
	Data backup				
	Data recovery				
	Data transfer				
	Database lock & archiving				

SADACC SOPS/GUIDELINES

Which of the following SOPs have been implemented at the site, and which have the site personnel been trained on?	Trained & implemented	Trained	Not yet trained	Comment
Username & Password Management				
User Rights & Data Access Groups Management				
Data Recording				
Data Receipt, Entry and Validation				
Database Design & Build (incl. Longitudinal setup)				

Database Backup, Lock, Export & Archiving				
Data transfer from SPARCO sites to SADaCC				
Remote data quality monitoring				
Recruitment & Informed Consent Process				

What is working particularly well at your site?

What is not working so well at your site, if anything?

What can we do to help?

Forms & Documents 2

SOP: Informed Consent for a Research Database and Biobank on Sickle Cell Disease

Patient Identifier []

Dear Madam/Sir

Invitation to be part of the Sickle Cell Disease research database and biobank

We invite you to be part of a research project that is aimed at establishing a Sickle Cell Disease database and biobank in Africa. A database is a collection of health information from people with the same or a similar health condition. A biobank is a collection of biological samples collected from different persons and which may be used for doing research. Examples of biological samples include blood, urine, and other samples that when you go to the hospital, they collect to test and know what you are sick of.

We are inviting you to participate in this research database and biobank because you (or your child) or a member of your family have been diagnosed with Sickle Cell Disease.

This form explains what will be expected of you if you agree to participate in the the research project. Please read it carefully and ask questions about anything you do not understand about the project.

Why do we we want to establish a sickle cell disease database and biobank?

To do more powerful research, it is helpful for researchers to share information and samples. This is made easier if there is a database and a biobank that has data and samples that can be used for research. We are therefore collecting medical information and biological samples from people who have been diagnosed with Sickle Cell Disease or who has a family member that has been diagnosed with sickle cell disease. This information will be put into a database and the samples will be stored in a biobank, together with information and samples that we will collect as part of this project from different African countries.

The database and biobank that we will establish as part of this project will include medical information and biological samples, from people staying in different countries in Africa. The database will be located at the University of Cape Town, South Africa and the biobank will be located at a university or research institution in an African country.

These samples and data will be used by researchers to do studies that will lead to a better understanding of Sickle Cell Disease and related complications. This may in the future provide information on best ways for improving healthcare for patients with Sickle Cell Disease.

What will be expected of me if I agree to be part of this study?

If you agree that you (or your child) want to be part of this database and biobank, one of our researchers will ask you to give them permission to extract information related to sickle cell disease from your medical file or that of your child. The researcher will also ask you (or your child) information related to how you have been feeling and any new and how your doctor has been managing your health.

When your doctor sends you (or your child) to the laboratory to do tests as part of your normal care in the hospital, we will request that some of the samples collected in the laboratory be stored in the biobank.

What information will be collected from me or from my child?

Our research team will go through your (or your child's) medical records and collect information specific to Sickle Cell Disease. This will include: when you were (or your child) diagnosed, the diagnostic test that was used; the treatment you have been taking, how often you come to consult the clinic or hospital staff for a sickle cell disease related problem and the reasons for the consultation. We would also ask you for your date of birth, where you currently stay, your contact information, lifestyle, gender and ethnicity/tribe.

How will my samples and information be used?

Your samples and data may be used by researchers from different parts of the world to do research on sickle cell disease and other diseases.

Who will have access to my data and samples?

In the future, researchers from different parts of the world who want to do research on sickle cell disease may request to use the medical information and samples that we have stored in the database and biobank. If this happens, we will not contact you to request for further permission. However, when researchers request to use the data and samples stored in the database and biobank, they will be required to provide documents explaining how they plan to use the samples and the data. These documents will be reviewed by a special committee to check that your rights and privacy are protected. If the committee approves the study, the samples and/or data will be given to the researchers free of charge.

How will my confidentiality be maintained?

Your name (or that of your child), contact information and any other information that may identify you will not be stored in the database and will not be given to the researchers who will be using your samples or the information stored in the database.

A code that is unique to you will be created by the research team when you join the study. This code will be used in the database instead of your name or identifiable information. Only your local research team can link the code to your name.

When researchers request to use your sample or data, we will not share with them any information that can identify you such as your name or the name of your child or parents. However, there is a small chance that someone could find out your name but this is very unlikely to happen.

If we want to contact you for more information or to provide you with information related to your care or or to ask if you if you will be willing to participate in other research projects, we will contact your local research team, your local research team will then be able to contact you based on the information in your unique code.

How long will my sample and information be stored?

We would like to store your information and samples for a minimum of 4 years as long as funds allow.

What are the benefits of participating in this research project?

There are no direct benefits to you for participating in this research. However, your participation will help researchers working on Sickle Cell Disease to better understand the disease. This will someday lead to better ways of caring for people with sickle cell disease.

If we get any information that we think may improve your clinical care, we will inform your doctor. Your doctor will then contact you to ask if you would like to receive the information.

What are the risks of participating in this study?

Most of the time, when blood is collected for laboratory tests in the hospital, it is safe. However, people may sometimes feel a bit weak or get a small wound at the site of the body where the blood is collected. If this happens please let us know and you will be treated.

There is also small chance that someone may be able to identify you in the future. This is because we do not know which research will be done in the future using the samples or data that we have in the database and biobank. However, we will do our best to keep your information private. No information that may identify you will be given to the researchers who request to use your samples or data.

Will I receive any payment or compensation if I accept to be part of this study?

You or your child would not be paid for accepting that your sample of data be added to the database and biobank.

[Additional information will be required if sites compensate participants for their time and/transport]

Can I refuse to participate in this research ?

Yes. Participating in this research project is voluntary. It is up to you to decide whether or not you want to be part of this project. If you decide not to participate, you will still receive the same health care.

Can I withdraw from this study?

Yes. You may decide, at any time, to stop participating in this study. This will not affect the care that your doctor or health care worker will continue to give you. All you need to do is to inform our researchers that you do not want to continue being part of the study. Once you tell us that you do not want to continue being part of the study, no new information about you or your child will be added to the database for research purpose. However, all information that was already in the database before you withdrew from the study will continue being used for research purposes. Also, your samples or that of child that were already in use for research before you withdrew will not be destroyed but any of your sample that is left in the biobank will be removed and destroyed.

Will I receive results from research that uses my sample or data?

We do not plan to share individual results with you except if we get results that may help to improve the care that you are receiving from your doctors. If we have such information, we will send it to your doctors. Your doctors will then contact you to ask if you would like to receive the information.

General research findings will be shared with your community. This will be through Sickle Cell Disease patient groups.

Assent

Permission for your child (less than 18 years) to be part of this database and biobank will be given by you or a guardian. We will also explain the objectives and procedures to your child in a language that she or he can understand, where appropriate.

If your child turns 18 during the course of this project, the research team will ask him/her if s(he) is still willing to be part of the study. If they accept, we will request that they sign a new consent form. If they refuse to continue being part of the research project, no new information about them will be added again to the research database. However, all information about them that was in the database before they withdrew from the study will continue being used for research purposes. Also, samples of your child that were still in the biobank will be removed and destroyed.

Who are the researchers leading this study?

The researchers leading this project are: Professor Julie Makani (Tanzania). Professor Alex Osei Akoto (Ghana) and Professor Ambroise Wonkam (South Africa). The researchers coordinating the

project in country AAA is (are): Prof Obiagele Nnodu (Nigeria), Dr Vivian Paintsil and Emmanuel Balandya (Tanzania). The project is currently funded by the National Institute of Health in the United States of America and is presently taking place in Ghana, Nigeria and Tanzania

Contact information

If you have any questions about this research project, please contact [local PI].....

If you have any questions about your rights as a research participant or you want to make complaints about your experience being part of the research study, please contact..... *[Sites to include contact information of the ethics committee that approved the project]*

Permission to be part of the research database and biobank

I have read this form and the information in this form has been explained to me in a language that I understand. I have been given the opportunity to ask questions and to consider whether I would like to be part of the research database and biobank.

I voluntarily agree to be part of the research database and biobank on Sickle Cell Disease and/or to allow my child (if you are the parent or guardian) to be part of the database and biobank. My data and sample and/or that of my child may be used in future research studies as explained in this consent form.

Please tick the appropriate box

Yes, I agree that my samples and data or that of my child should be shared with researchers working on Sickle Cell Disease in other countries through the Sickle Cell Disease Coordinating center (SADaCC) at the University of Cape Town, South Africa.

Yes, I will be happy to be contacted in the future to ask if I would like to be part of other research projects on Sickle Cell Disease.

No, do not contact me to ask if I will like to be part of other sickle cell disease projects

*Signature of participant or Parent/Guardian
(In the case of a minor)*

Date (DD/MM/YYYY)

OR

Thumb Print

Date (DD/MM/YYYY)

Name of participant or Parent/Guardian

Administrative use only (To be completed by the research team)

Certification of research staff

I certify that the nature and purpose, the potential benefits, and possible risks associated with submitting data and samples to the research database or biobank have been explained to the above individual and that all questions about this project that they asked have been answered.

Name and signature of research staff

Date

Complete this section in the case where an interpreter was used to explain the purpose of the research database and biobank to a participant because he/she cannot read/comprehend the information presented in this consent form

Name and signature of interpreter

Date

Forms & Documents 3

SOP: Assent Form (7-18 years)

Patient Identifier []

Invitation to be part of the Sickle Cell Disease research database and biobank

We invite you to be part of the Sickle Cell Disease research database and biobank, where we will store samples and information, collected from different people with sickle cell disease. A database is a collection of health information from people with the same or a similar health condition. A biobank is a collection of samples such as blood and urine which has collected from different persons. The information and samples may be used by researchers to do research.

We are presently doing this project in Ghana, Nigeria and Tanzania but we also plan to include other countries in the future.

Who are the people that are responsible for this project?

In country XXX, this project is led by XXX Other researchers who are involved in the project are: Prof Julie Makani (Tanzania) and Prof Ambroise Wonkam (South Africa).

What will happen if I accept to be part of this database and biobank?

If you accept to be part of this project, a doctor or a nurse or a researcher will look at your medical file and take out information about sickle cell disease. They will also store some of your blood samples. This sample and information will be shared with researchers in different parts of the world who are doing research on sickle cell disease. We hope that this will someday lead to better ways of caring for people with sickle cell disease

Your name and any other information that may identify you will not be given to researchers who will be using your samples and information.

We will inform your parents or guardian about this research project. If you accept to be part of this database and biobank, we will also ask your parents or guardian to give us their permission for you to be part of the research project.

Can I refuse to be part of this database and biobank?

Yes, you can refuse to be part of this project even if your parents/guardian has accepted that you can join the project. The doctors and nurses will continue to care for you as they always do when you come to the hospital.

Can I ask a question?

Yes, please ask the doctor or the nurses or the researcher any question about this project that you do not understand or want to know more.

What should I do if I want to be part of the research project?

If you decide you want to be part of this database and biobank, please write your name below.

Add Site PI and ethics contact numbers should be provided instead of just the names.

I, _____, want to be of this database and biobank

(Sign your name here/Thumbprint)

(Write the date)

Administrative use only (To be completed by the research team)

Certification of research staff (Tick all that apply)

I certify that I have explained the purpose of the research database or biobank to the child whose name is written above and that I have answered the questions that they asked.

I certify that informed consent has been obtained from their parent/guardian

Name and signature of research staff

Date

Version No.	Date	Internal Reviewer(s)	Author	Details of changes

Data and Database Management 4

SOP: Data Access & Sharing Policy - SickleInAfrica

PURPOSE

To describe the policy for accessing and using data in the SickleInAfrica registry.

SCOPE

This policy serves to guide members of the SickleInAfrica consortium on access and use of data in the SickleInAfrica registry.

GLOSSARY/DEFINITIONS

SADaCC	Sickle Africa Data Coordinating Centre at the University of Cape Town
SOP	Standard Operating Procedure: Detailed written instructions to achieve uniformity of the performance of a specific function.

PRINCIPLES

In developing this policy, the SickleInAfrica aims to:

- Maximise the availability of data on sickle cell disease in Africa.
- Promote the sharing of data on sickle cell disease whilst protecting the rights and privacy of persons in each country enrolled in the SickleInAfrica registry and should follow the rules and regulations stipulated in each country.
- Recognize the scientific contribution of researchers and healthcare personnel who collected and or generated data in the SickleInAfrica registry.
- Promote deposition of research data in public data repositories.

TYPES OF DATA

Two types

- **Raw data:** This is the raw data of individual registry participants and which can be directly attributed to them. For example: Socio-demographic and clinical data.
- **Research Data:** This includes data derived from the registry data but which can no longer be attributable to any individual participant.

DATA MANAGEMENT

Principal Investigators (PIs) of SickleInAfrica data shall ensure that informed consent used for the enrolment of registry participants allows for the sharing of data (both raw and research data). Equally, research ethics clearance for all studies should allow

for data sharing. It is critical that the consent form clearly outlines the intended use of the data and implications of data sharing, the IC should be comprehensible enough to the patient. It is the duty of the researcher to make sure that the descriptions of the IC are understood by the patient.

In designing research studies, site PIs must ensure that there are resources in place to protect the confidentiality of patients and to minimise the risk of identification of participants. Data from all SickleInAfrica sites must be anonymised before it is shared with other SickleInAfrica site PIs or third parties.

All SickleInAfrica sites must use a data management plan that ensures that the data meets the FAIR (Findability, Accessibility, Interoperability, reusability) principles. This will allow for wider use of SickleInAfrica data for healthcare and health research.

SADaCC will provide stewardship for the aggregated data stored in the SickleInAfrica registry. SADaCC therefore has the responsibility of ensuring that there are mechanisms in place to protect against potential breaches of privacy. SADaCC will also ensure that the data is used in ways that promote the research interests of all SickleInAfrica researchers.

SADaCC, on behalf of the SickleInAfrica consortium, shall manage access requests from third parties. SADaCC therefore has the responsibility of ensuring the ethical integrity and scientific merit of proposals seeking to use data from the SickleInAfrica registry.

SADaCC and the SPARCO-hub, with the support of all site PIs, will build the research capacity of SPARCO staff in data management and analysis, this will be done through site visits and conducting a series of training collectively known as “Data Stewardship practices in clinical research and translation”. This will ensure that the different SPARCO sites have the required capacity to analyse datasets from their sites, as well as the aggregated data from all sites.

USES OF DATA

Use of SickleInAfrica registry data is limited to projects on Sickle cell disease, related haemoglobinopathies and other co-morbidities related to sickle cell disease.

Data from SickleInAfrica projects will only be used for general research, health innovation, training or for advocacy purposes related to health and health research, on condition that:

- a) Participants are re-consented with a consent form that allows to select appropriate options

ACCESS TO DATA

SickleInAfrica distinguishes between internal users and external data users with respect to access and secondary use of data.

- Internal users are members of the SickleInAfrica Consortium and include: Project PIs of the different sites and their co-investigators, researchers, students and staff members as named on the grant proposal. SADaCC will maintain and update a list of all researchers and students at the different SPARCO and SADaCC sites.
- External users (or third parties) are non-SickleInAfrica consortium members. This includes collaborators of SickleInAfrica PIs who are not directly involved in the SickleInAfrica consortium.

Members of the consortium shall request access to raw data from a site, via the PI of that site. SADaCC will only grant access to the data when they receive written permission from the site PI. This written permission will be logged by SADaCC.

Data and datasets (including details of the analysis plan) from SickleInAfrica projects may also be shared with third parties who request to use the data for research, healthcare, innovation and health-related advocacy. Data will only be shared with third parties after a data sharing agreement (DSA) has been signed between SickleInAfrica and the potential data recipient. The signed DSA will be logged by SADaCC. [\[Insert link to sample for data sharing agreement\]](#)

SickleInAfrica shall establish a Data Access Committee (DAC) at least a year before the end of Phase 1 of the SickleInAfrica Project. The DAC will review requests for access to data in the SickleInAfrica registry by third parties. Members of the DAC will include SickleInAfrica project PIs or their nominees; two representatives of patient support groups; and other non-SickleInAfrica members. The Non SickleInAfrica members will be nominated by SickleInAfrica consortium members.

Non-SickleInAfrica Consortium members (Third parties) will request access to SickleInAfrica data through SADaCC. Once a request for access is received by SADaCC, PIs of the different sites will be informed of the data request. Thereafter, the request will be submitted to the SickleInAfrica DAC

[\[Insert link to application form for data access-to be developed\]](#)

Before data is made accessible to third parties, there must be a signed data sharing agreement between SADaCC and the third party user.

[\[Insert link to data-sharing agreement-Draft available\]](#)

SADaCC, on behalf of SickleInAfrica, will sign all the DSAs between the SickleInAfrica and the third party that has been granted access to SickleInAfrica data. SADaCC will inform all SickleInAfrica PIs of all DSA that it signs on behalf of SickleInAfrica. SADaCC also will keep an inventory of all DSAs via the SickleInAfrica website.

[\[Insert link SickleInAfrica website\]](#)

The basic descriptive and aggregate summary information of data generated by each SickleInAfrica site will be made publicly available through the SickleInAfrica website using data visualization tools.

DATA ANONYMIZATION

At any one time, data in the SickleInAfrica repository will be anonymized and de-identified. However an identifier will be kept by the data generating site who may provide back linking when the need arises.

INTERNAL ACCESS PROCEDURES

SickleInAfrica consortium members who wish to use data collected from the different study sites may contact the site PIs directly. SADaCC will only grant access to the data once they receive written permission from the site PIs

Third parties who wish to use data from the SickleInAfrica registry will make a written request for data access, via SADaCC. SADaCC will inform the SickleInAfrica PIs of the request. In cases where a Site PI expresses an interest to collaborate on the project, SADaCC and the SPARCO Hub will facilitate discussions between the site-PIs and the third party. Collaboration with SickleInAfrica investigators is not a requirement for granting access to the data. However, projects that involve collaborations with SickleInAfrica investigators will be prioritised.

Written request for third party access should include:

- An abstract of the study and how the requested data will be used;
- The methods that the data requestor plans to use to ensure data security and to safeguard the confidentiality of registry participants;
- Statement on project funding;
- Statement indicating whether they have obtained REC approval for the proposed study.
- A detailed description of variables for which the data required and;
- The timing/duration or specified events for which the data requested must comply to.
- Affiliations of the requestor and their collaborators on the project.
- **At any time, researchers should be encouraged to demonstrate a research question first, instead of first looking at the data structure.**

Upon receipt of a request for data access, SADaCC will perform an initial review to check if the study proposal meets access requirements as set out in this policy. Applications that meet access requirements will be sent to the DAC for scientific and ethics review. In cases where the application does not meet access requirements, SADaCC will communicate the details to the applicant.

A researcher or institution that is granted access to SickleInAfrica data may designate specific persons who will have access to the data. The primary data recipient shall keep a list of people who have access to the data, and their respective affiliations. This list, and any changes to it, should be made available to SADaCC, by the data requestor.

The DSA between SickleInAfrica investigators and the data recipient will only be signed after the request for data access has been approved by the DAC and upon receipt of research ethics approval for the proposed study.

Data will be provided in a format mutually agreed upon by the Recipient and SickleInAfrica. Any extra cost to data extraction, formatting, encryption and transfer would be borne by the data recipient.

In the case where a request for data access is not approved, the DAC shall provide reasons for their decision. SADaCC will communicate this information to the applicant. Applicants to the The applicant may appeal the decision. Appeals will follow the same procedures for new request including a reason for the appeal.

Researchers or institutions who plan to use data in the SickleInAfrica registry and need a letter of support for funding applications, may contact SADaCC to discuss their intended use of the data. A request for a support letter from the SickleInAfrica consortium should include the title of the proposed study; an abstract describing the proposed study; how they intend to use the data; and a list of all researchers involved in the project (and their affiliations). If SADaCC, in consultation with SPARCO principal investigators, determines that the proposed use of the data is in line with the goals of the SickleInAfrica consortium, SADaCC, on behalf of SickleInAfrica, will provide a letter of support to the potential data recipient. A letter of support does not mean access to SickleInAfrica data has been approved. Once funding is obtained, the potential recipient would have to apply for access to data as described above.

ACCESS TO RE-IDENTIFIABLE DATA AND CONTACT WITH REGISTRY PARTICIPANTS

Where necessary, individuals may request access to re-identifiable data. Data recipients who wish to access re-identifiable data for the purposes of enrolling

research participants into a research study, or for the provision of better healthcare, must first register their interest with SADaCC. Researchers or innovators are not permitted to recruit registry participants and/or their families without first receiving written permission from SADaCC.

Contact with registry participants will require collaboration with SickleInAfrica investigators at the sites where participants will be enrolled for the proposed study.

LIMITS ON DATA SHARING

Data will only be shared with third parties after a period of one year embargo following deposition of the site data at SADaCC. This is to allow SickleInAfrica researchers to analyse data generated from their projects and to publish this data. This exclusive publication period will be managed by SADaCC and the SPARCO Hub. Publication exclusivity extends to: meeting/conference abstracts or proceedings, presentations at conferences, articles in peer reviewed journals and publicly accessible electronic submissions (e.g. websites, web blogs).

PUBLICATIONS AND AUTHORSHIP

The SickleInAfrica consortium should be acknowledged in all publications and scientific outputs that report information based on data in the SickleInAfrica registry.

[Include sample acknowledgment statement- e.g The data used in this study was obtained from the SickleInAfrica Registry. SickleInAfrica Registry received funding from the National Heart, Lung and Blood Institute of the National Institutes of Health (Award Number.....). The content of this paper does not necessarily represent the official views of the National Institutes of Health or the SickleInAfrica consortium]

INTELLECTUAL PROPERTIES AND PATENTS

The SickleInAfrica is in support of initiatives aimed at developing interventions that may improve healthcare for persons with SCDs. Any IPs, patents or royalties arising from SickleInAfrica data should be equally divided among the SickleInAfrica consortium and the innovator. IP and patents will only be acceptable if they will not impede further research. SickleInAfrica investigators and third parties have the responsibility to honor their respective institutions in the acquired IPs.

BENEFIT SHARING

SickleInAfrica will adopt a power sharing approach to benefit sharing. This is because the current use of the registry may not provide any direct benefit to study participants in the short term. SickleInAfrica would engage SCD patient support groups in the countries where enrolment is taking place. The patient support groups will be empowered to effectively participate in decisions on data access and use. The essence is to give study communities a voice in decision making and to have them as partners in the SickleInAfrica registry.

SickleInAfrica will support public engagement as part of its core activities. Researchers who are given access to data in the SickleInAfrica registry must make a commitment to provide SADaCC with a lay summary summary of the project and the project outcomes. This lay summary will be made available on the SickleInAfrica website as part of its public engagement activities.

For each approved request for access to data, SADaCC will check if the requestor may derive any form of commercial gains from the use of the data or develop a new intervention, SickleInAfrica shall assess the potential benefits and risks of this activity to registry participants or their communities. SADaCC will then make a decision on whether or not the DSA should be modified to include specific benefit sharing arrangements. Such arrangements will be mutually agreed upon by SickleInAfrica and the data recipient.

ACCOUNTABILITY

SADaCC, on behalf of the SickleInAfrica consortium, will consistently monitor and evaluate that data use is in line with the goals of the SickleInAfrica consortium and the consent provided by registry participants. Monitoring and evaluation activities will also aim at providing a situation analysis of SickleInAfrica at the beginning of the project and how the consortium is working towards achieving its goals. This may include outputs related to: number of approved requests for secondary use of data/samples; research that has been carried out using samples; number of research publications; authorship of papers by SickleInAfrica consortium members; and completed training or research fellowships.

SADaCC will make publicly available on its website, or other media outlets, a scientific and lay summary of all ongoing projects in SickleInAfrica, as well as projects that are using data from the SickleInAfrica registry.

SADaCC will monitor publications, patents and IPs arising from the use of data from SickleInAfrica. This information will be made available on the SickleInAfrica website.

For the benefit of registry participants and the general public, a lay summary of project outcomes and progress will be made available to study communities and support groups through the SickleInAfrica website, newsletters or other media, as deemed appropriate by the support groups and the SickleInAfrica consortium. For this purpose, all users of data in the SickleInAfrica registry shall provide a lay summary of their projects and where available, the project outcomes.

Researchers or entities who use data from the SickleInAfrica registry will provide SickleInAfrica with a list of publications and other outputs arising from the use of the data. This information should be made available to SADaCC once it is available.

SickleInAfrica will periodically, but continuously, engage patient support groups in the participating countries. SickleInAfrica will provide them with information on the registry and will make efforts to address concerns they may have in relation to: The consortium's policies and processes; and the use of data in the SickleInAfrica registry.

ARCHIVING OF DATASETS

Where the SickleInAfrica consortium is unable to continue administration and management of data from SickleInAfrica projects, SADaCC, after consulting with all SickleInAfrica partners, will submit data to a public repository and/or the institutional databases of participating SickleInAfrica institutions, making sure that such deposited data is compliant with the FAIR data principles. The data will therefore be FAIRfied before archiving to make sure that in the future it will be accessible and re-usable.

RESOLVING DISPUTES ARISING FROM MISUNDERSTANDINGS

Refer to the Conflict Resolution document

[Insert link]

REFERENCES

1. Wilkinson MD et al. The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data. 2016;3:160018.<https://doi.org/10.1038/sdata.2016.18>

Version No.	Date	Internal Reviewer(s)	Author	Details of changes
1		Daima, Buikini, Furahini Tluway	Syntia	n/a (first version)

Administration, Coordination and Management 6

SOP: Conflict Resolution Policy

PURPOSE:

To describe procedures to be taken to resolve conflict involving individuals supported by the SPARCO and SADaCC grants.

SCOPE

GLOSSARY/DEFINITIONS

SADaCC Sickle Africa Data Coordinating Centre at the University of Cape Town

SPARCO **XXX**

SOP Standard Operating Procedure: Detailed written instructions to achieve uniformity of the performance of a specific function.

UCT University of Cape Town

RESPONSIBILITIES & PROCEDURES

This policy applies to all investigators, mentors, post-doctoral fellows, graduate/undergraduate students, and staff supported by the SPARCO and SADaCC grants. This policy is not intended to supersede the individual's institutional conflict resolution policy. The best practice is to prevent conflicts or to prevent escalation of disagreements to conflicts.

RECOMMENDED BEHAVIOURS TO REDUCE DISAGREEMENTS AND PREVENT CONFLICTS INVOLVE:

Routine communication between partners.

- True appreciation of each partner's contribution to the project.
- Recognition and respect of the role of the partner at his/her home institution and therefore, patience for competing demands on his/her time.
- Communication at the first twinge or suspicion of a misunderstanding on any issue or a sense of non-alignment with the project goals and objectives.
- Revisiting the proposal to review what was proposed; this document should be used to guide the communication between parties.

Situations where conflicts or disagreements could arise might involve:

- Authorship on manuscripts or patents or other intellectual property claims.
- Expectations of who is responsible for project activities, or experimental results
- Expectations of who is responsible for project reports or presentations.
- Decisions involving funding and budgeting.

To prevent these types of misunderstandings from escalating into disagreements and conflicts, please remember to communicate with all parties and negotiate fairly.

INFORMAL AND FORMAL RESOLUTION OF CONFLICTS

I. The Steering Committee (SC) encourages the active resolution of conflicts in order to maintain and promote a fair and productive environment for all participants. Participants are encouraged to maintain and utilize open communications to promote and maintain a civil and humane resolution of any and all conflicts. The SC encourages the resolution of conflict at the earliest opportunity, or the lowest step in whatever dispute resolution process is used. Both formal procedures and informal conflict resolution mechanisms are provided. Participants are encouraged to use informal resolution whenever appropriate. Together, the resources listed below are intended to provide fair, thoughtful, and effective means to manage and/or resolve conflict situations.

II. We define conflict as **a disagreement through which the parties involved perceive a threat to their needs, interests or concerns.**

III. Informal Resolution of Conflicts

We recognize that resolving a conflict requires personal integrity on the part of all parties involved. It takes courage to honestly and clearly articulate needs and to sit down and listen to alternative/adversarial perspectives. It takes courage to look at one's own role in the dispute, and to approach others with a sense of empathy, openness, and respect for their perspective. The following general suggestions may help to informally resolve conflicts with co-workers, collaborators, and/or students constructively and respectfully.

A. Treat the other person with respect

Although respecting the other person during a conflict is challenging, one must try. Words of disrespect block communication.

B. Confront the problem

A time and place to discuss the conflict with the other person should be identified. The best time is when the parties are not arguing or angry. The place should be comfortable and away from either party's "turf."

C. Define the conflict

The conflict should be described in clear, concrete terms. Specificity of the “who, what, when, where, and why” of the situation should occur. Identify the problems for discussion, without focusing on the individuals. Behaviors, feelings, consequences, and desired changes should be discussed.

D. Communicate understanding

Listening to understand the other person's feelings, needs, and so forth is needed. The parties should be encouraged to step back and try to imagine how the other person sees things. After discussing the issues, the parties should explain how they see the problem after discussion. It is especially important to discuss any changes in how the parties feel about the issues.

E. Explore alternative solutions

Alternative solutions should be offered. The consequences of each solution should be examined.

F. Agree on the most workable solution

A solution that all parties involved can understand and can live with should be adopted. Win-win solutions are most appropriate. Commitment to resolving the conflict is required.

G. Evaluate after time

After a certain time has passed (specify length of time), the parties should evaluate how well the solution is working. Adjustment of the resolution should occur if and when necessary.

IV. Formal Resolution of Conflicts

When informal conflict resolution is ineffective (Any length of time to determine if the informal approach has failed?), formal mechanisms to resolution may be required. These steps will be used to formally resolve conflicts.

A. SPAN, SADaCC & SPARCO Directors/ Chairs

The Chairs will form the first point of contact in the formal conflict resolution process. Written statements describing the conflict issues will be collected from all participants. The Chairs will suggest possible strategies for resolution.

B. Steering Committee

Conflicts that cannot be resolved among the affected participants through the SPAN, SADaCC & SPARCO Directors/Chairs will be referred to the SC. The committee will consider the documentation, discuss the issues with the participants either separately or jointly, and may initiate other inquiries or request additional information to facilitate comprehension of the situation and identify recommendations for resolution. The SC will discuss the issues with the participants either separately or jointly and will make a

recommendation for resolution. The SC recommendation will be considered binding for the participants involved in the conflict.

In the event that the Directors/chairs are involved in the said conflict, 3 members from the Advisory body will be co-opted in the SC to help resolve the conflict. The NHLBI will be the last resort.

NIH Dispute Resolution:

Any disagreement that may arise in scientific/programmatic matters (within the scope of the award); between award recipients and the NHLBI may be brought to arbitration. An arbitration panel will be composed of three members--one selected by the Steering Committee (with the NHLBI member absent from the discussion) or by the individual awardee in the event of an individual disagreement, a second member selected by NHLBI, and the third member selected by the two prior members. This special arbitration procedure in no way affects the awardees' right to appeal an adverse action that is otherwise appealable in accordance with the PHS regulations at 42 CFR part 50, Subpart D and DHHS regulation at 45 CFR Part 16.

Source: <https://www.nminbre.org/index.php/conflict-resolution-policy>

- Document Name_Draft version 0.3 (21 Feb 2017) – reworked by author
- Document Name_Draft version 0.4 (28 Feb 2017) – by external reviewer
- Document Name_Published version 1.0 (03 Mar 2017) – by authorizer

Version No.	Date	Authorizer	External Reviewer	Internal Reviewer	Authors	Details of changes
1.0					Furahini Chinenere & Vicky Nembaware	n/a (first version)

Administration, Coordination and Management 7

SOP: Authorship Policy

PURPOSE:

To describe procedures for deciding upon authorship of publications.

SCOPE

SickleInAfrica Consortium aims to publish its research findings widely and without delay in order to optimize appropriate translation that will contribute towards improving the health of Africans and the world in general. This Publications Policy will guide this overarching aim by clarifying the specific objectives of the Consortium and the processes to be followed in publication of its research. Implementing and interpreting this policy is the primary responsibility of the Steering Committee (SC) of SickleInAfrica. The SC will be advised by a Publications Group, when necessary. The SC will use a Manuscript Concept Document as a tool to promote publication, encourage collaboration, ensure transparency and avoid unnecessary duplication. This will also promote fair and equitable attribution of credits in authorship and acknowledgements to investigators, study participants and the funding organizations.

GLOSSARY/DEFINITIONS

SADaCC Sickle Africa Data Coordinating Centre at the University of Cape Town

SOP Standard Operating Procedure: Detailed written instructions to achieve uniformity of the performance of a specific function.

UCT University of Cape Town

OBJECTIVES

The Publications Policy aims to promote and facilitate the following objectives of the SickleInAfrica Consortium:

- Timely dissemination of findings
- Fair and appropriate authorship and attribution of credits
- Equity in opportunity to earn authorship in consortium publications
- Transparency of planned publications to avoid overlap and encourage collaboration
- Encouragement of African-led publications without sacrificing the principles governing authorship practices
- Encouragement of open access publication
- Provision of opportunities for junior investigators on the continent

- Assurance of appropriate communication of publications

Promotion of publications as a vehicle for implementing the data access policy of SickleInAfrica. SickleInAfrica supports and encourages a variety of types of publication. All publications involving grantees (from one PI group or several) and the use of SickleInAfrica resources will have the key authors listed and are expected to include the “The SickleInAfrica Consortium” as an author at the end of the list of authors. Details for each category of publication are given in Table 1. The intention is to have a mechanism for tracking publications emanating from the Consortium, to assess the impact of the Consortium’s research for reporting purposes and to use this information when applying for additional funding.

The Manuscript Concept Document will be used as a tool for researchers to indicate their intention to publish specific manuscripts. Its use will promote transparency and information sharing within the consortium, and will help to avoid overlap while promoting collaboration. The intention is that the Manuscript Concept Document (MCD) will encourage collaboration by allowing recognition of synergistic opportunities without being prescriptive. It will also form part of a quality assurance process to ensure valid scientific outputs for the Consortium. Participants in the SickleInAfrica Consortium will complete the MCD as suggested in Table 1. See Appendix 1 for the MCD template.

Publications arising from collaborations between individuals or groups of SickleInAfrica Consortium members and external partners will follow the process outlined in this policy. The criteria that will be considered are the scientific merit of the proposed publication, whether the appropriate attribution and acknowledgement is given and that the publication will not bring the Consortium into disrepute.

Table 1: Categories of Publication with guidelines for authorship and the use of the MCD

Publication Category	Authorship	MCD	Manuscript Concept Document
SickleInAfrica	1. Marker paper: Author shall be written as “The Sickle In Africa Consortium” with the writing group and all members of the Sickle In Africa Consortium listed alphabetically by groups and then by surnames at the end of the manuscript.	Yes	Submit to Coordinating Centre for review by the SC prior to submission

	<p>2. Consortium-wide: Members of the research and writing groups are listed as authors followed by "and The Sickle In Africa Consortium". All members of the Sickle In Africa Consortium involved with the publication listed at the end of the manuscript according to groups and surnames, alphabetically.</p> <p>Sickle In Africa Consortium Working Group: Authors followed by "and The Sickle In Africa Consortium". In a footnote or at the end of the manuscript, refer to the working group and other members of that group who did not meet the criteria for authorship.</p>		
Multiple grantee group	<p>Authors followed by "as members of The SickleInAfrica Consortium". As more than one Grantee Group is involved, and many of their group members have participated in generating the resources, there will be an expansion of "The SickleInAfrica Consortium" either as a footnote or at the end of the manuscript to include the other SickleInAfrica members who have been involved with the study, but do not meet the requirements for authorship.</p>	Yes	Submit to Coordinating Centre following publication
Single grantee group or intra-grantee group (e.g. single site)	<p>Authors followed by "for xxx Research Group, as members of The SickleInAfrica Consortium". There is no need to expand "the SickleInAfrica Consortium" as all relevant authors will be included and an acknowledgement could be used to cover additional group members who have contributed to generating resources, but were not involved in this specific study (i.e. do not qualify for authorship).</p>	No	Submit to Coordinating Centre following publication

<p>Consortium Collaborations (including external partners)</p>	<p>SickleInAfrica Consortium members collaborating with external partners on a research project will ensure that an MCD is duly completed and submitted once a publication is mooted. The publication guidelines will be as above, either for a “multiple grantee groups” or a “single grantee group”, as appropriate</p>	<p>Yes</p>	<p>Yes (multiple grantee group) No (single grantee group)</p>
--	---	------------	--

AUTHORSHIP CRITERIA

Authors of all papers should meet the common criteria for authorship used by journals. Specifically, the individual:

1. must have made a significant contribution of scholarly effort in conception, design, data generation, or analysis;
2. should have been involved in drafting reviewing and/or revising of the manuscript; and
3. must have approved the manuscript prior to publication.

The accepted principles governing assignment of Lead (First) Author and Senior (Last) Author should be followed while promoting African-led publications and opportunities for junior investigators. PIs will follow rules of “fairness” in authorship attribution when selecting members for inclusion of authors in the expanded group authorship sections, being mindful of the contributions of all their group members. This will be left to the discretion of the PIs, but they are encouraged specifically to facilitate contributions from young scientists and inclusion of scientists dedicated to this project and funded through SickleInAfrica. It is important to help build the careers and reputations of our talented young scientists and thereby to expose them to future opportunities.

The SC will be assisted by a Publications Group when there is a dispute regarding authorship that cannot be resolved by the PIs and/or Steering Committee, and when negotiations are needed between PIs in cases where there is significant overlap between proposed research publications.

PROCESS

The following operational process is envisaged to streamline the objective of timely publication and maximizing the research outputs (read in conjunction with Table 1):

- a. Authors identify a study which is likely to result in a publication and complete the MCD template (downloadable from www.SickleInAfrica.org) with as much detail as possible. <https://www.sickleinafrica.org/node/182>
- b. The MCD is submitted to SADaCC via an internet-based process.
- c. Receipt is acknowledged and the MCD is sent to the Steering Committee for comment.
- d. The Steering Committee acknowledges and a member is nominated to provide feedback to the submitting author within 10 days of receipt.
- e. Possible outcomes:
 1. Steering Committee approves.
 2. Steering Committee recommends revision and/or possible collaboration(s) and the researchers agree, a revised MCD is submitted.
 3. Steering Committee refers MCD to the Publications Group for comment and advice.
 4. If necessary, the Publications Group becomes involved, and following dispute resolution, the revised MCD is submitted.
 5. Steering committee rejects the proposed publication

APPENDIX A: MANUSCRIPT CONCEPT SHEET

Project/Manuscript Concept Sheet

Submission Date	
Project Title	
Tentative Lead Investigator (first author)*	
Tentative Senior Author (last author)*	This is not always straightforward for the kind of work we are doing. This is a proposal, not final point
All other authors (actively participating in manuscript creation)*	
Sites Contributing Data	

Background / Significance	
Outline of Project <ul style="list-style-type: none"> • Main Manuscript Question • Data to be Used • Analysis Plan 	
Identification of possible overlapping Sickle In Africa analyses and manuscripts	
Target Journal	
Milestones**	

* Please also list authors' project sites

** This section should include: Timeline for completion of project, including approval, project duration, first and second draft of the paper and submission.

Version No.	Date	Authorizer	External Reviewer	Internal Reviewer	Authors	Details of changes
1.0					Furahini Tluway & Vicky Nembaware	n/a (first version)

Research and Ethical, Legal and Social Implications 2

SOP: Development of Research Projects: Manuscript Concept Note

PURPOSE

To describe the procedure for developing manuscript concept note as a tool for manuscript development.

SCOPE

This procedure applies to all scientists/researchers under Sickle Pan African Research Consortium (SPARCO) and the Sickle Africa Data Coordinating Centre (SADaCC).

GLOSSARY/DEFINITIONS

SPARCO Sickle Pan African Research Consortium

SADaCC Sickle Africa Data Coordinating Centre at the University of Cape Town

SOP Standard Operating Procedure: Detailed written instructions to achieve uniformity of the performance of a specific function.

RESPONSIBILITIES & PROCEDURES

1. Determine the title of the manuscript.

The leading author and co-authors must agree on the title of the manuscript.

2. Authorship.

The responsible author should come up with a list of authors by following the SPARCO/SADaCC and NIH authorship guidelines.

3. Abstract.

The authors should prepare an abstract of a maximum of 250 words which highlights the background, objectives, methods, results and conclusion of the study.

4. Selection of Journals

The authors should list at least five journals in a preference order in which the manuscript will be submitted for publication.

5. Proposed milestone for manuscript development

The responsible author in agreement with other authors should develop a milestone and working schedule for the development of the manuscript. This should reflect specific activities, responsible person/people and dates.

Version	Date	Authorizer	External Reviewer	Internal	Author	Details	of
---------	------	------------	-------------------	----------	--------	---------	----

No.				Reviewer		changes
1.0				Mario Jonas	Siana Nkya	n/a (first version)

Research and Ethical, Legal and Social Implications 3

SOP: Access to Research Resources

PURPOSE

To set minimum conditions for access to research resources within the consortium.

SCOPE

This procedure applies to research resources within the consortium sites. It does not apply to research resources that are publicly available.

GLOSSARY/DEFINITIONS

SPARCO Sickle Pan African Research Consortium

SOP Standard Operating Procedure: Detailed written instructions to achieve uniformity of the performance of a specific function.

Research can be thought as critical and exhaustive investigation or experimentation, aiming essentially at discovering new facts and their correct interpretation, revising accepted conclusions, theories, or laws in the light of newly discovered facts, or the practical application of such new or revised conclusions, theories, or laws.

Research resources are different sources of information used when conducting a specific research. These sources are generally subdivided into primary and secondary sources.

Primary sources refer to actual events written or original literary or artistic works. They may be letters, official records, interviews, survey results, or data. These sources contain raw data and information, such as the original work of art or immediate impressions.

Secondary sources, on the other hand, are usually discussions, evaluations, syntheses, and analyses of primary- and secondary-source information.

RESPONSIBILITIES & PROCEDURES (INTERNAL COLLABORATIONS)

1. Review and approval of access to research resources

The Consortium PIs will review and approve or reject all requests from the research community for access to research resources. They will only consider requests for access to research resources submitted using the Research Resource Access Request Form. PIs may request modifications to request before approval. If necessary, the main PI will delay a decision pending a request for additional information from the submitter.

2. Verification of request for access to research resources

The main PI must ensure that all requests conform with Consortium policies and procedures including compliance with the terms of informed consent and any limitations stipulated by the submitting institution/investigators for each study. Researchers who qualify to submit a request are defined as scientists or medical professionals employed at, or legitimately affiliated with academic, non-profit or government institutions, or commercial companies.

3. Minimum conditions for access to research resources

Submission of Resource Access Request indicates that the Requester and his/her institution have agreed to the terms and conditions for resource use specified in the Resource Access Agreement for the SPARCo Resource Access Policy. This requires that Requesters, at a minimum, agree to:

1. Abide by the agreed upon research uses of the requested resource
2. Not distribute the resource in any form to any entity or individual other than his/her research staff or trainees or independent collaborating investigators listed in the resource access request
3. Adherence to SPARCo, institutional and ethics regulations as per the resource transfer agreement, unless a modification thereof is required, which would be evaluated on its own merit
4. Compliance with the rules that guide procedures for disposal of research resources following said research or study, where applicable
5. Keep resources secure, and
6. Acknowledge the SPARCo consortium as appropriate in publications and presentations.

Version No.	Date	Authorizer	External Reviewer	Internal Reviewer(s)	Author	Details of changes
01	08-11-2018				Gaston	n/a (first version)

Layperson Ontology 2.1

SOP: Incorporating Layperson Annotations

Note:

*This SOP is specifically for when the input ontology file does NOT have language tags for labels and descriptions of ontology terms. Use **Layperson Ontology 2.2 SOP** if the ontology input file already contains language tags for labels and descriptions of ontology terms.*

PURPOSE

To describe the procedure for incorporating layperson annotations (labels and descriptions) of ontology terms into the ontology file, using the *creating_layperson_ontology_files_NO_existing_lang_tags.py* python program.

SCOPE

This procedure is used by curators when producing ontology files that include layperson annotations.

GLOSSARY/DEFINITIONS

- Python** The programming language in which the program used in this SOP was written.
- SOP** Standard Operating Procedure: Detailed written instructions to achieve uniformity of the performance of a specific function.

SYSTEM REQUIREMENTS TO RUN PROGRAM

Instructions below are for Windows 10.

For the python program to run successfully, the following need to be installed:

1. Python.
Ensure that Python is setup correctly on your PC (Use **Program Running 1 SOP**)

REQUIRED INPUT FOR PROGRAM

1. Input files:
 - a. A .txt file containing the layperson annotations (for labels and descriptions of terms) of the ontology.
 - b. The owl file of the version of the ontology to which layperson annotations must be added. (Note: for this SOP, this file will NOT contain language tags for labels and descriptions of ontology terms)
2. These additional variables are required in the python program code before it is run:

- a. **layperson_annotations_file_cell1** The content of the first cell in the input file containing layperson annotations.
- b. **lang_tag** Specifies the language tag to be added to labels and descriptions.
- c. **iri_path** The specific IRI generic path that is used for terms in the ontology being translated.
- d. **layperson_tag** The tag to be added to layperson annotations.
- e. **outputfile1_name** name of rdf file created with all ontology terms and layperson annotations included where necessary.
- f. **outputfile2_name** name of rdf file created with only terms with layperson annotations

PREPARING INPUT FOR PROGRAM

1. Preparing input files and/or specifying them with relevant variables in the *creating_layperson_ontology_files_NO_existing_lang_tags.py* program code:
 - a. To prepare the .txt file with the layperson annotations of the ontology:
 - a) Open the shared Google spreadsheet containing layperson annotations (product of **Layperson Ontology 1 SOP**).
 - b) Download the file as a Microsoft Excel file (.xlsx).
 - c) Open the downloaded file and save it as a .txt (tab delimited) file.
 - d) Move the .txt file to the same location as the python program.
 - e) Ensure that the **layperson_annotations_file** variable in the python program code (row 135) points to the name (including file extension) of this text file.
 - b. Ensure that the **owl_file_path** variable in the python program code (row 136) points to the name (including file extension) of the file of the version of the ontology to which layperson annotations must be added.
2. Specifying additional required variables in the *creating_layperson_ontology_files_NO_existing_lang_tags.py* program code:
 - a. For the **layperson_annotations_file_cell1** variable in row 139, supply the content of the first cell in the input file containing layperson annotations.
 - b. For the **lang_tag** variable in row 140, supply the language tag to be added to labels and descriptions of all ontology terms (e.g. "en" for English)
 - c. For the **iri_path** variable in row 141, supply the specific IRI generic path that is used for terms in the ontology being translated.
 - d. For the **layperson_tag** variable in row 142, supply the tag to be added to layperson annotations (e.g. "layperson").

- e. For the **outputfile1_name** variable in row 143, supply the name of the rdf file that will be generated by the program and will contain all ontology terms, with layperson annotations included where necessary.
- f. For the **outputfile2_name** variable in row 144, supply the name of the rdf file that will be generated by the program and will contain only terms with layperson annotations.

RUNNING THE PROGRAM

Run the *creating_layperson_ontology_files_NO_existing_lang_tags.py* python program by following the procedure in the **Program Running 2 SOP**.

PROGRAM OUTPUT

Two files are generated:

1. An rdfxml file containing all terms in the ontology, including layperson annotations where necessary.
2. An rdfxml file containing only terms in the ontology that have layperson annotations.

USING THE OUTPUT FILES

The output files will appear in the same location as the python program file and can now be used as necessary, e.g. used to create a new project in Web Protégé or uploaded to a relevant repository.

Version No.	Date	Authorizer	External Reviewer	Internal Reviewer	Author	Details of changes
1.0					Jade Hotchkiss	n/a (first version)

Administration, Coordination and Management 8

SOP: Using Hootsuite

PURPOSE:

To describe the procedure of using Hootsuite to set up scheduled posts.

SCOPE:

This procedure is used to schedule timely posts on all SickleInAfrica social media platforms.

GLOSSARY/DEFINITIONS:

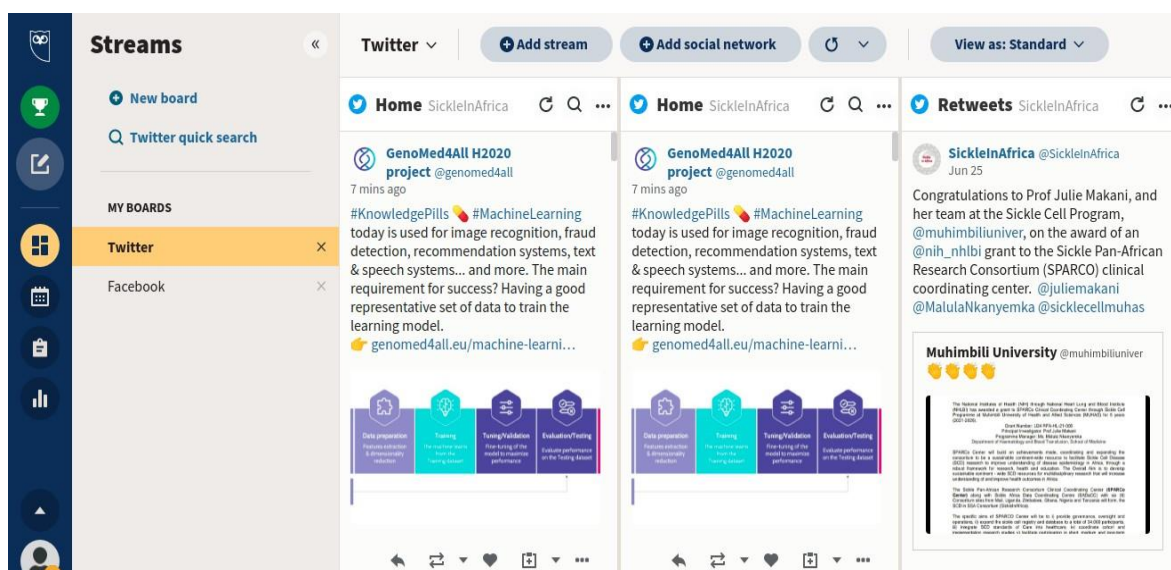
SOP Standard Operating Procedure: Detailed written instructions to achieve uniformity of the performance of a specific function.

Hootsuite Social media management tool.

Procedure:

1. Setting Up Hootsuite

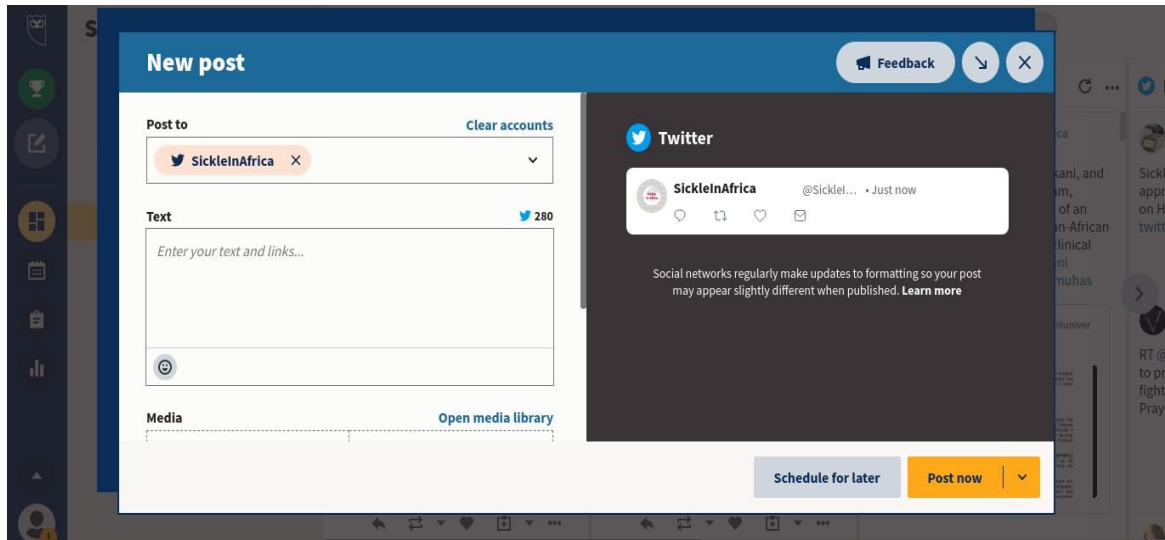
Create an account on the [Hootsuite](https://hootsuite.com) website and follow the instructions to add your social network platforms/accounts e.g. Twitter, Facebook, Instagram etc.



Setting Up Hootsuite:
Tutorial Video: [Setting up Hootsuite](#)

2. Scheduling posts

Schedule timely posts on your social media accounts.

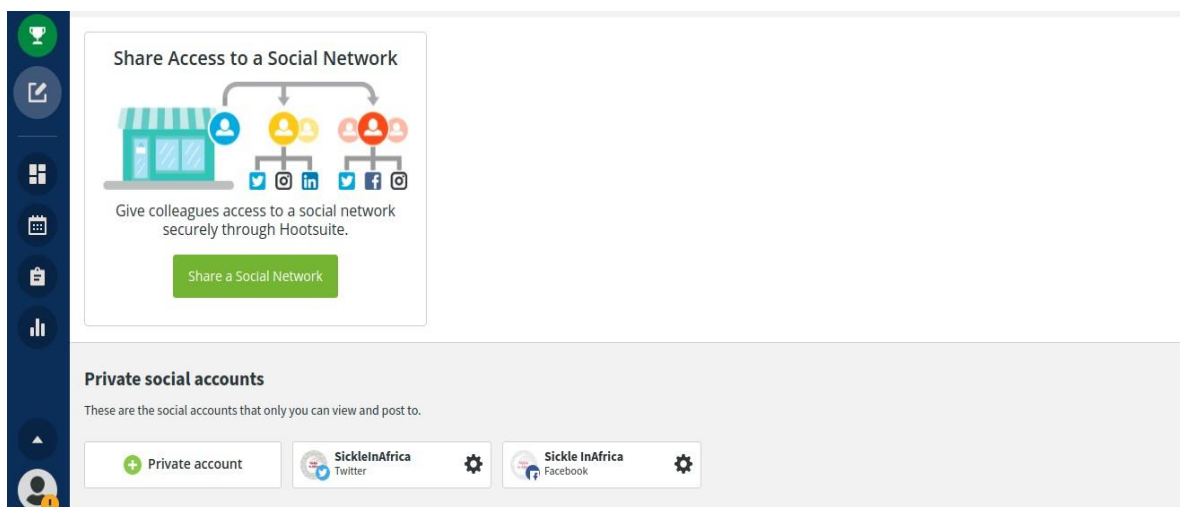


Scheduling posts:

Tutorial Video: [Scheduling posts](#)

3. Setting Up organization

Create an organization within Hootsuite account that controls user access and privileges.

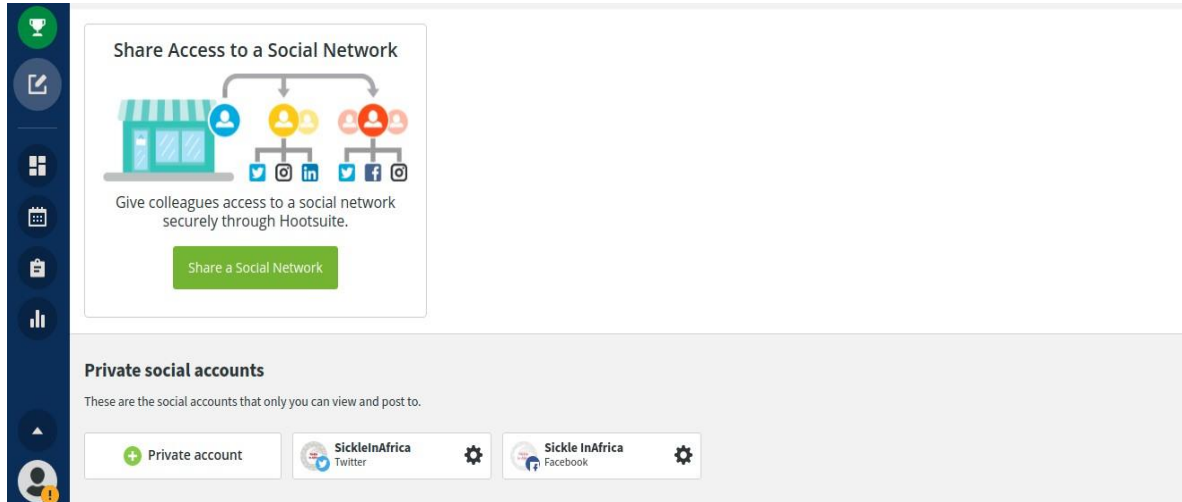


Creating organizations:

Tutorial Video: [Hootsuite organization and permissions](#)

4. Approving posts composed by team members

Approve posts scheduled by other team members.



Approving posts:

Tutorial Video: [Approving scheduled posts by team members.](#)

Plans:

Plan	Cost	Core features
Professional	R290/month	<ul style="list-style-type: none"> • 1 user • 10 social accounts • Schedule unlimited posts • access messages in one inbox
Team	R1500/month	<ul style="list-style-type: none"> • 3 users • 20 social accounts • Schedule unlimited posts • access messages in one inbox • Manage team access and roles with permissions • Assign posts or comments to team members
Business	R9000/month	<ul style="list-style-type: none"> • 5+ users • 35 social accounts • Schedule unlimited posts • access messages in one inbox • Manage team access and roles with permissions • Assign posts or comments to team members • Review messages before they publish

Version No.	Date	Authorizer	External Reviewer	Internal Reviewer	Author	Details of changes
1.0	22 July 2021				Wilson Mupfururirwa	n/a (first version)
2.0	3 Dec 2021				Jade Hotchkiss	Updated file heading and file name

Ontology Development 9

SOP: Internal Curator's Guide

PURPOSE

To provide documentation on aspects of maintaining the SCDO that are relevant to curators of the SCDO.

SCOPE

This document is used to guide curators of the SCDO in their curation work.

GLOSSARY/DEFINITIONS

SOP Standard Operating Procedure: Detailed written instructions to achieve uniformity of the performance of a specific function.

GENERAL SCDO RULES

1) Naming/Labelling Convention

- a) All terms should be in singular form (plurals to be avoided if possible).
- b) No symbols are to be used in names/labels of terms.
- c) Names of terms must not end with a full stop ('.').
- d) All class/term names should have capitalisation according to these rules:
 - i) The first letter of each word should be capitalised. This excludes small words (a, an, the, but, and, if, as, or, nor, to name a few), which should not be capitalised unless they are the first or last words in the name.
 - ii) All parts of hyphenated words should be capitalised, e.g. X-Ray.

2) Definition Convention

- a) Descriptions should preferably not include the term name.
E.g. instead of "Hearing impairment is the inability to detect sound lower than..." rather describe "Hearing Impairment" as "The inability to detect sound lower than..."
- b) Descriptions should be as succinct as possible, providing only the minimum necessary information to define the term adequately.

3) Annotation Properties

- a) These annotation properties are compulsory for each term:
 - ***rdfs:label*** - The term's label/name.
 - ***definition*** - A formal definition of the term.

Sentences in a term's definition must start with a capital letter and must end with a full stop.

- **definition source** - A source from where the formal definition was derived. This can be in the form of a link to a relevant webpage, the IRI of a term in an existing ontology, a reference to specific literature, etc.
- If information other than the formal definition was taken from the same or a different source, this should be recorded using the **dc:source** property.

Note: There can only be one URL per each description source property and each dc:source property.

dc:source http://www.ebi.ac.uk/efo/EFO_0001444; http://www.ebi.ac.uk/efo/EFO_0001333 ❌

If there is more than one relevant source URL, a separate property should be assigned to each URL.

dc:source http://www.ebi.ac.uk/efo/EFO_0001444 ✓
dc:source http://www.ebi.ac.uk/efo/EFO_0001333

- **existence in other ontologies** – Indicates the level of existence of the term in other ontologies (useful when wanting to request inclusion of terms in other ontologies).

One of these annotations is chosen from this list:

- Sufficient (exists and has appropriate description)
 - Suggest update to description
 - Suggest update to label
 - Suggest update to label and description
 - Negligible (nothing useful/outdated ontology)
 - Not relevant to context of SCDO
 - Few but definitions not specific enough
 - Few but definitions not freely available
 - Few but definitions not available
 - None
- If the SCDO was responsible for creating the description, the **dc:creator** annotation property should be used to annotate that the SCDO was the creator, e.g. SCDO (name of curator). In such a case, the

existence in other ontologies annotation property should be used to indicate "None".

dc:creator SCDO (Jade Hotchkiss)
 existence in other or None

- b)** Where appropriate, these annotation properties should be applied:
- **abbreviation** – for providing an abbreviation (including acronyms)
 - **database cross reference** - for providing IDs/accessions/CURIEs as references to external resources.
 - **rdfs:comment** - for additional information about the term.
 - **skos:prefLabel** - makes the Display name different from that provided by the rdfs:label property. (Sometimes the experts prefer a different term and want this to be visible in the class tree.)
 - **curator note** – for notes only relevant to the curator(s).
 - Synonym annotation properties. Each synonym provided for a term must be specific, i.e. it must be either related, exact, broader or narrower.
This should be specified using these properties:
 - **hasExactSynonym** - used when the synonym is an exact synonym of the term.

 - These types of synonyms are of lesser importance:
 - **hasRelatedSynonym** - used when the synonym is related but not exact, broader or narrower.
 - **hasBroaderSynonym** - used when the synonym is a broader term that includes the term being annotated.
 - **hasNarrowerSynonym** - used when the synonym is a narrower term that is a further classification of the term being annotated.
- c)** A number of other annotation properties and/or object properties may be given for a term. It is important that such properties are used uniformly across the terms they apply to.
- d)** Some of the properties already mentioned are only used for curation and validation purposes. These should be updated as curation progresses.
- **existence in other ontologies** – if a term, or its definition and/or label, has been updated in the external ontology based on a request from the SCDO, the status here needs to be updated accordingly (e.g. it could go from being "Suggest update to description" to "Sufficient").
 - **curator note** – an annotation property for notes only applicable to the curator/curation process.

CREATING NEW TERMS AND PROPERTIES

Wherever possible, the terms/classes and properties in the SCDO should link to external ontologies. Thus:

- Before creating a new property, existing ontologies should be searched for a similar property and in the case where one does exist, the existing property should be used if appropriate.
- Likewise, if a term exists in an external ontology, it should be referenced appropriately.
- An effort should be made to include new terms and properties generated by the SCDO into relevant existing ontologies (by contacting the relevant ontology developers and requesting inclusion of new terms and/or properties). For example, new properties can be requested for inclusion into the Relations Ontology (RO).

1) Creating a New Term in WebProtege

- Click on the term in the class tree under which you would like to create a new term. Click "Create" (above the ontology tree). Enter the name of the new term. Click "Create".
- When annotating a new term, apply the General SCDO Rules.
- Ensure the IRI is in the correct format (it should already be setup in the settings for new terms' URIs to start with http://purl.obolibrary.org/obo/SCDO_1

2) Creating a New Property in WebProtege

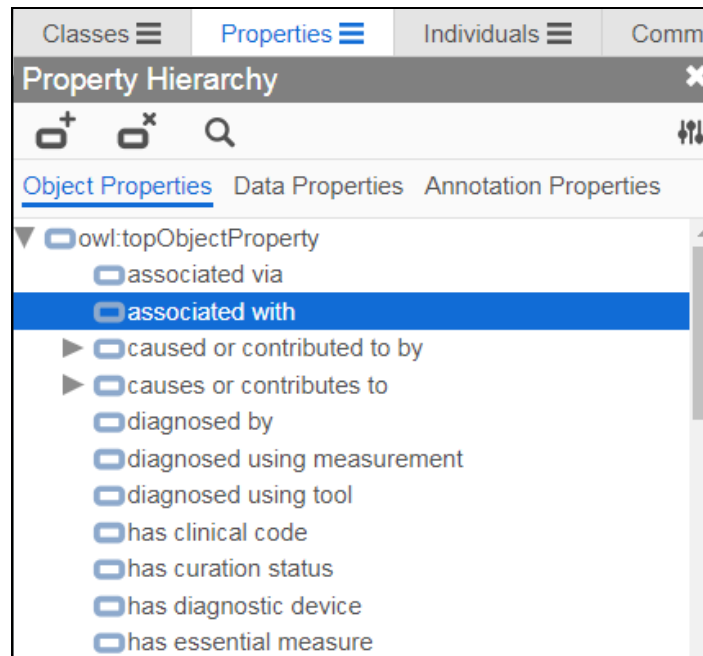
New properties can be created in two different ways:

- a)** When annotating a term, type the new property name in the left-hand side of a blank row, press "Tab", then type in the property description (selecting from the available options if an object or data type property) and press "Tab" again. This will automatically add the new property to the property tree in the "Properties" tab. To further describe the new property, navigate to it in the "Properties" tab.
- b)** In the "Properties" tab, select the appropriate property tree (Object, Data type or Annotation) or existing property under which you wish to create a new property. Click on "Create", provide a name for the new property and click "Ok". Then further describe the new property in the new property entry that has been created.

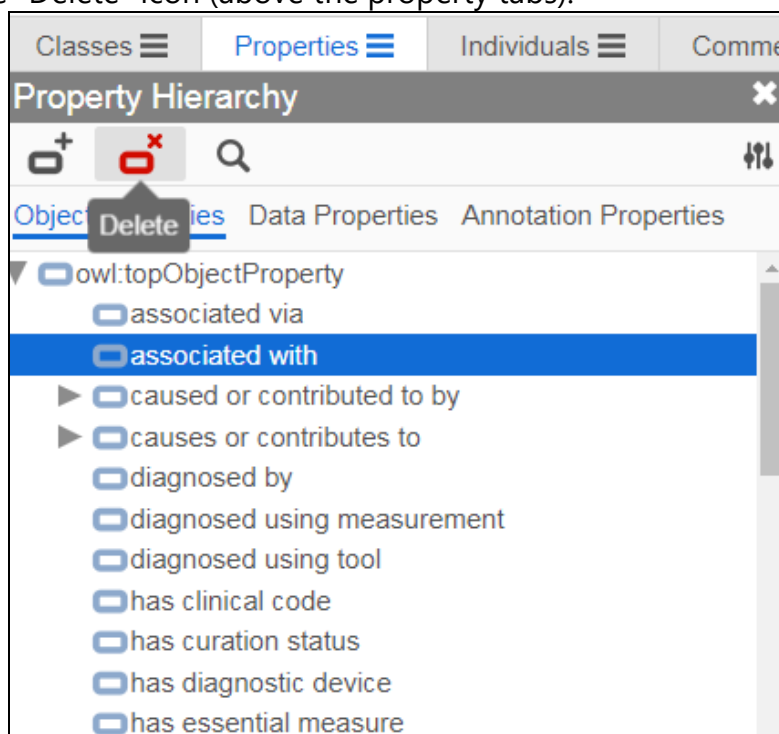
DELETING PROPERTIES

Properties can only be deleted if they are not in use, i.e. if they are not being used to annotate one or more other properties or terms/classes.

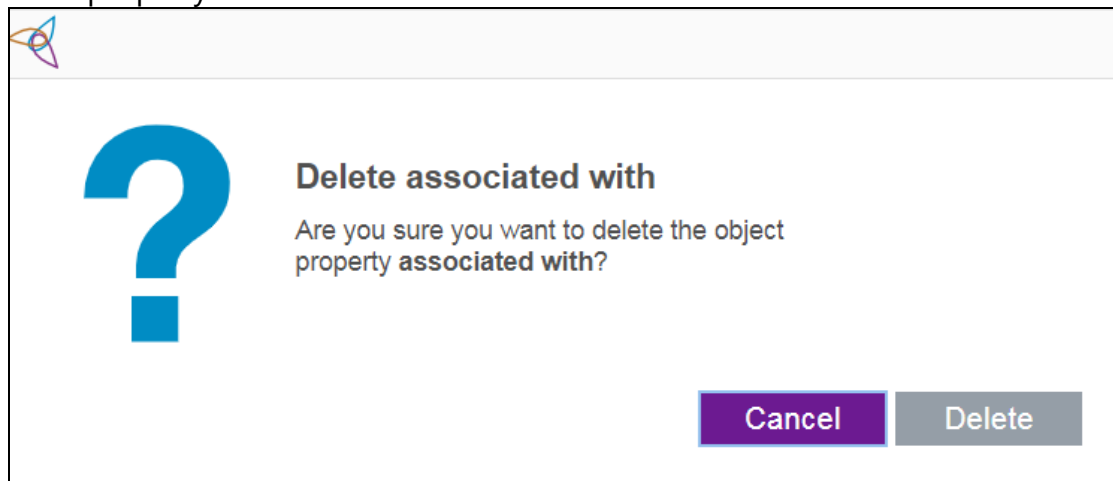
To delete a property, navigate to the "Properties" tab and select the appropriate tab for the type of property to be deleted (i.e. either "Object Property" or "Data Property" or "Annotation Property" tab), followed by the specific property to be deleted (in the example below, the "associated with" object property is to be deleted).



Then select the "Delete" icon (above the property tabs).



A pop-up window will appear where you must confirm that you want to delete the selected property.



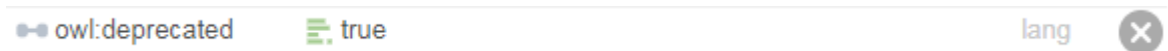
DELETING VS DEPRECATING A TERM

Deleting a term is the complete removal of a term from the ontology, whereas deprecating a term marks the term as being deprecated and moves it to the "deprecated terms" class.

DEPRECATING TERMS

To deprecate a term:

- 1) Add an "owl:deprecated" annotation property with the text "true".



- 2) Move the term to the "deprecated terms" class.
- 3) If the term is replaced by another term:
 - a) Specify this with the "term replaced by" annotation property. Use it to point to the IRI of the term it is replaced by.

Class: Disease Caused by A/Jamaica Plain Genotype

IRI
http://scdontology.h3abionet.org/ontology/SCDO_1000212

Annotations

rdfs:label	Disease Caused by A/Jamaica Plain Genotype	lang	X
created_by	Jade Hotchkiss	lang	X
creation_date	2018-09-07T00:05:55.126Z	lang	X
owl:deprecated	true	lang	X
term replaced by	http://scdontology.h3abionet.org/ontology/SCDO_1000164	lang	X

The IRI will automatically be replaced by the replacing term's name.

Class: Disease Caused by A/Jamaica Plain Genotype

IRI
http://scdontology.h3abionet.org/ontology/SCDO_1000212

Annotations

rdfs:label	Disease Caused by A/Jamaica Plain Genotype	lang	X
created_by	Jade Hotchkiss	lang	X
creation_date	2018-09-07T00:05:55.126Z	lang	X
owl:deprecated	true	lang	X
term replaced by	<input checked="" type="radio"/> Sickle Cell Disease A/Jamaica Plain	lang	X

One can redirect to that term if you click on the "Go to entity" arrow.

term replaced by	<input checked="" type="radio"/> Sickle Cell Disease A/Jamaica Plain	lang	X
Enter property	Enter value	<input type="button" value="Go to entity"/>	

- b)** If the term replacing it has the same label:
 - i) remove the term's "rdfs:label" annotation property before creating the new term
- c)** But if the replacing term has a different label:
 - i) Include the word "deprecated" in front of the text for the "rdfs:label"

Class: deprecated Disease Caused by A/Jamaica Plain Genotype

IRI
http://scdontology.h3abionet.org/ontology/SCDO_1000212

Annotations

rdfs:label	deprecated Disease Caused by A/Jamaica Plain Genotype	lang	X
created_by	Jade Hotchkiss	lang	X
creation_date	2018-09-07T00:05:55.126Z	lang	X
owl:deprecated	true	lang	X
term replaced by	Sickle Cell Disease A/Jamaica Plain	lang	X

- 4) If the term is not being replaced by another term, include the word "obsolete" in front of the text for the "rdfs:label".

Class: obsolete Diffusion Capacity by Biochemical Methods

IRI
http://scdontology.h3abionet.org/ontology/SCDO_0000383

Annotations

rdfs:label	obsolete Diffusion Capacity by Biochemical Methods	lang	X
dc:description	'Not described [Curator team: []]'	lang	X
owl:deprecated	true	lang	X
Enter property	Enter value	lang	

DEPRECATING TERMS

Care should be taken when deleting terms/classes.

Deleting a term that occurs elsewhere in the ontology deletes the term from the other location(s) as well. If a term is to be removed from one of multiple locations in the ontology, remove the relevant "parent" term but do not delete the term itself.

Parents

- Abnormal Hemoglobin
- Abnormal Hemoglobin in SCD

Enter a class name

Delete

CHANGING A TERM NAME

If a term's name needs to be changed, the term will need to be deprecated and replaced by a new term with the new name.

Steps:

- 1) Create a new term (in the relevant location) with the new name.
- 2) Copy all annotations (except for `rdfs:label`) across from the old term to the new term.
- 3) For the old term:
 - a) Put the word "deprecated" in front of the old term's name (in `rdfs:label`).
 - b) Add an "owl:deprecated" annotation property with the text "true".



- c) Add the "term replaced by" annotation property, pointing to the IRI of the new term. (The IRI will automatically be replaced by the replacing term's name)
 - d) Move the old term to the "deprecated terms" class.
- 4) To make sure all object property relationships of the old term are carried across to the new term, update the term name in all sheets where it appears in this google spreadsheet:
https://docs.google.com/spreadsheets/d/1uUqcZSlww-Y1AhoqWdNi3fZaU4y_2qd0Qk3lzmBZchM/edit?usp=sharing

CLEANING UP TERM FORMAT

This might be necessary if someone created new terms or a new section of the ontology without applying the *General SCDO Rules*.

- Check that all required properties are populated in the *Annotations* section of each term (See under *General SCDO Rules*).
- Check that links provided are functional when clicked on. If *https* was given, it should be changed to *http* (this makes the link "clickable").
- Check that synonyms are specific (See under *General SCDO Rules*).

CREATING NEW WEB PROTÉGÉ PROJECT

In the life-cycle of an ontology, sometimes the ontology must be exported to Desktop Protégé for certain edits and then back into Web Protégé as a new project.

When doing this, it is important to:

- 1) Use the naming ONT_EDIT_DATE protocol for naming the project, where ONT is replace by the ontology abbreviation (SCDO), and DATE is replaced by the date of creating the new project.

- 2) Edit the settings of the new Web Protégé project so that for the New Entity Settings:
 - a) **IRI Prefix** is set to the ontology IRI followed by the prefix to be used for new entities, such as classes (e.g. http://purl.obolibrary.org/obo/SCDO_1)
 - b) **IRI Suffix** is set to **Auto-generated OBO Style ID**
 - c) **Digit Count** is set to **6**

New Entity Settings

IRI Prefix

OBO IRI prefixes should end with an underscore

IRI Suffix

Auto-generated OBO Style Id ▼

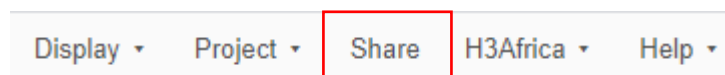
Digit count

User specific ranges

Type user name	start	end
----------------	-------	-----

MANAGING USER RIGHTS AND OTHER SHARING SETTINGS

This is done through the “Sharing Settings” page, which is accessed by clicking on “Share” on the top right of the screen when in the WebProtégé project you want to share or edit user rights for.



- 1) To share a curator’s version of the ontology with other curators so that they can also make edits to the same WebProtégé project:
 - a) Best is to add them on the “Sharing Settings” page, under the “Share with specific people” section, rather than sharing a link (which is not as secure). The WebProtégé user who created the project will automatically be added here and will have “manage” rights.
 - b) Someone with “manage” rights will be able to add specific users to share the project with, by:
 - i) Adding the person’s WebProtégé username (they will need to already have created a WebProtégé account and share their username with you)

- ii) Setting their rights to "edit"
- iii) Clicking "Apply"

c) User rights can be edited by someone with "manage" rights.

- i) To revert a person's rights (i.e. giving them fewer rights, e.g. from "edit" to "view"), the person needs to be removed and re-added with the new rights.

2) To share a WebProtégé project in a way that allows others to view it using a link (sign-in still required):

- a)** Select the "Link sharing enabled" checkbox
- b)** Set the "Anyone with the link can" multiselect field to "view"
- c)** Click "Apply"

Version No.	Date	Authorizer	External Reviewer	Internal Reviewer	Author	Details of changes
1.0	21/09/2022				Jade Hotchkiss	n/a (first version)

Ontology Development 10

SOP: Updating Term Webpages on SCDO Website

PURPOSE

To describe the procedure for updating term web pages for the SCDO website, using an updated OWL file containing translations of one or more languages (**currently this is only functioning for one language translation, French. The program will need to be updated for additional translations*) for the SCDO and the python program "generating_class_webpages.py" and its associated template files.

SCOPE

This procedure is used by curators when producing updated term web pages for the SCDO website. This is done when a new version of the SCDO, containing translations, is released.

GLOSSARY/DEFINITIONS

- Python** The programming language in which the program used in this SOP was written.
- SOP** Standard Operating Procedure: Detailed written instructions to achieve uniformity of the performance of a specific function.

SYSTEM REQUIREMENTS TO RUN PROGRAM

Instructions below are for Windows 10.

For the python program to run successfully, the following need to be installed:

1. Python.
Ensure that Python is setup correctly on your PC (Use **Program Running 1 SOP**)

REQUIRED INPUT FOR PROGRAM

Before the program can be run:

1. A copy of the most recent SCDO OWL file needs to be in the same location as the "generating_class_webpages.py" python program and needs to be named in the program code.
2. An empty "output" folder needs to be in the same location as the "generating_class_webpages.py" python program.

PREPARING INPUT FOR PROGRAM

1. Use the "owl_file_path" variable in the python program code to specify the name of the SCDO OWL file to be used (include the file extension (.owl)).

2. If you have run the program before and the "output" folder already contains files from the previous run, you can move those previous files to a different location or delete them if you don't want to keep them.

RUNNING THE PROGRAM

Run the *generating_class_webpages.py* python program by following the procedure in the **Program Running 2 SOP**.

USING THE OUTPUT FILE

The output files will appear in the "output" folder (which is in the same location as the python program).

Note: In order to use these files to update the SCDO website, you will need to have admin access rights to edit the files on the SCDO website.

1. Delete the old html files from the "Term_files" folder on the website where these files are stored.
2. Upload the updated html files from the "output" folder to the "Term_files" folder on the website.

Version No.	Date	Authorizer	External Reviewer	Internal Reviewer	Author	Details of changes
1	13/06/2022				Jade Hotchkiss	n/a (first version)

Data and Database Management

SOP: Data backups at sites, SPARCO Hub and SADaCC

PURPOSE

To describe the procedure for backing up data.

SCOPE

This procedure applies to data and databases of the Sickle Africa Data Coordinating Centre (SADaCC) as well as those of the Sickle Pan African Research Collaboration (SPARCo).

GLOSSARY/DEFINITIONS

Backups	For the purposes of this document, “backups” will refer to both database and non-database data backups.
Data	For the purposes of this document, “data” will refer to all database as well as non-database data.
Data at rest	Data that is physically stored in electronic form on computer hardware.
Data in transit	Data that is actively being moved from one location to another.
GDPR	The General Data Protection Regulation ¹ is a European Union (EU) legal framework that sets guidelines for the collection and processing of personal information of EU citizens.
POPIA	The Protection of Personal Information Act ² or POPI Act is a South African privacy law that seeks to protect and regulate the processing of personal information.
REDCap	Research Electronic Data Capture (REDCap) is a secure web application for building and managing online surveys and databases. While REDCap can be used to collect virtually any type of data (including 21 CFR Part 11, FISMA, and HIPAA-compliant environments), it is specifically geared to support online or offline data capture for research studies and operations.
SADaCC	Sickle Africa Data Coordinating Centre (SADaCC) at the University of Cape Town
SOP	Standard Operating Procedure: Detailed written instructions to achieve uniformity of the performance of a specific function.
SPARCo	Sickle Pan African Research Collaboration (SPARCo) is a sister collaboration of SADaCC, existing of multiple sites across three countries, including one site per country that acts as a country-level data centre or ‘hub’.

RESPONSIBILITIES & PROCEDURES

A. The Principal Investigator (PI) delegates responsibility for ensuring data and database

¹ <https://gdpr.eu/what-is-gdpr/>

² <https://www.gov.za/documents/protection-personal-information-act>

security as well as backup functions, database close/lock and releasing data to a competent data person (data clerk/coordinator and/or data manager as applicable).

- B. The delegated data person records the identifying details of the data (e.g. name and physical location of the data) as well as the distinct online/network location details of where backups are being saved, in a local log and/or the online database backup log on the SADaCC site³. *It is very important that the online/network storage location is on a different server to where the database is physically located.*

Backup datatypes

1. **REDCap clinical data:** Regardless of whether the REDCap server is managed by your institution of yourself, it is advised that the delegated data person downloads an XML copy of the entire REDCap project. Under the “**Other Functionality**” tab (Fig 1.1), select the ‘**Download metadata & data**’ option (Fig 1.2), selecting to include user roles, record status dashboards, reports, alerts, and surveys (if applicable) in the XML file. *The location of this backup should be recorded in a local log and/or the online database backup log on the SADaCC site.*
2. **SQL databases:** For databases in MySQL or MariaDB, a regular data-dump must be performed. This can be automated to occur daily, especially where frequent data capturing happens. This process is ideally managed either via (a) command-line on Unix by the Systems Administrator responsible for the database server or via (b) a graphical user interface like MySQL WorkBench by either a Systems Administrator or designated data person. *The location of this backup should be recorded in a local log and/or the online database backup log on the SADaCC site.*
3. **Sequence data:** Due to the large size associated with genetic data, this data will ideally reside on a secure server with large data disks. These servers will either be managed by your site, your institution, or SADaCC. The Systems Administrator for the data server should implement a regular backup via a utility like *cron* (used on most Unix-like servers). *The location of this backup should be recorded in a local log and/or the online database backup log on the SADaCC site.*

Backup Frequency

Backups should be done daily, weekly, or monthly. The frequency of data capture determines the default frequency of data backups e.g. daily instances of data capture would imply daily backups. In essence, datasets that remain relatively static will have lower backup frequencies.

³ https://sadacc.org/db_bcklog

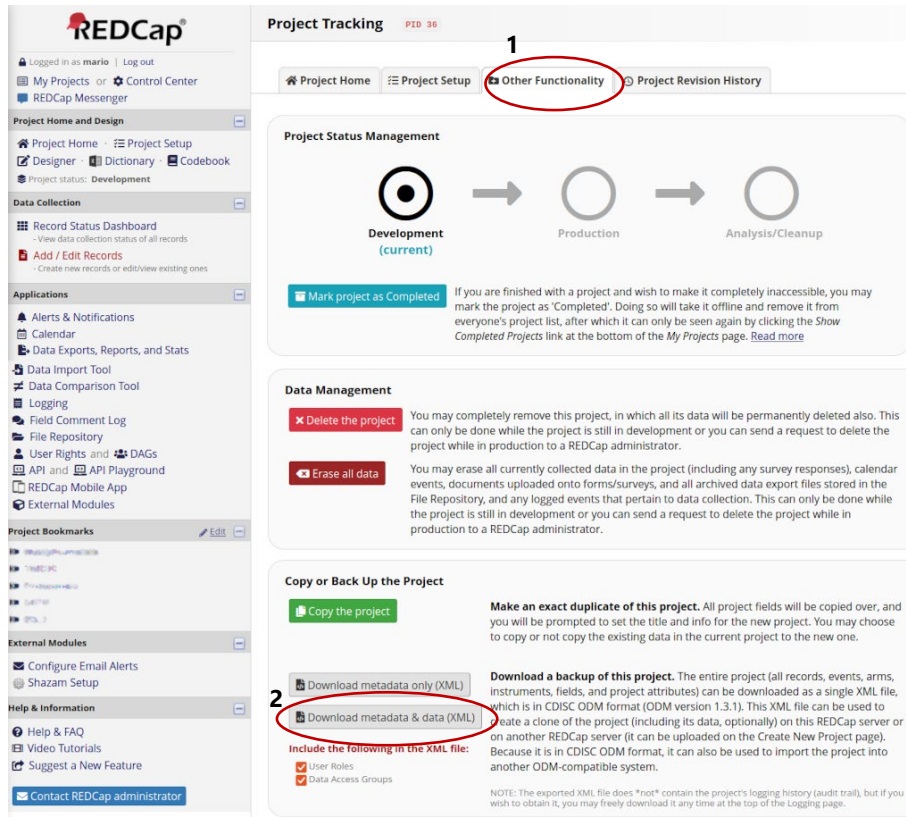


Figure 1: REDCap project back-up options

1. Daily online/network backups:
 - 1.1. If the physical location of the data is on a server managed by the institutional IT department, then regular backups of the institutional server are made following institutional IT policies. *Please familiarise yourself with the institutional back-up frequency with the aim to arrange a frequency suitable for your needs.*
 - 1.2. If the physical location of the data is on a server managed by the department or research group itself, then daily backups should be made to the backup location. The data person records the backup in the local log and/ or the online database backup log on the SADaCC site⁴.
 - 1.3. If the physical location is on one of the SADaCC backup resources, upload the data to either or both, of the Storage Cloud or Network Attached Storage options provided as described in **SOP: Backing up to a Storage Cloud** and **SOP: Backing up to a Network Attached Storage Device**.
2. Weekly removable storage backups:

Regardless of the physical location of the data, weekly backups must be made to a form of removable storage media, e.g. flash drive or external hard drive. The data person

⁴ https://sadacc.org/db_bcklog

records the backup in the log. The removable storage media should be password protected and be kept in a secure location.

3. The data person checks the security and integrity of the back-up process periodically by checking that the current removable storage back-up may be restored and interrogated without error. The data person records the test and the test outcome in the log.

Data security

1. On-site security:
 - 1.1. **Hardware Security:** All data storage devices, including desktop computers, servers, as well as external devices should be stored in a space that is secured behind a locked door. This lock can be physical and/ or electronic.
 - 1.2. **Software Security:** All desktop computers, servers, as well as REDCap should have password access enabled. Each permitted user should have a unique username and password, and this shall not be shared. (*See **SOP: Username and Password Management***)
2. Off-site security:
 - 2.1. **Anonymised data:** In compliance with the South African POPI act and the European GDP Regulation, data shared amongst sites should exclude any identifiers that will compromise the identity of the individual from whom data was sampled. *Only de-identified data should be shared.*
 - 2.2. **Encrypted data:** Data shared amongst sites or stored off-site should be encrypted. Encryption should *ideally* be enforced for data *in transit* or *at rest*. Where the data sharing agreement allows, the decryption key for encrypted data should also be provided in conjunction with the encrypted data *at rest*.

DOCUMENT VERSION HISTORY

Version	Date	External Reviewer	Internal Reviewer	Author	Details of changes
1.0	Sep 2020		Mario Jonas	Annemie Stewart	n/a (first version)
2.0	1 Aug 2021		Victoria Nembaware, Jack Morrice,	Mario Jonas	Add <ul style="list-style-type: none"> • REDCap screenshots • prescriptions for genetic data • section on data security • Reference to NextCloud and Synology backup resources

Data and Database Management

SOP: Data backups using Synology Drive Client

Purpose

To describe the procedure for making database backups to a Synology storage device using Synology Drive Client.

Scope

This procedure applies to databases of the Sickle Africa Data Coordinating Centre (SADaCC) as well as databases of the Sickle Pan African Research Collaboration (SPARCo).

Glossary/Definitions

NAS	Network Attached Storage: a storage device connected to a network that allows storage and retrieval of data by authorised network users.
SADaCC	Sickle Africa Data Coordinating Centre at the University of Cape Town
SOP	Standard Operating Procedure: Detailed written instructions to achieve uniformity of the performance of a specific function.
SPARCo	Sickle Pan African Research Collaboration (SPARCo) is a sister collaboration of SADaCC, existing of multiple sites across three countries, including one site per country that acts as a country-level data centre or 'hub'.

Introduction

SADaCC has several Synology storage devices that are used to back up users' data. There are several ways to interact with these servers, 1) a web interface and 2) a desktop application called Synology Drive Client. The instructions in this SOP were created using Synology Drive Client for Ubuntu, but most of the instructions are generic and should equally apply to Windows and Mac.

Download **Synology Drive Client** for your operating system from the link below:

<https://www.synology.com/en-global/support/download/DS418#utilities>

Install Synology Drive Client

Follow the install process for your operating system. In general, for most operating systems, a simple "double-click" on the downloaded file would suffice.

Log in to Synology Drive Client

If an account has been created for you on the Synology storage device, your Systems Administrator would have provided you with:

- 1) Log-in details and
- 2) The web address for the Synology server.

You will need the above-mentioned information if you want to log in to the web interface or the desktop application.

1. Start the Synology Drive Client (Fig1).
2. At the welcome screen, click on "Start Now" (Fig1.1)
3. At the "Connect to you Synology NAS" screen, enter the address for the server (Fig2.1), as well as you log-in details provide to you by your Systems Administrator.
4. Click "Next"
5. In the pop-up window, select the "Backup Task" (Fig3.1). This will pop up a window that will allow you to select the source directory or folder that you want to back up.

Select the source directory

To specify which directories or folders need to be backed up, you need to specify the source directory or folder (Fig4).

It is a good idea to think carefully about which directories you want to back up. For the purposes of this back-up resource, your personal files would not be good backup candidates since you may need to share the data with your supervisor, fellow students or collaborators.

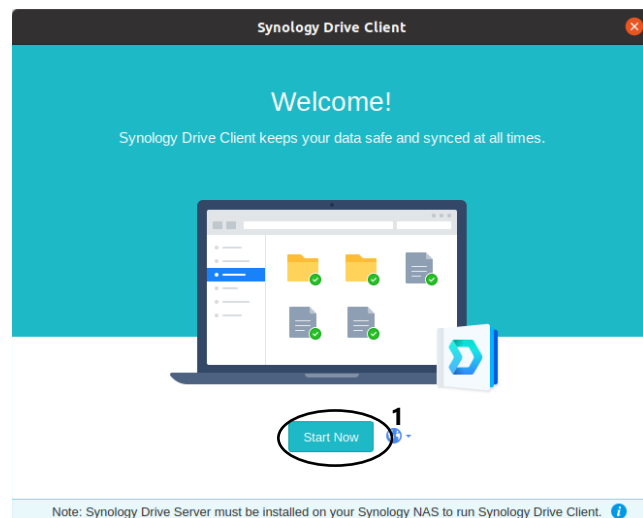


Figure 1: Starting up Synology Drive Client

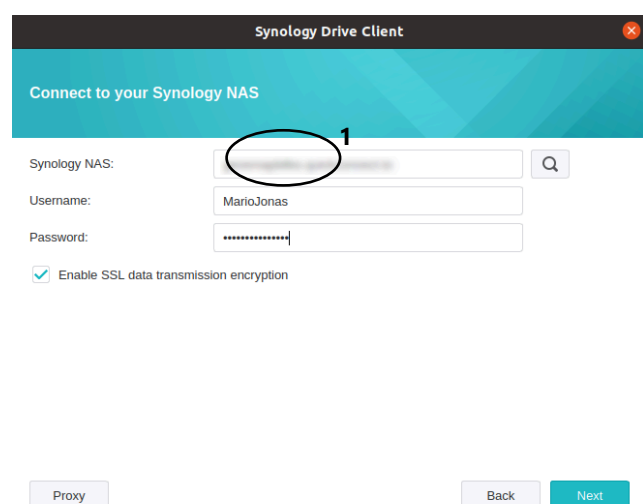


Figure 2: Logging into Synology Drive Client

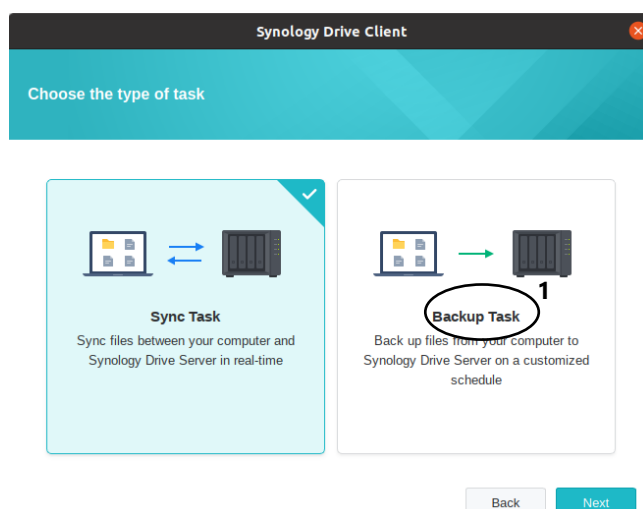


Figure 3: Selecting the backup option

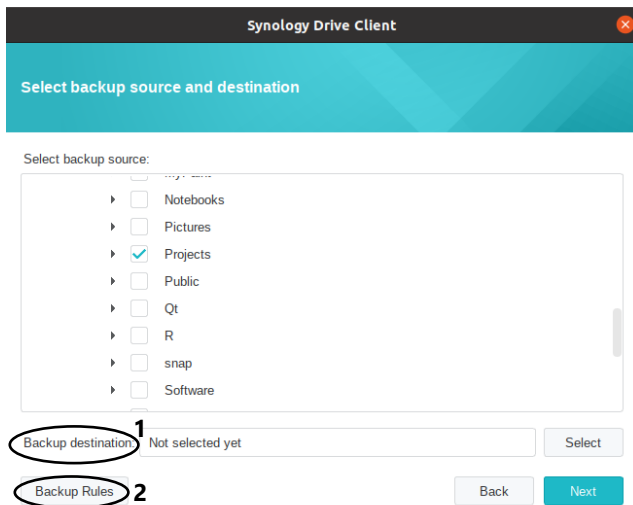


Figure 4: Select the source directory

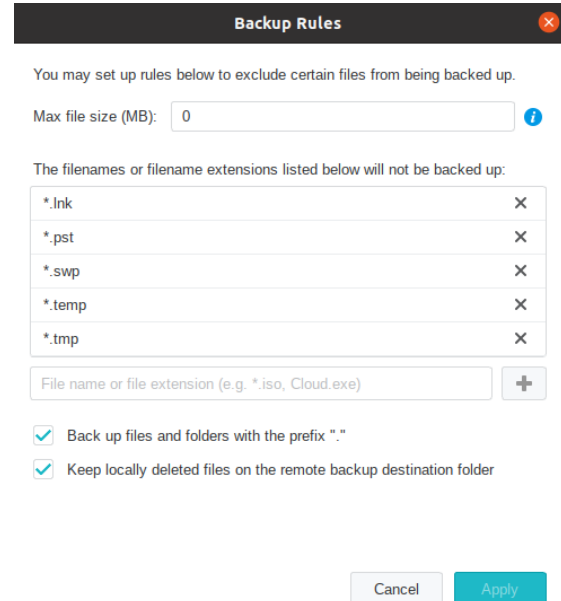


Figure 5: Selecting which file types should not be backed up

Define Backup Rules

Thinking about which files need to be backed up also then determines which files do not need to be backed up. These may include temporary files (eg .swp, .tmp, .temp) or even video (eg .mp4, .mkv, .avi) and audio (eg .mp3, .ogg, .aac) files.

Clicking "Backup Rules" (Fig4.2) allows you to specify which kinds of files should **not** be backed up (Fig5). Excluding specific filetypes decreases the amount of space your backup data will need, as well as the decrease the amount of time needed to transfer your data across the network.

Select the destination directory

You need to select where exactly on the server you want to save your back-up. Your options for location would generally be limited by the kind of access or permissions your account has been given.

1. Next to the "Backup destination" (Fig4.1) click on "Select"

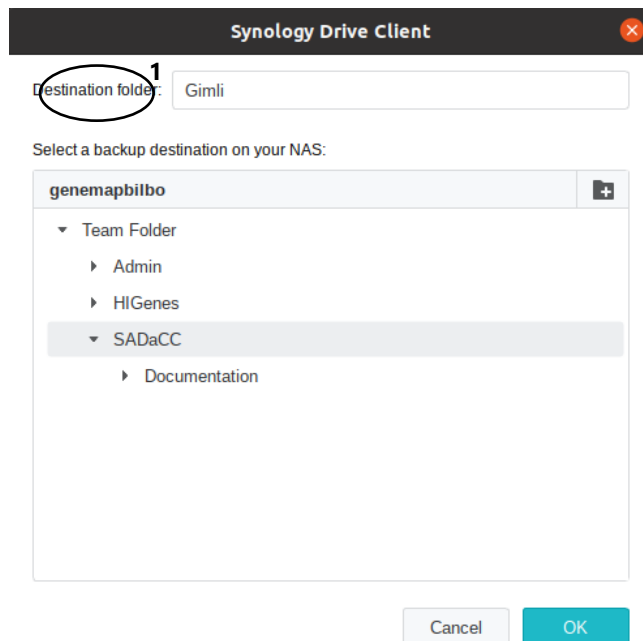


Figure 6: Select the destination folder

- In the pop-up window, the default destination folder is named for the machine name you are connecting from (Fig6.1). You can either accept that default name or change both the location and the name of the backup folder.

Select the back-up mode and schedule

Depending on how often your datasets change, you may choose to back up your data every time a file changes, every day, week, or month. This back-up schedule can be set up as follows:

- Select your preferred backup mode. For convenience and simplicity, I will select to have a "Scheduled backup" (Fig7.1)
- Click "Next"
- In the next pop-up window (Fig8), set your back-up schedule e.g. I have selected to backup at midnight over weekends.
- Click "Next"
- You will be prompted on whether to run the backup right now (Fig9). Since this will be your first backup, click "OK".
- The backup process will start and a dialog box will provide you with a status update on which files are being backed up (Fig10).

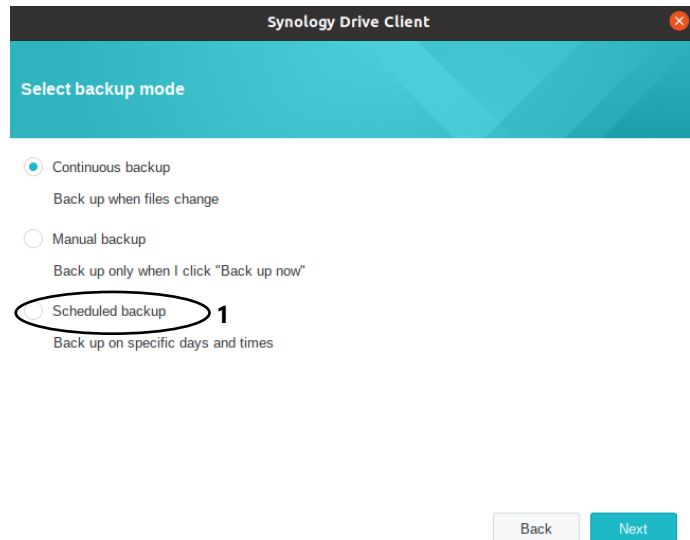


Figure 7: Select the backup mode

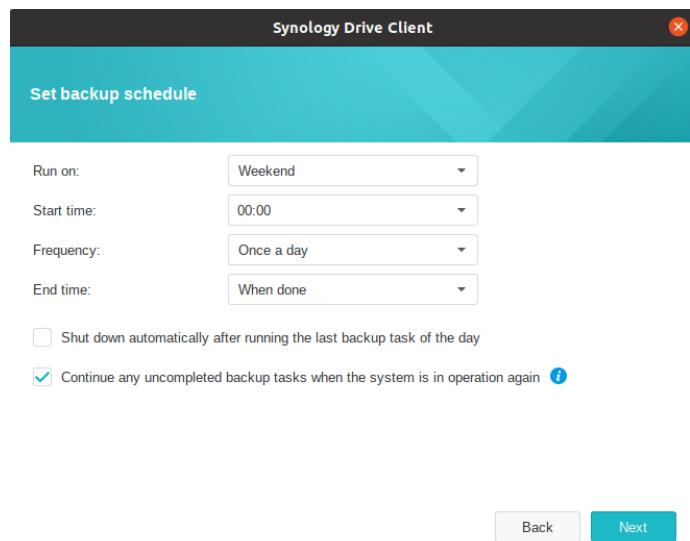


Figure 8: Specify the time schedule for backing up

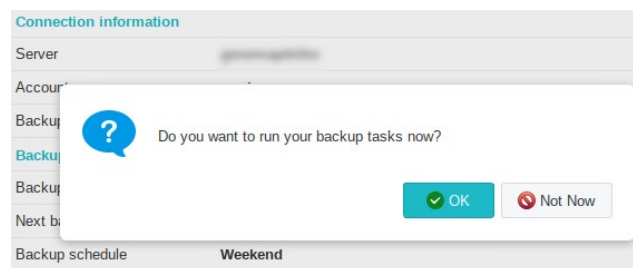


Figure 9: Select to run the first backup

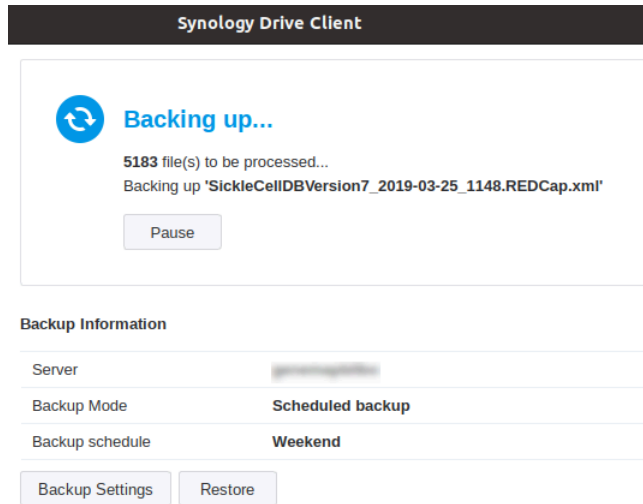


Figure 10: Status report of the first backup run

Document version history

Version No.	Date	Authorizer	External Reviewer(s)	Internal Reviewer(s)	Author(s)	Details of changes
1.0	21 Sep 2021	Victoria Nembaware		Samuel Adadey, Kevin Esoh, Jack Morrice, Noluthando Manyisa	Mario Jonas	n/a (first version)

Data and Database Management

SOP: Backing up to a Network Attached Storage Device

PURPOSE

To describe the procedure for backing up data to a Network Attached Storage device with specific reference to Synology.

SCOPE

This procedure applies specifically to data of the Sickle Africa Data Coordinating Centre (SADaCC), the Sickle Pan African Research Collaboration (SPARCo) and to the SickleInAfrica Consortium generally.

GLOSSARY/DEFINITIONS

NAS	A Network Attached Storage (NAS) is "...a term used to refer to storage devices that connect to a network and provide file access services to computer systems. These devices generally consist of an engine that implements the file services, and one or more devices, on which data is stored..." ¹
SADaCC	Sickle Africa Data Coordinating Centre at the University of Cape Town
SOP	Standard Operating Procedure: Detailed written instructions to achieve uniformity of the performance of a specific function.
SPARCo	Sickle Pan African Research Collaboration (SPARCo) is a sister collaboration of SADaCC, existing of multiple sites across three countries, including one site per country that acts as a country-level data centre or 'hub'.
URL	A Uniform Resource Locator (URL) is the network address of a "...resource (such as a document or website) on the Internet..." ²

Introduction

A Network Attached Storage (NAS) device, as the name indicates, is attached to a computer network, and provides file access to that network. SADaCC uses several such devices for back-up purposes and, by extension, provides this to the SickleInAfrica Consortium at large.

The NAS will, by its very definition, be accessible via the network and hence, will have a specific network or web address or Uniform Resource Locator (URL). The URL for the specific NAS backup location will be provide to you once your account has been created.

¹ <https://www.snia.org/education/online-dictionary/term/network-attached-storage>

² <https://www.merriam-webster.com/dictionary/URL>

There are several options when using a NAS. This document will focus on the use of the web interface for file backups.

Contents

Log in.....	1
Create a folder	1
Set (or Change) Folder Permissions	3
Set (or Change) User Permissions	4
Upload a file	5
Document version history	5

LOG IN

The Systems Administrator for your selected NAS will create an account with you and provide you with you login details. Follow the URL provided to you by your Systems Administrator and enter your login details in the textboxes provided (Fig 1). Once logged in, you should see, at the very least, the two icons labelled “File Station” and “DSM Help” (Fig. 2).

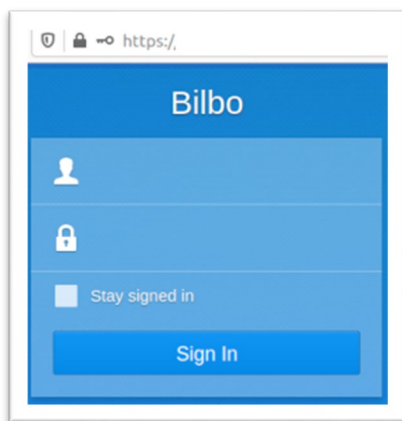


Figure 1: Login screen

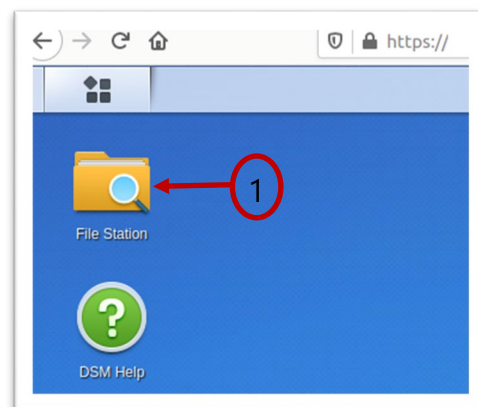


Figure 2: Post-login screen

CREATE A FOLDER

Depending on your permissions, you should be able to create folders under the main folder you were assigned to.

1. Click on the “File Station” icon (Fig 2.1)
2. Click in the “Create folder” button (Fig 3.1)
3. Provide the Folder name (Fig 4.1) and click “Ok” (Fig 4.2). The created folder should now appear (Fig 5.1).

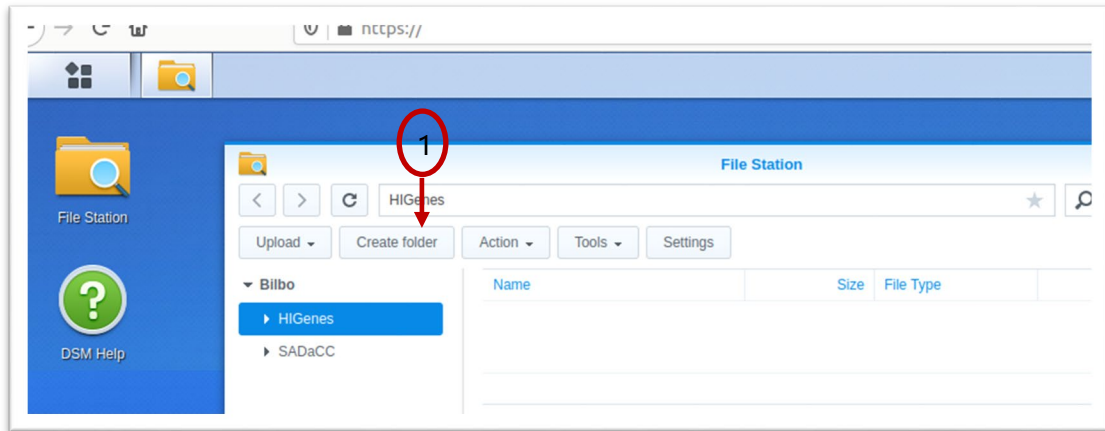


Figure 3: File Station dialog box will allow you to create folders and upload files

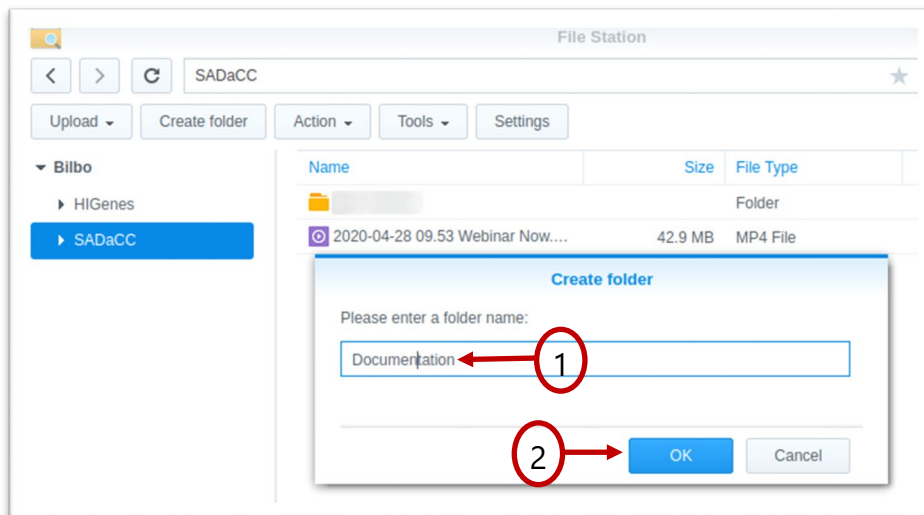


Figure 4: Creating a folder named "Documentation"

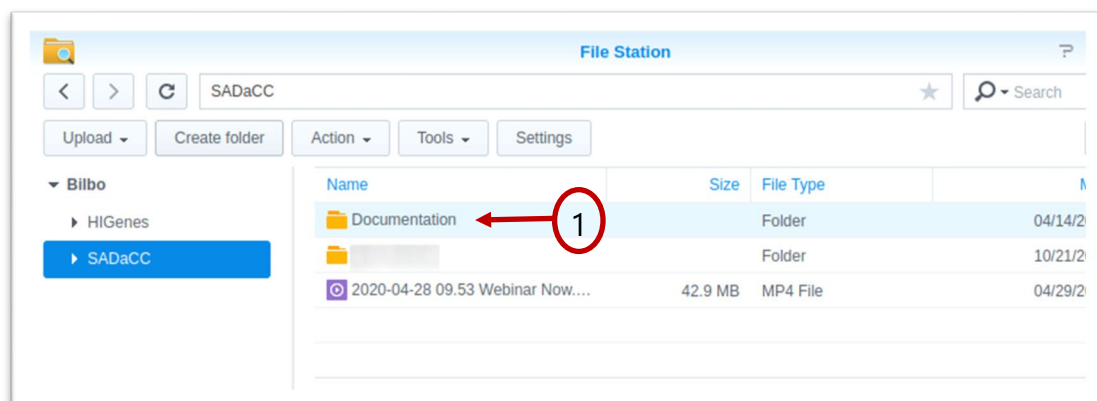


Figure 5: The "Documentation" folder has now been created

SET (OR CHANGE) FOLDER PERMISSIONS

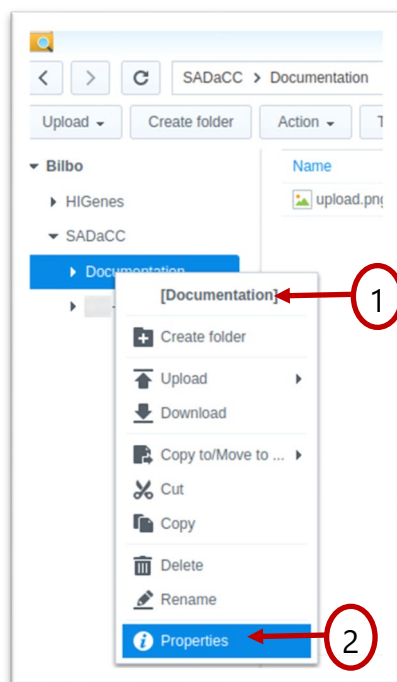
There are at least two approaches to granting permissions. One would be to select a folder and edit the permissions to the folder. This section will deal with this approach. The section “*Set (or change) User Permissions*” will deal with the approach of selecting a specific user and editing the permissions or access for that user.

Once you have created a folder, you can select who has access to that folder (or file). Regardless of the documents you are backing up, there may be a need for you to restrict (or extend) the access to that document (or folder). Changing the permissions on a folder (or file) allows you to exclude everyone but yourself, or perhaps just give your group “read” access.

PERMISSIONS

Types of permissions:

- ✓ **Read:** allows someone to see inside a directory or read the contents of a file.
- ✓ **Write:** allows someone to write or create files in a directory or edit a file.
- ✓ **Administration:** determines whether someone can change the above permissions or even take ownership of a directory or file



1. Select the folder or file to set or adjust permissions for, the just created “Documentation” folder (Fig 6.1) in our case.
2. Select the “Properties” option (Fig 6.2)
3. From the pop-up dialog box, select the “Permission” tab (Fig 6.3)
4. Select the “Create” button (Fig 6.4)

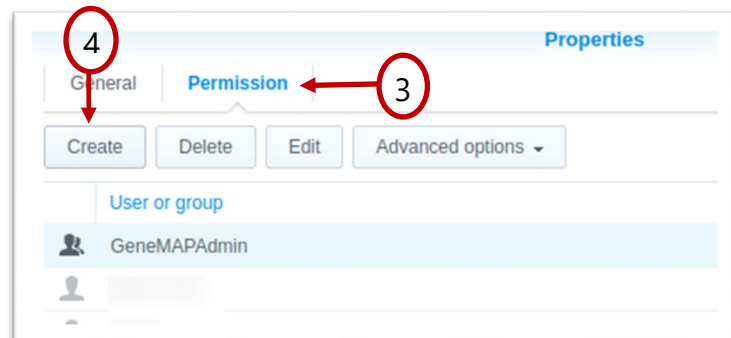


Figure 6: Modifying permissions for the “Documentation” folder

For my purposes, I am going to just give GeneMAPAdmin permissions to this folder, and only allow them to be able to **write** to this folder, nothing else.

SET (OR CHANGE) USER PERMISSIONS

UPLOAD A FILE

The File Station allows you to upload files or data archives (compressed or zipped folders). If a file already exists in the directory, you have the option to *Overwrite* the file or to move on (*Skip*) to the next file in the copy list (Fig 7.1). Once a file has been selected and the upload completed, you will be notified as such (Fig 8.1).

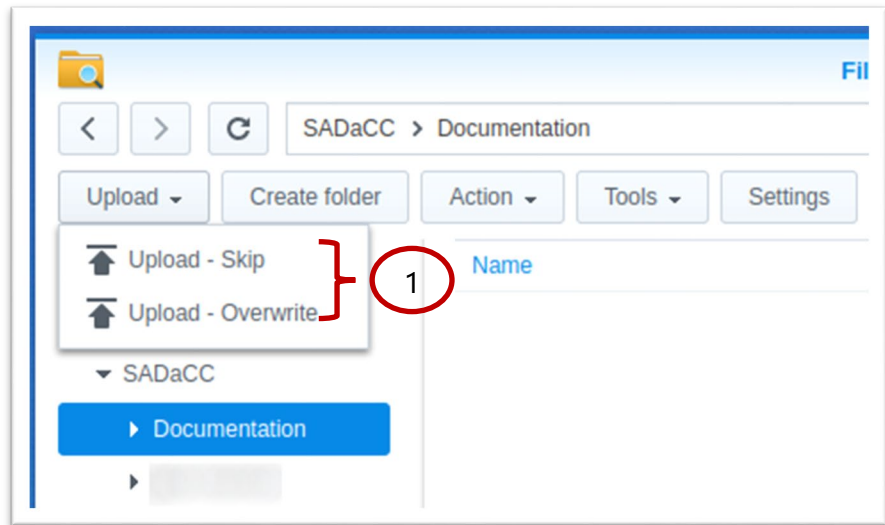


Figure 7: Select the file or file archive to upload as a backup

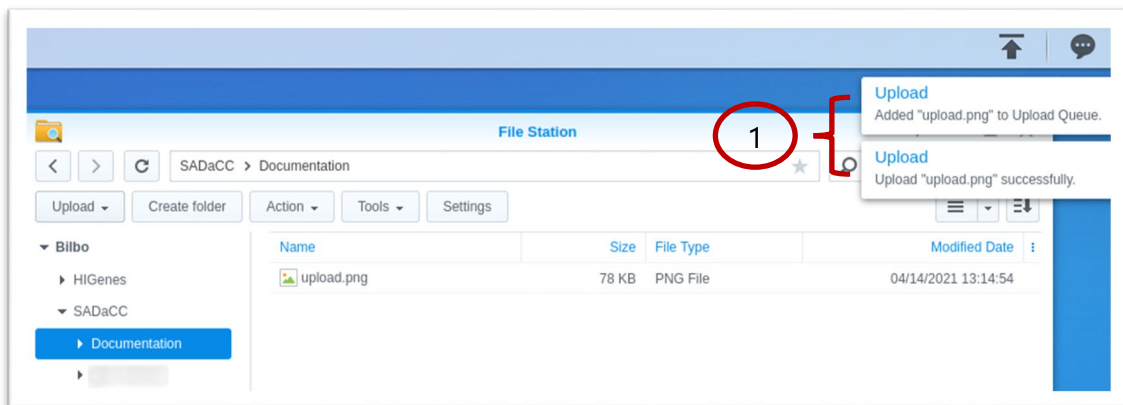


Figure 8: File upload and hence, backup, has been completed successfully

DOCUMENT VERSION HISTORY

Version No.	Date	Authorizer	External Reviewer	Internal Reviewer	Author	Details of changes
1.0	xx July 2021	Ambroise Wonkam	?		Mario Jonas	n/a (first version)

SADaCC GoToMeeting Security SOP

Only a few people will have admin permissions for the GoToMeeting platform. The admin password will be changed every x months.

Securing a meeting room

1. Log into GoToMeeting and navigate to **"Meetings"** on the left-hand side. *(If this is for the default SADaCC meeting room, this step is not necessary)*
2. Click on **"Rooms"** (Figure 1#1) and select the meeting room you want to secure
3. Click on **"Edit"** (Figure 1#2)
4. Click the **"More"** (Figure 2#1) tab and select the **"Require a meeting password"** box and save (Figure 2#2).

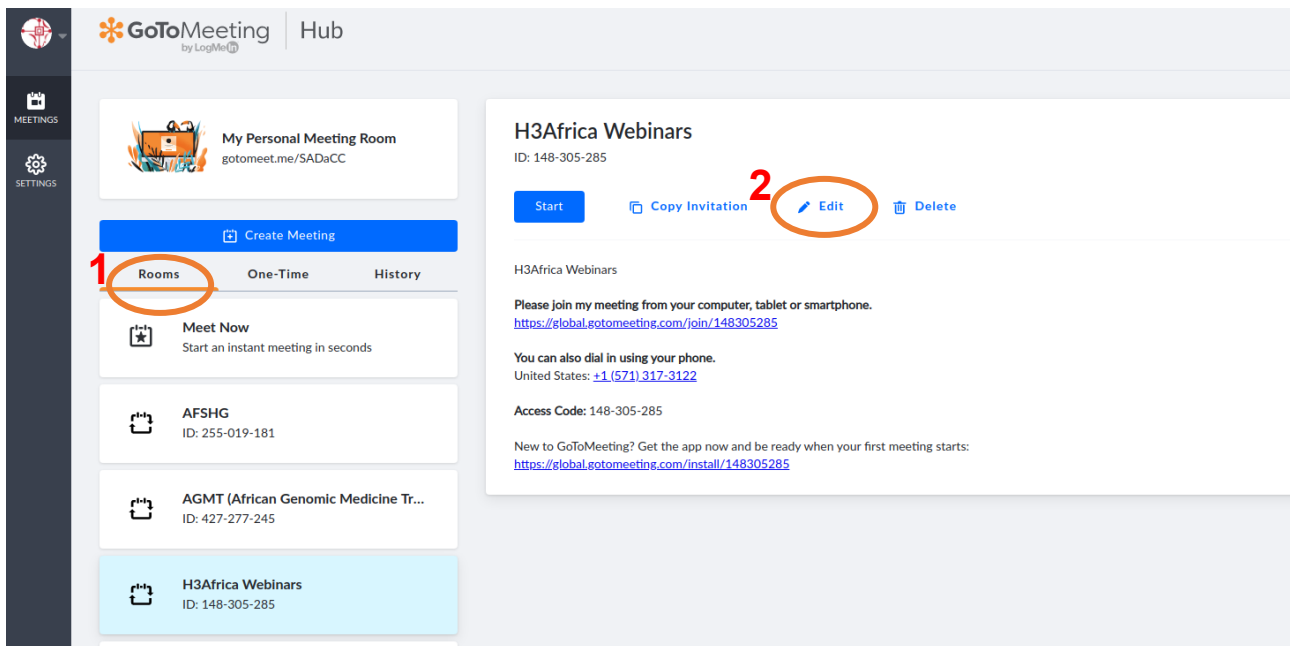


Figure 1: Select the GoToMeeting room to be secured

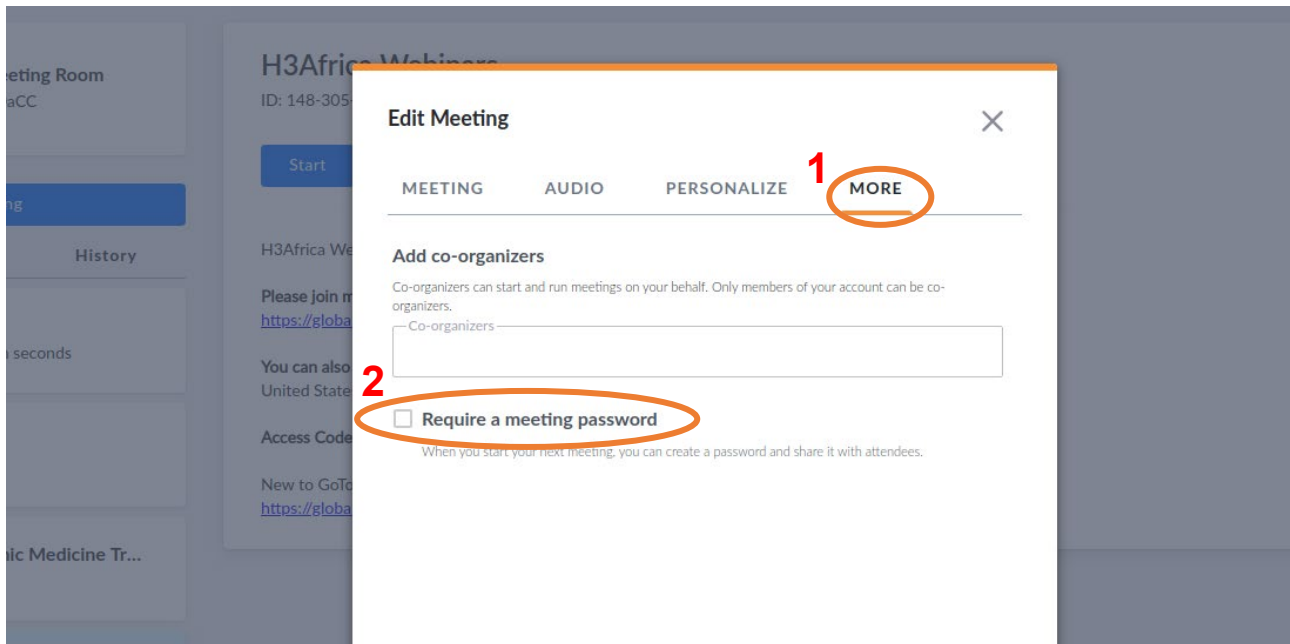


Figure 2: Change settings such that joining the meeting requires a password

Using a secure meeting room

Log in early enough to give ample time for users to receive the meeting password.

1. Log into GoToMeeting and navigate to **"Meetings"** on the left-hand side. (If this is for the default SADaCC meeting room, this step is not necessary)
2. Click the **"Start"** link to start the meeting (Figure 3#1)
3. The meeting will be started in another tab and you will be prompted to set the meeting password (Figure 4). **This password is only valid for one meeting. Each subsequent meeting should have a different password.**
4. Once this is done, share either the password with participants explicitly or use the standardised GoToMeeting invitation provided and substitute the text next to *"This meeting is locked with a password"* with the password set in step 4 above (Figure 5#1).

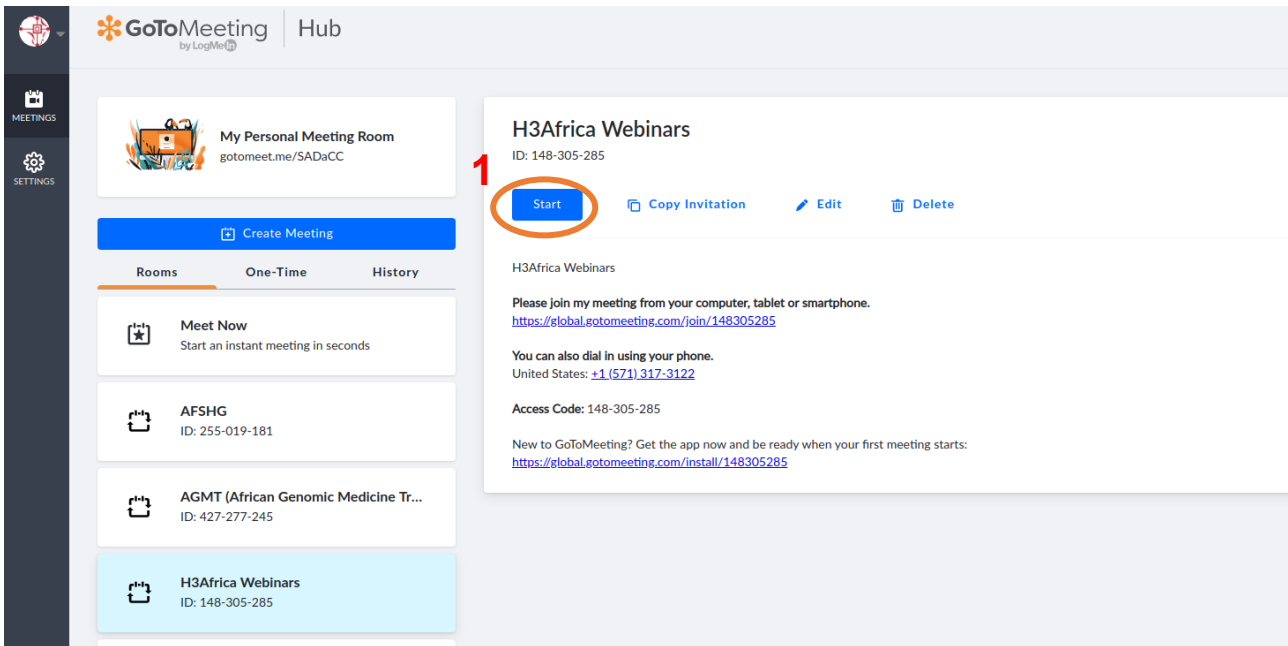


Figure 3: Starting the GoToMeeting

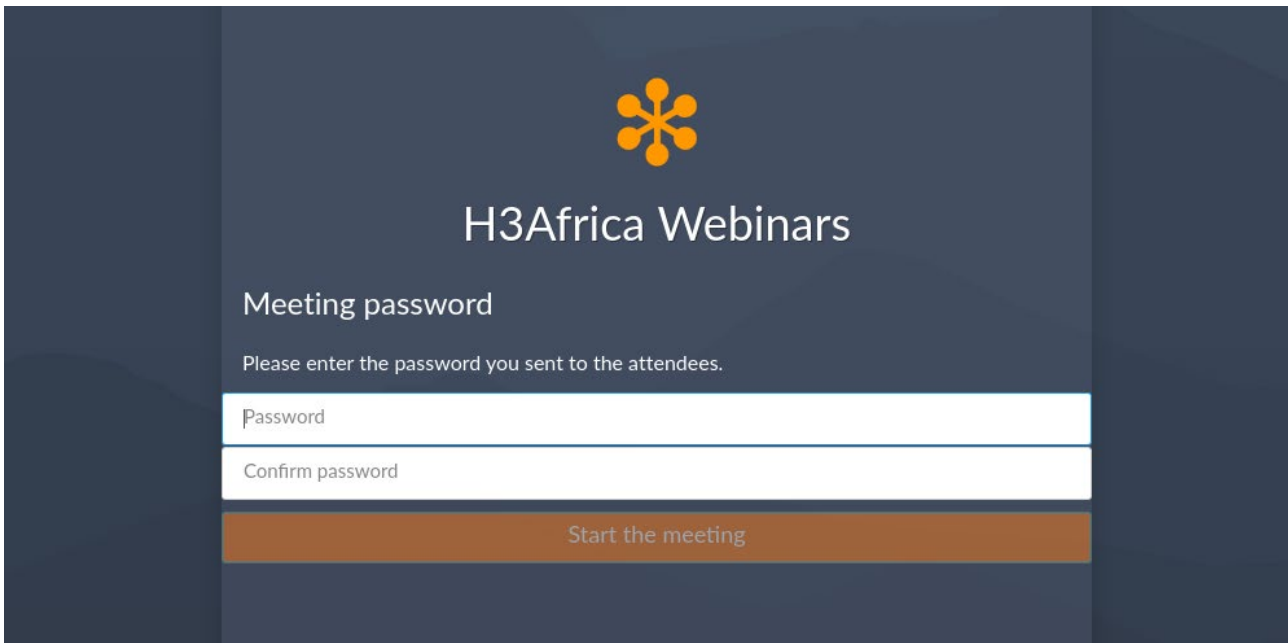


Figure 4: Setting the meeting password

H3Africa Webinars
ID: 148-305-285

Start **Copy Invitation** **Edit** **Delete**

H3Africa Webinars

Please join my meeting from your computer, tablet or smartphone.
<https://global.gotomeeting.com/join/148305285>

This meeting is locked with a password: ##meetingPassword **1**

You can also dial in using your phone.
United States: [+1 \(571\) 317-3122](tel:+15713173122)

Access Code: 148-305-285

New to GoToMeeting? Get the app now and be ready when your first meeting starts:
<https://global.gotomeeting.com/install/148305285>

Figure 5: Sending out the standard GoToMeeting invitation

Locking a room

GoToMeeting has the option to “lock” a room. One of the reasons may be that the allowed number of participants have arrived for example.

1. When the meeting has started and it is determined that the room can be locked, click on the “**People**” icon (Figure 6#1) towards the right of the meeting screen.
2. In the bottom right of the screen, below the “*MEETING ID*”, the little lock indicates that the meeting is unlocked (Figure 6#2).
3. To lock a room, click on the little slider next to the “*Meeting Unlocked*” text (Figure 6#2).
4. A closed lock, with the text “*Meeting Locked*” (Figure 7#1) should now be visible. This means that no-one can subsequently enter the meeting room.

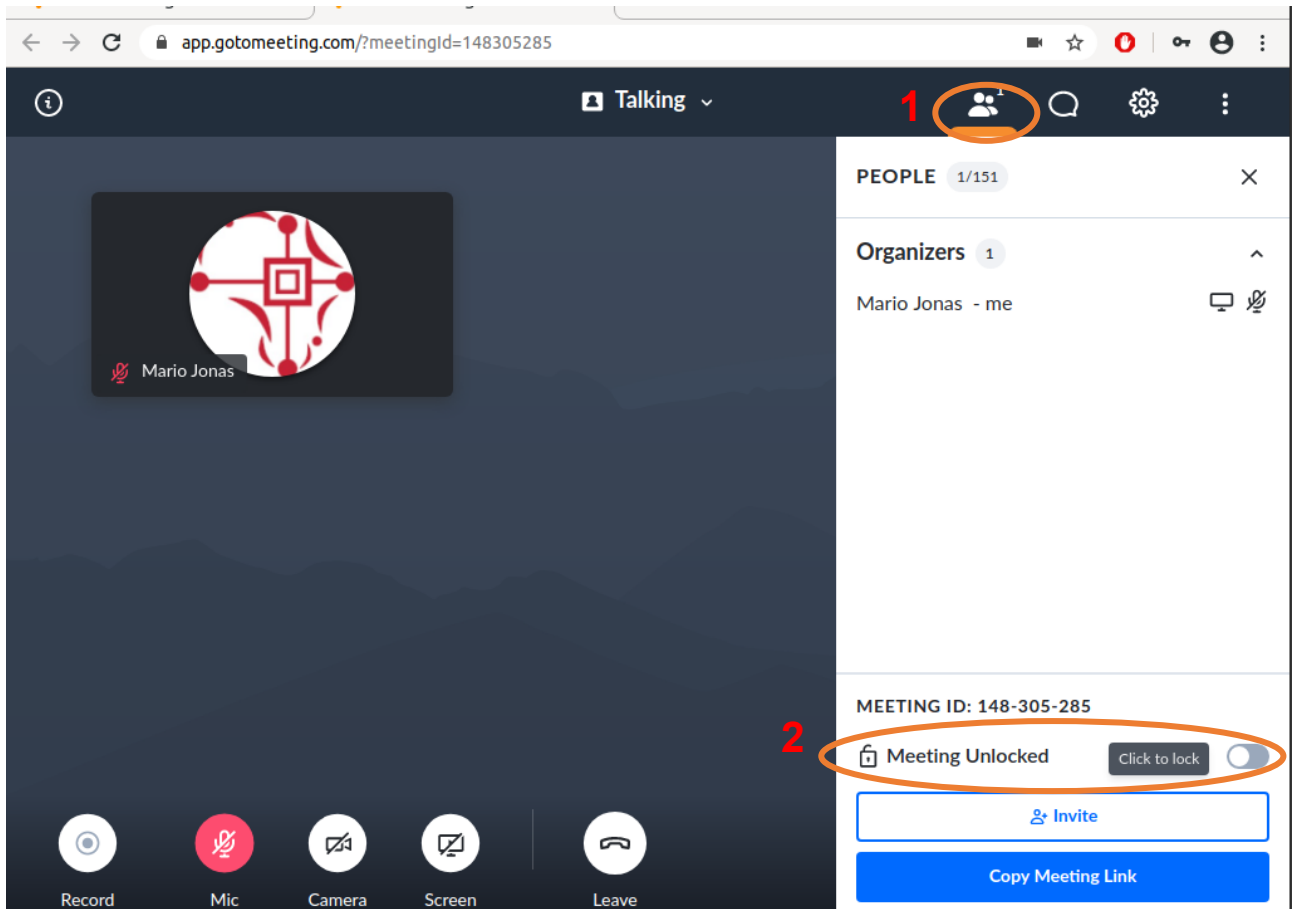


Figure 6: Locking a meeting room

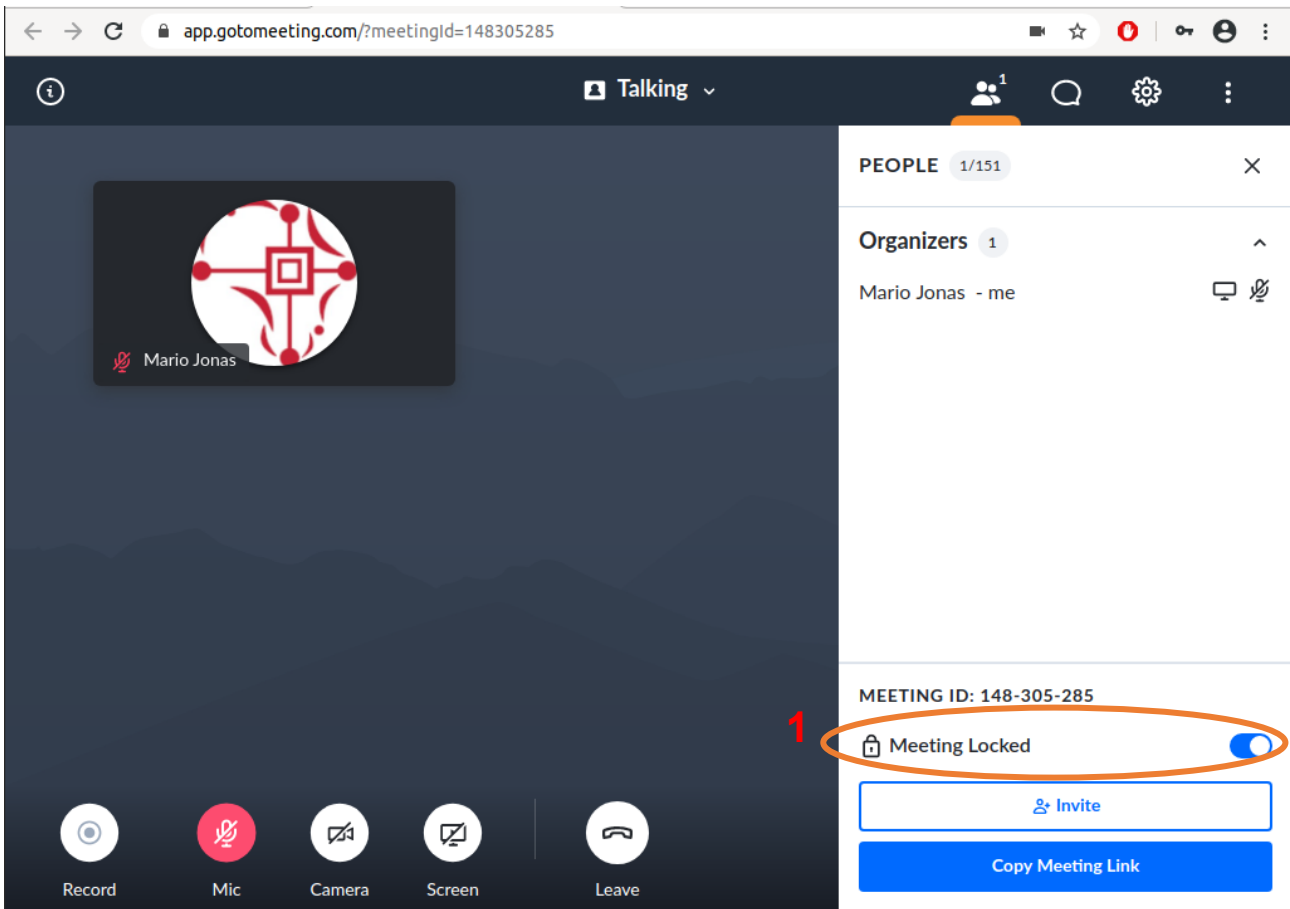


Figure 7: Locked GoToMeeting room

Version No.	Date	Internal Reviewer(s)	Author	Details of changes
1	5 April 2020	Zimkita Magangana, Jack Morrice, Victoria Nembaware, Keano Natus	Mario Jonas	None, First draft

SickleInAfrica Webinar Series

User Guide: GoToWebinar Platform

Contents

Registration	1
Using Firefox (no audio support).....	2
1. Joining the webinar	2
2. Logging in.....	2
3. Interacting	2
4. Exiting the Webinar.....	4
Using Chrome (with audio support).....	5
Using Mobile application.....	5

Recommendations:

1. Join 15 mins before the time to test audio and network connectivity
2. These instructions were only tested on Chrome, Firefox and Android Mobile
3. When you join the webinar, by default your microphone is muted. The organiser will unmute your microphone during the discussions if you raise your hand.

Registration

Register for the webinar by accessing the registration link provided to you. You will be prompted to enter some basic details, including a valid e-mail address. This address will be used to relay all webinar details.

For security purposes, registrations are moderated i.e. rejected or accepted. **Once your registration is accepted, you will receive an e-mail with the log-in details for the specific webinar**

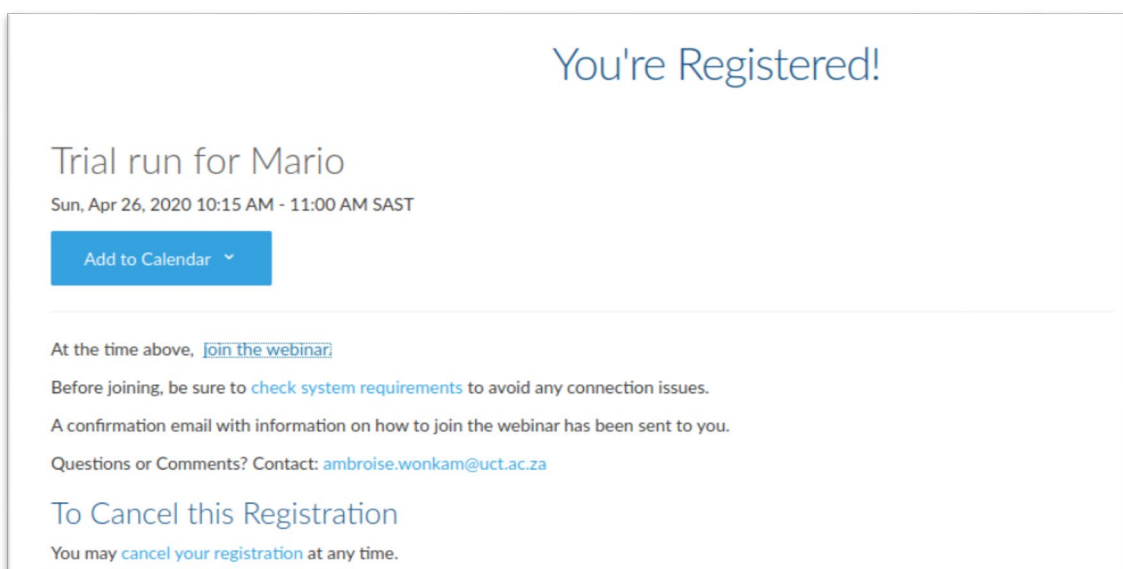


Figure 1: Dialog received at registration

Using Firefox (no audio support)

1. Joining the webinar

You would have either have received the link to the webinar or, upon registering as in Fig 1. Regardless of how you received the link, following it would bring you to a log-in screen that will require the password sent to you in the confirmation e-mail sent to you.

2. Logging in

Enter the password that you have been provided with in the textbox provided (Fig 2). You should be logged in and see a warning about a lack of audio for Firefox as well as details for using audio via phone (Fig 3) if you should so choose. The number to dial should vary according to the country you are connecting from. Please pay attention to the PIN provided if you should choose to connect using your phone.

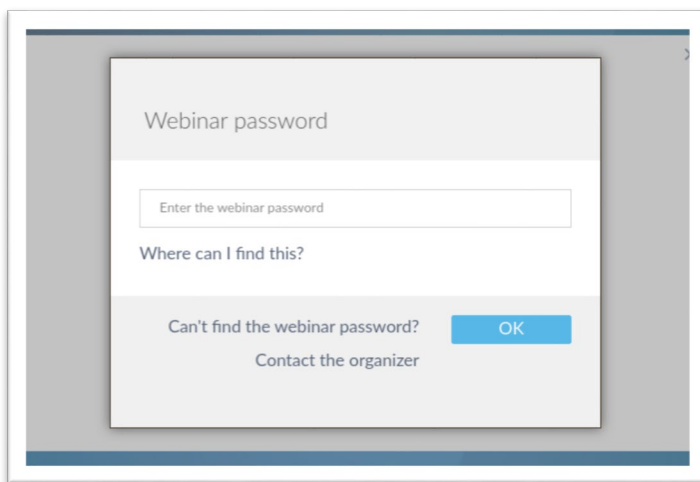


Figure 2: Log-in dialog

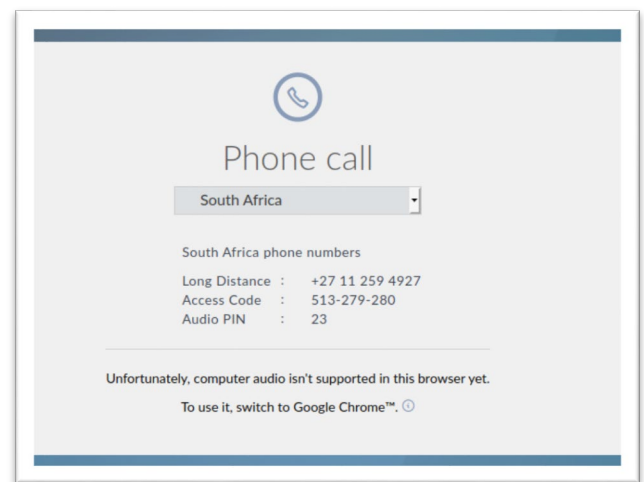


Figure 3: Additional information post-login

3. Interacting

After logging in, you will either see a shared screen from the presenter or a message stating that it is awaiting a screen to be shared.

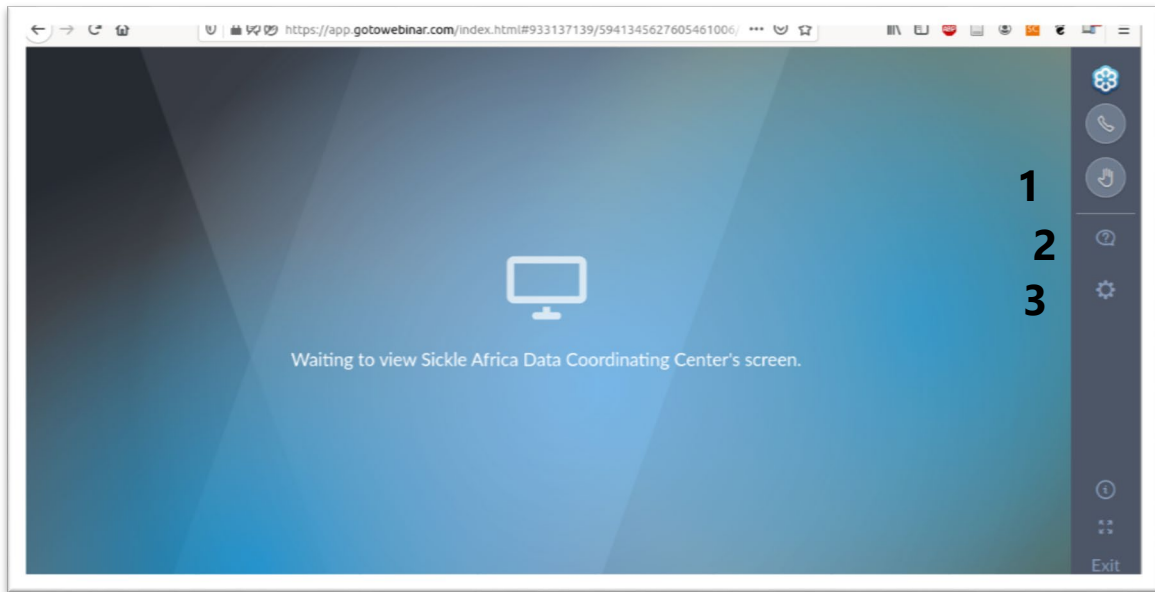


Figure 4: Post-login screen

Icons:

- **1) Raising your hand** (Fig 4.1): Clicking on it turns it green (Fig 5.1) indicating that your hand is raised, and the organiser will address your raised hand in due course. The organiser would then unmute your microphone to allow you to ask a question.
- **2) Asking a question** (Fig 4.2): This question will be asked to the organiser who is labelled as "Staff" (Fig 5.2). The question will be passed onto the presenter at an appropriate time who will then address the question.
- **3) Settings** (Fig 4.3): This is where you can select the correct audio devices and camera if these are detected.

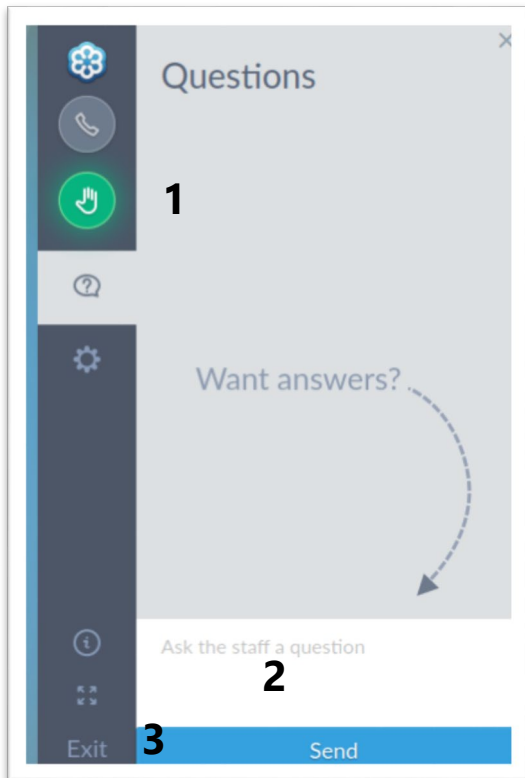


Figure 5: Asking questions

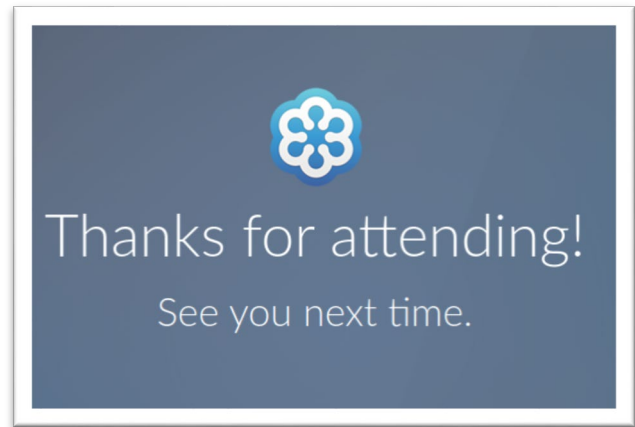


Figure 6: After exiting of your own accord

4. Exiting the Webinar

You can either terminate the webinar by explicitly clicking on "Exit" (Fig 5.3 and Fig 6) or by the organiser ending the meeting (Fig 7).

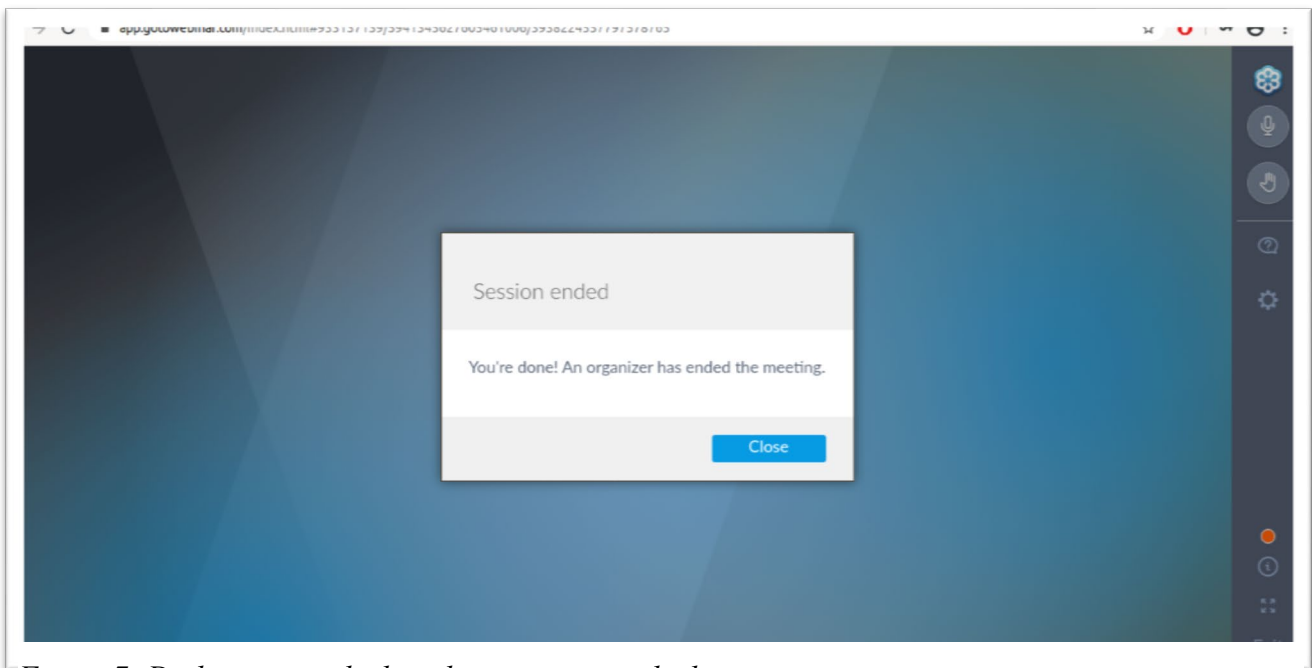
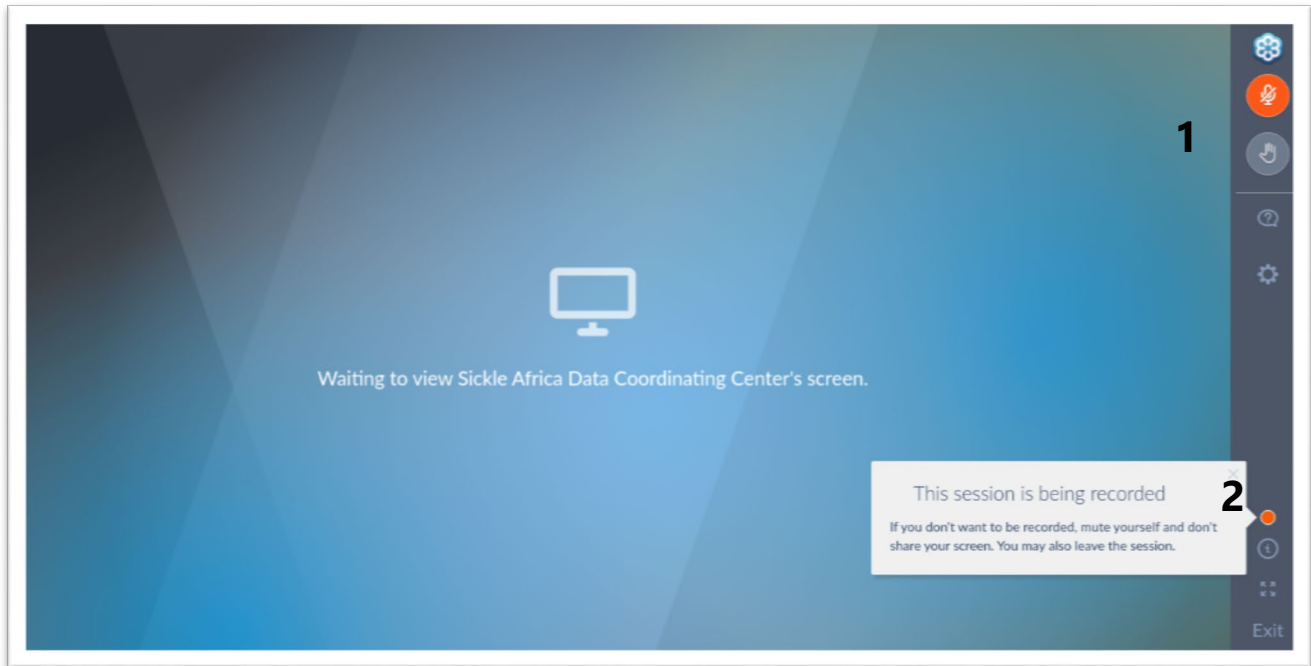


Figure 7: Dialog received when the organiser ends the meeting

Using Chrome (with audio support)

Joining, Log-in, Interaction and Exiting is as for Firefox but the audio support in Chrome provides the options for interacting with the webinar via your microphone (Fig 8)



- **Microphone** (Fig 8.1): By default, you are muted by organiser and will only be able to unmute yourself once the organiser has unmuted you.
- **Recording** (Fig 8.2): The SickleInAfrica webinars are recorded by default and you should see the recording icon at the bottom right of your screen (This icon is also visible in Firefox). If you do not want to be recorded, you can keep your microphone muted. You could also leave the webinar or meeting.

Using Mobile application

1. Install the GoToWebinar mobile application from your relevant mobile app store.
2. Once the application is started, you can attend the webinar by selecting the "Sign in" button (Fig 9).
3. Although you are presented with several options to sign in, these instructions will use the "Join with Webinar ID" option. You will find the Webinar ID under point 2 ("**Choose one of the following audio options**") of the registration mail sent to you (Figs 11 & 12).
4. You will be asked to verify (Figs 13 & 14)) and enter the password (Fig 15) provided in the registration mail sent to you.

Figure 8: Webinar or meeting screen in Chrome

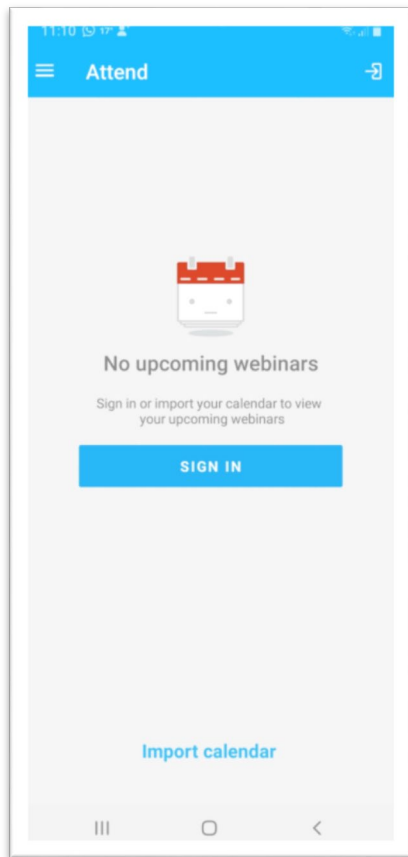


Figure 9: Sign in

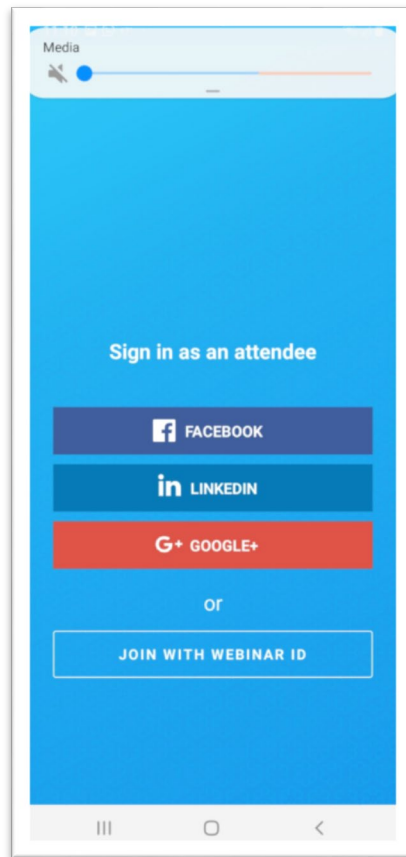


Figure 10: Use Webinar ID

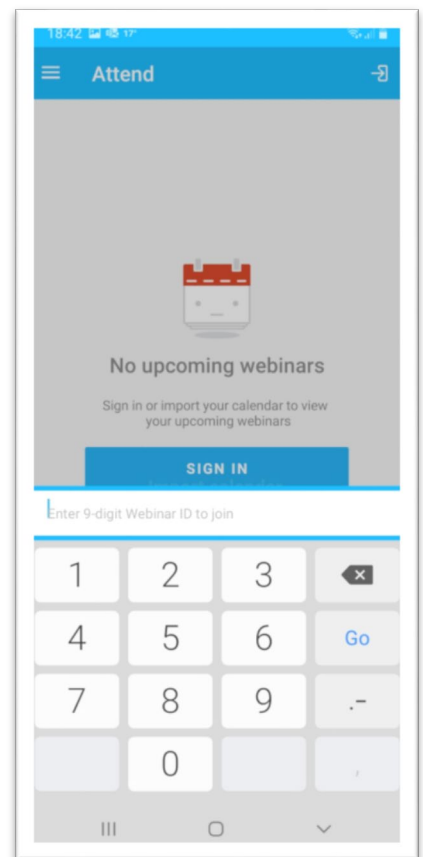


Figure 11: Entering Webinar ID

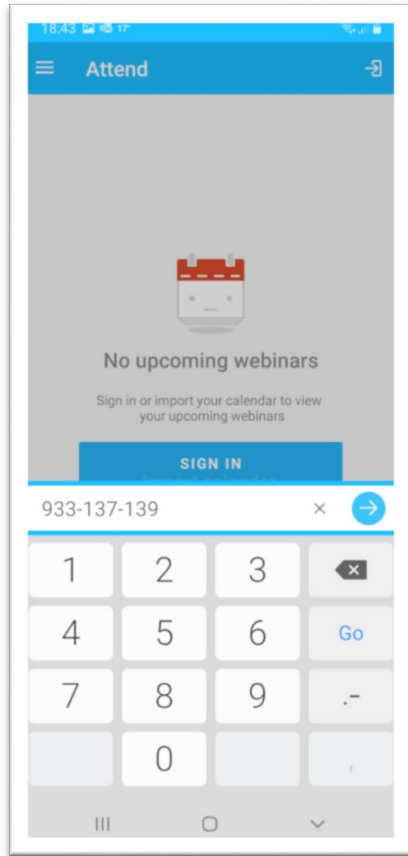


Figure 12: Webinar ID entered

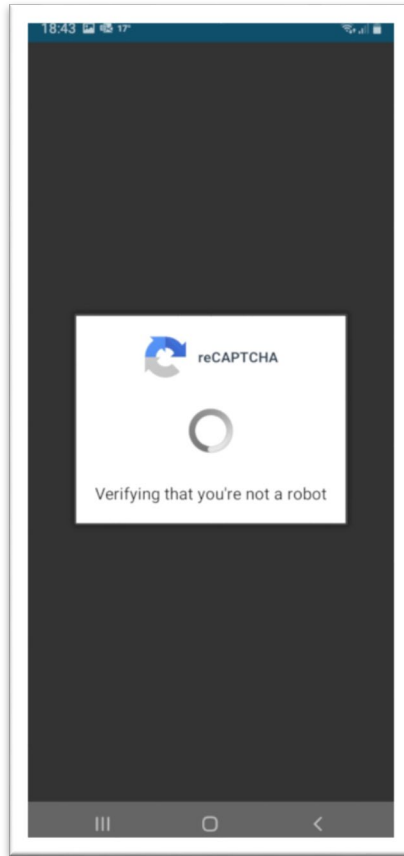


Figure 13: Verification

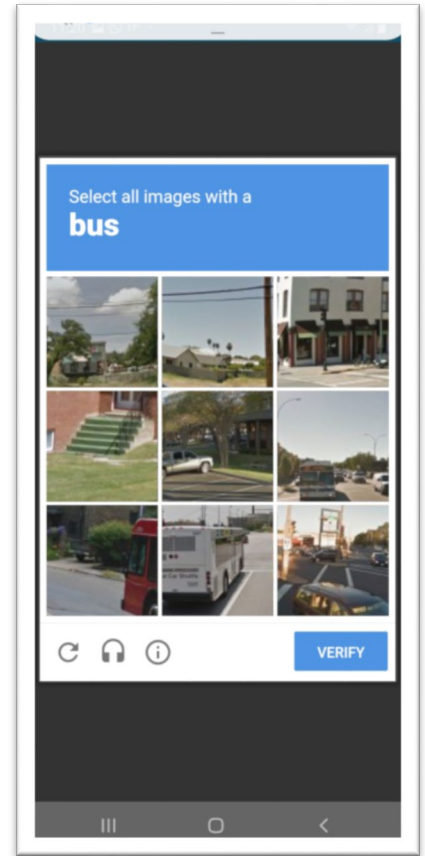


Figure 14: Verification

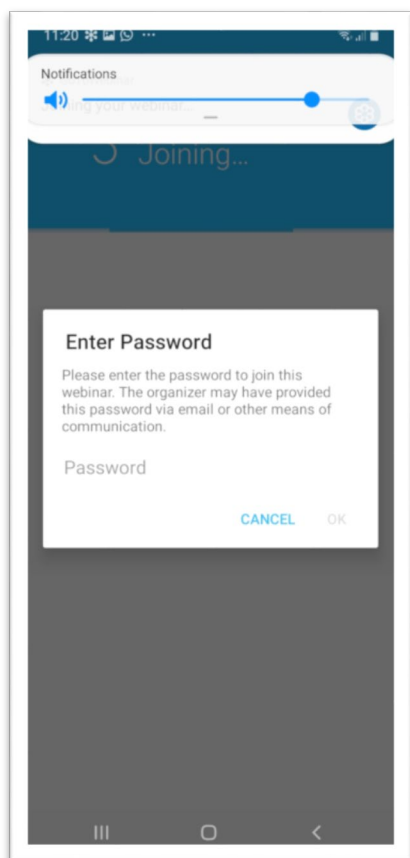


Figure 15: Enter password

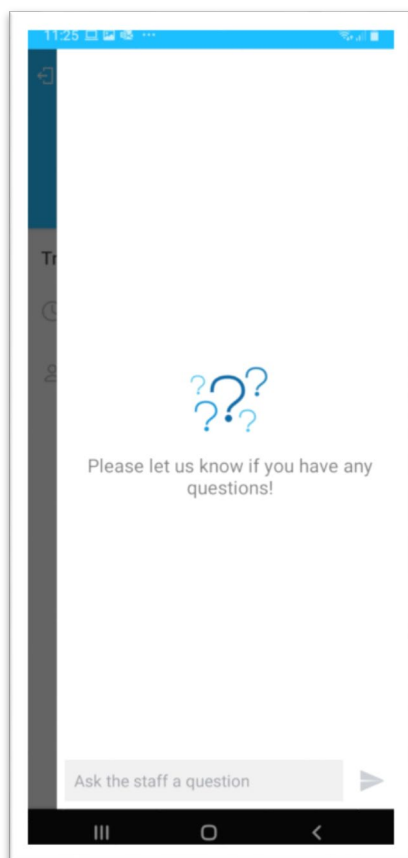


Figure 16: Asking questions

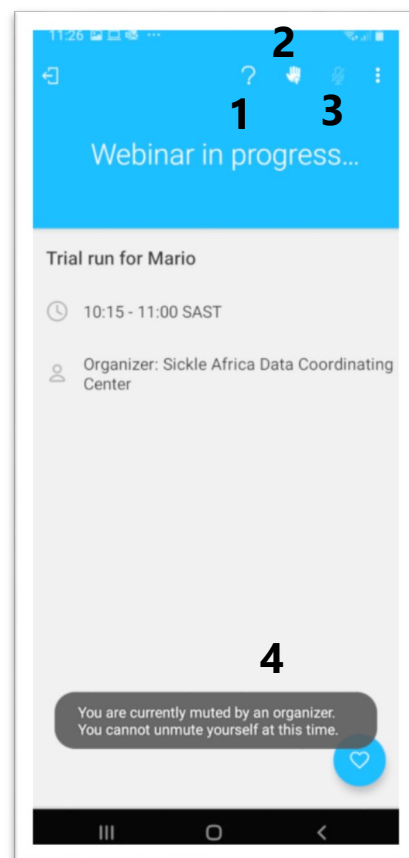


Figure 17: More interaction tools

Icons:

- **Question mark** (Figs 16 & 17.1): This allows you to ask questions during the webinar. These questions will be addressed to "staff" who will then relay the question to the presenter to respond at an appropriate time.
- **Raise your hand** (Fig 17.2): Click this icon when you want to ask a question. Depending on whether Q&A is allowed during the session or after, the organiser will allow you to pose your question via your microphone.
- **Microphone** (Fig 17.3): Provided that your microphone is unmuted by the organiser (Fig 17.4) as well as yourself, you would be able to interact with the presenter in this manner.

Version No.	Date	Internal Reviewer(s)	Author	Details of changes
1	26 April 2020	Jade Hotchkiss, Keano Natus, Victoria Nembaware, Jack	Mario Jonas	None; First draft

Purpose

To provide instructions on how to convert your Powerpoint presentation into a recorded video.

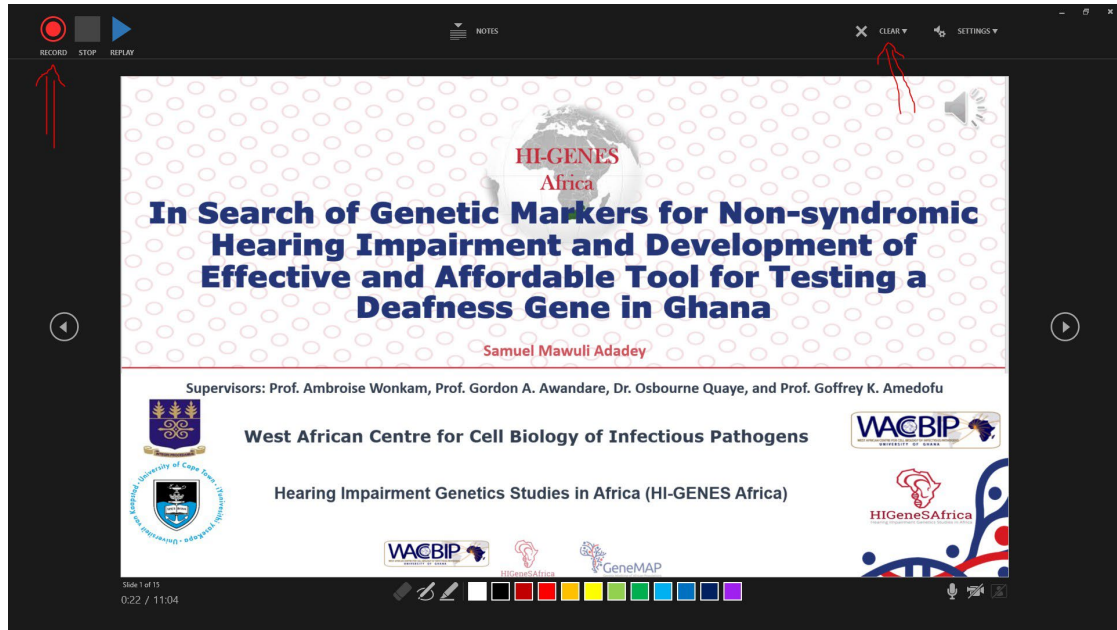
Instructions

Record narration and timings

1. Select **Slide Show > Record Slide Show**.
2. Choose from two options:
 - **Start Recording from Beginning** - to record from the beginning of a presentation.
 - **Start Recording from Current Slide** - to record from a specific slide.



3. Select or clear what you'd like for your recording, and then select **Start Recording (red button)**.
4. Start speaking or add markings to your presentation.



HI-GENES Africa

In Search of Genetic Markers for Non-syndromic Hearing Impairment and Development of Effective and Affordable Tool for Testing a Deafness Gene in Ghana

Samuel Mawuli Adadey

Supervisors: Prof. Ambroise Wonkam, Prof. Gordon A. Awandare, Dr. Osbourne Quaye, and Prof. Goffrey K. Amedofu

West African Centre for Cell Biology of Infectious Pathogens

Hearing Impairment Genetics Studies in Africa (HI-GENES Africa)

VACBIP

HIGeneSAfrica

GeneMAP

Slide 1 of 15
0:22 / 11:04

Note: Audio won't record when slide transitions happen so let these play first before you start speaking.

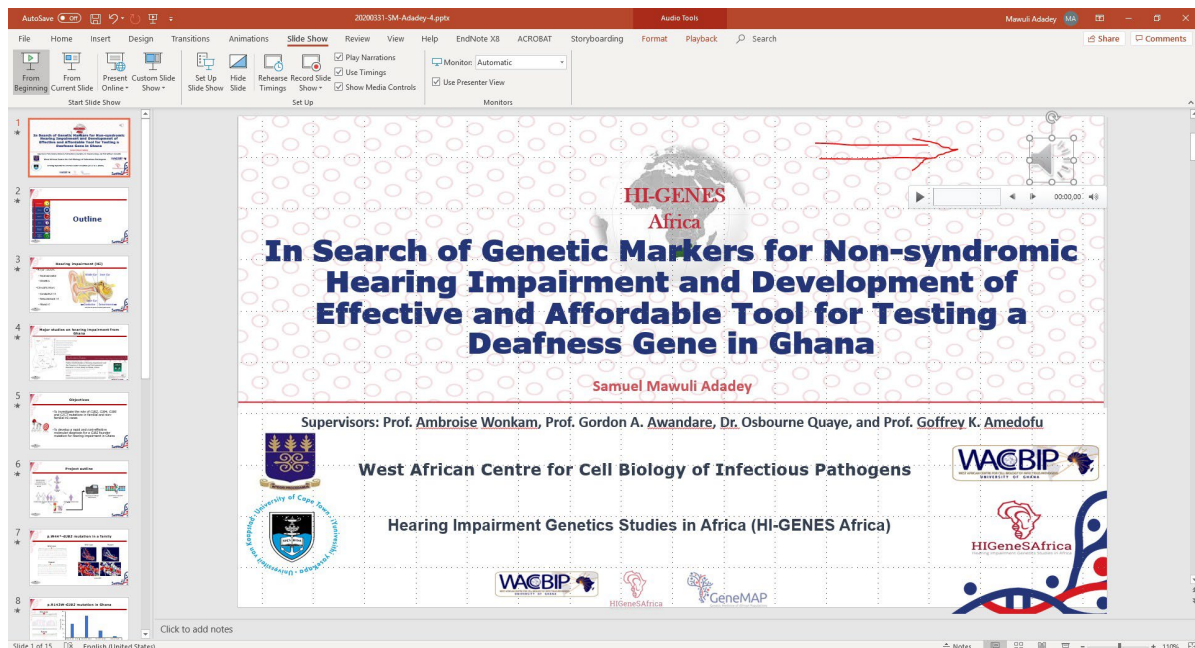
Manage recordings

There are different ways to manage recordings in your presentation:

- **Pause** - to pause a recording.
- **Close** - to end a recording.
- **Laser Pointer, Pen, Highlighter, or Eraser** - to use the pointer, ink, eraser, or highlighter tools in your recording.

Remove recordings

A sound icon appears on a slide to show when a recording is available.



You can also remove recordings from slides.

1. Select **Record Slide Show > Clear** (on the right corner).



2. Choose from four options:
 - **Clear Timing on Current Slide** - to delete timings on the current slide.
 - **Clear Timings on All Slides** - to delete timings from all slides at once.
 - **Clear Narration on Current Slide** - to delete narration on the current slide.
 - **Clear Narration on All Slides** - to delete narration on all slides at once.

Save recordings

When you're done recording, save and share your presentation as a PowerPoint Show. Your recording will automatically play when someone opens the presentation.

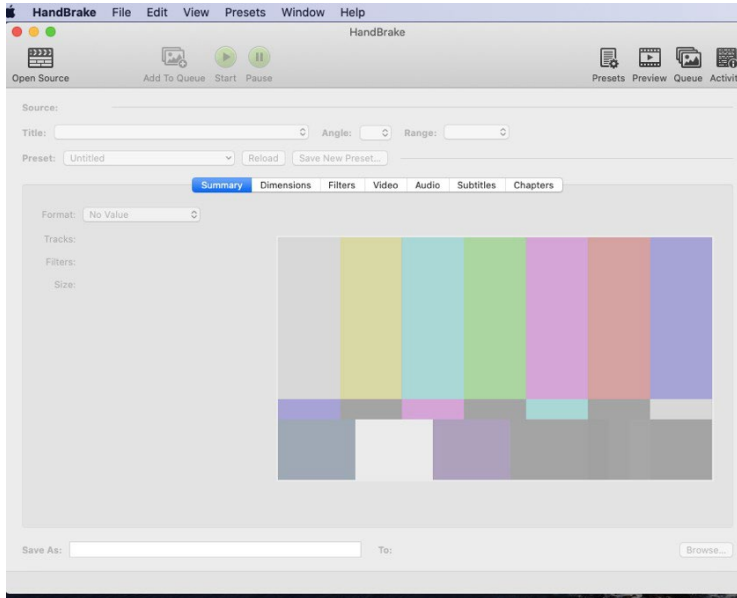
1. Select **File > Save As**.
2. Select where you'd like to save your presentation to.
3. Under **Save as type**, select the dropdown arrow and then select **PowerPoint Show**.
4. Select **Save**.

Source: <https://support.office.com/en-us/article/video-record-presentations-2570dff5-f81c-40bc-b404-e04e95ffab33>

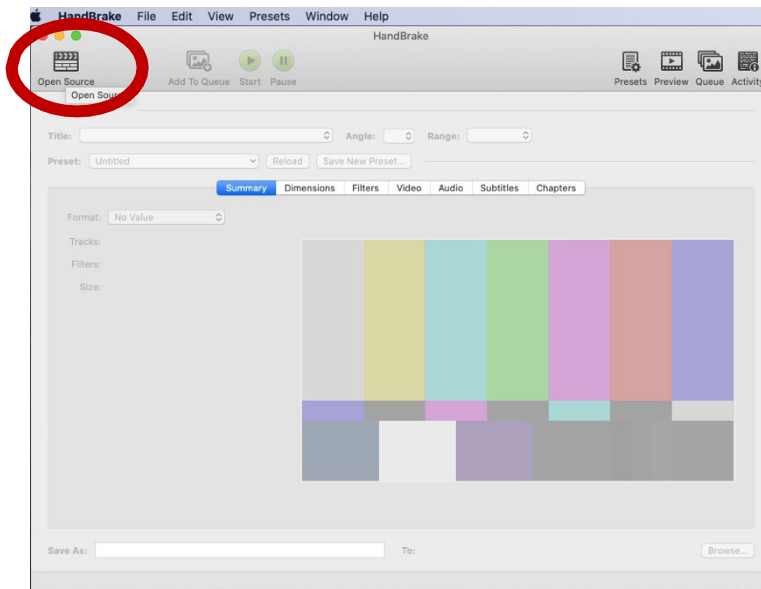
How to resize your recorded slides

Download a HandBrake version that is compatible to your pc

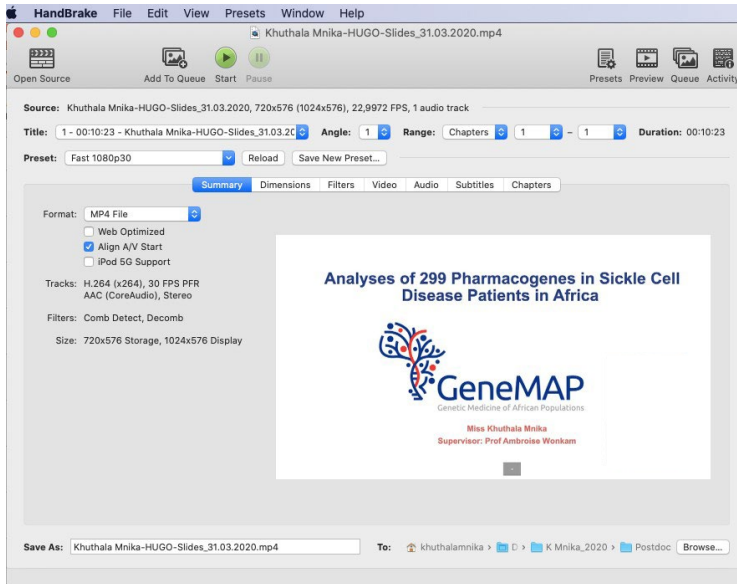
1. **Open** HandBrake



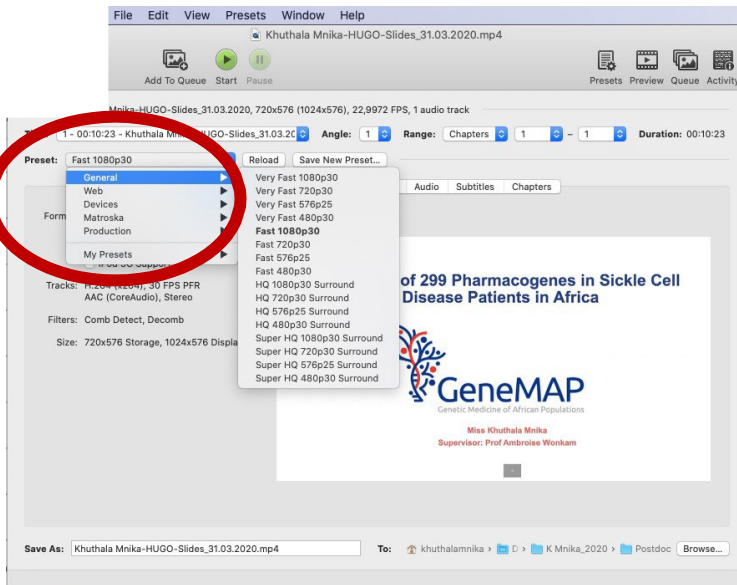
2. **Click open source** and select your saved mp4 recoded slides



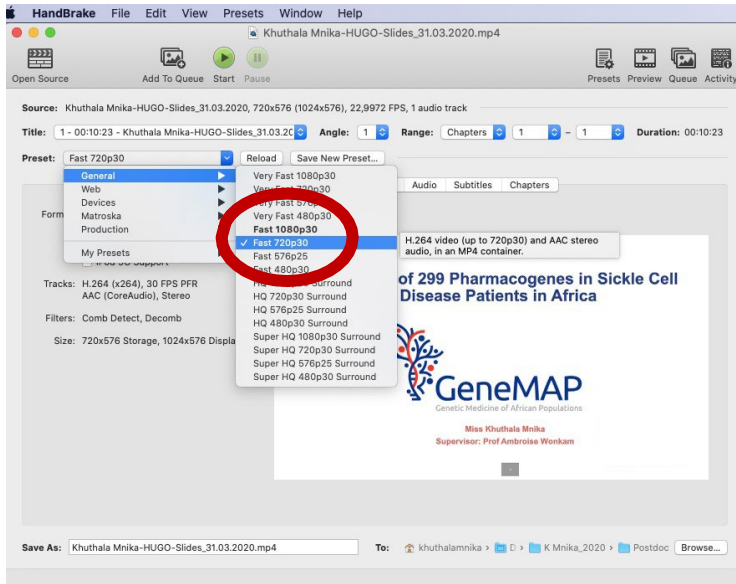
3. Output after selecting your mp4 recorded slides



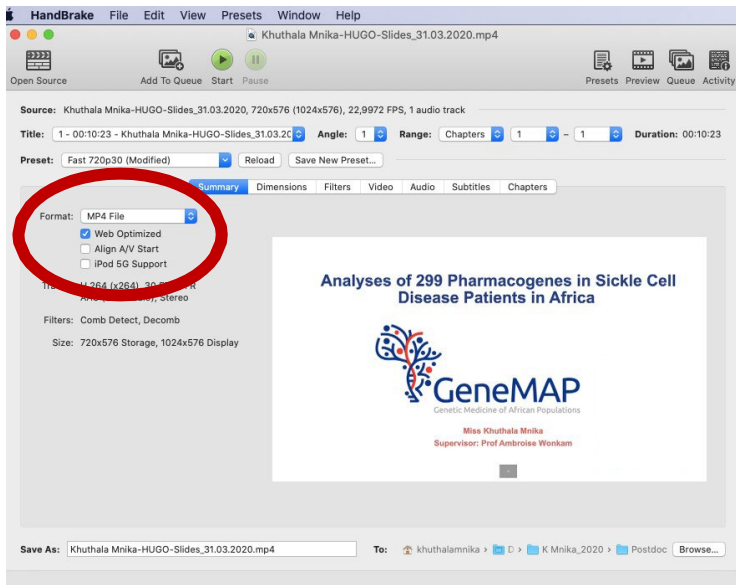
4. Click preset dropdown arrow then select general.



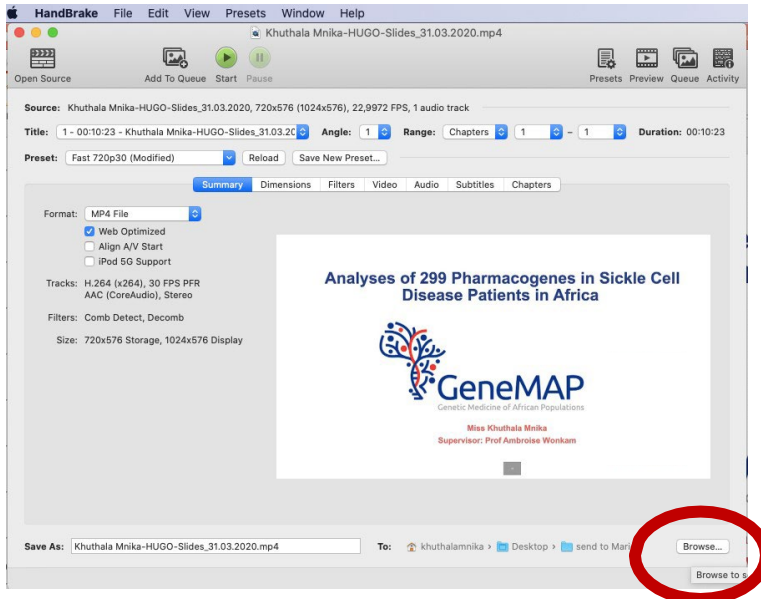
5. Select Fast 720p30



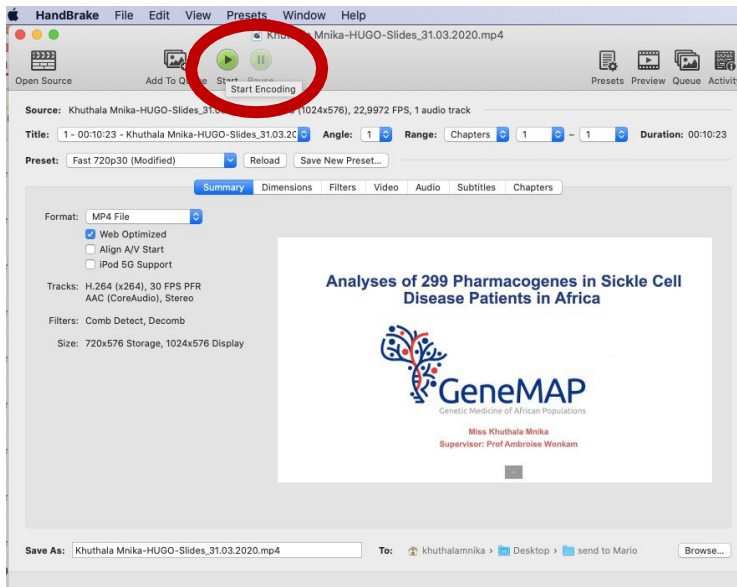
6. Select web optimized under Format



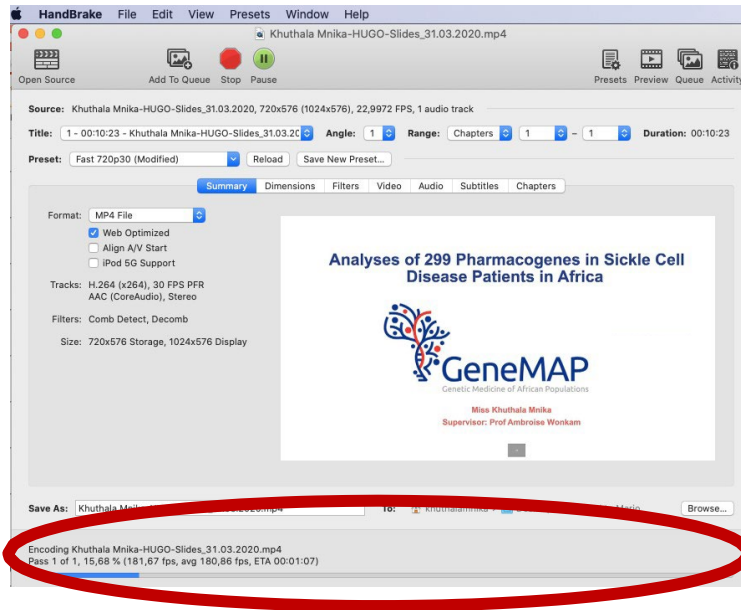
7. **Click** browse to select destination path for the encode (resized video)



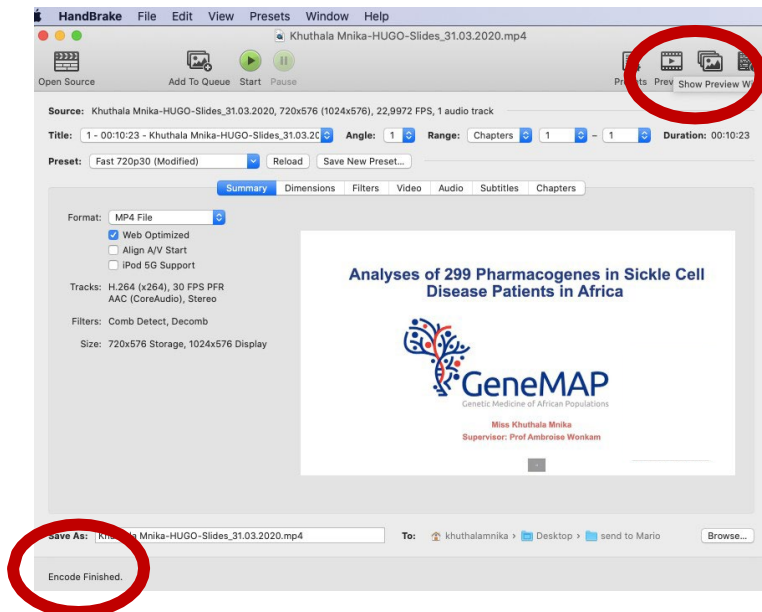
8. **Click** start, to start encoding



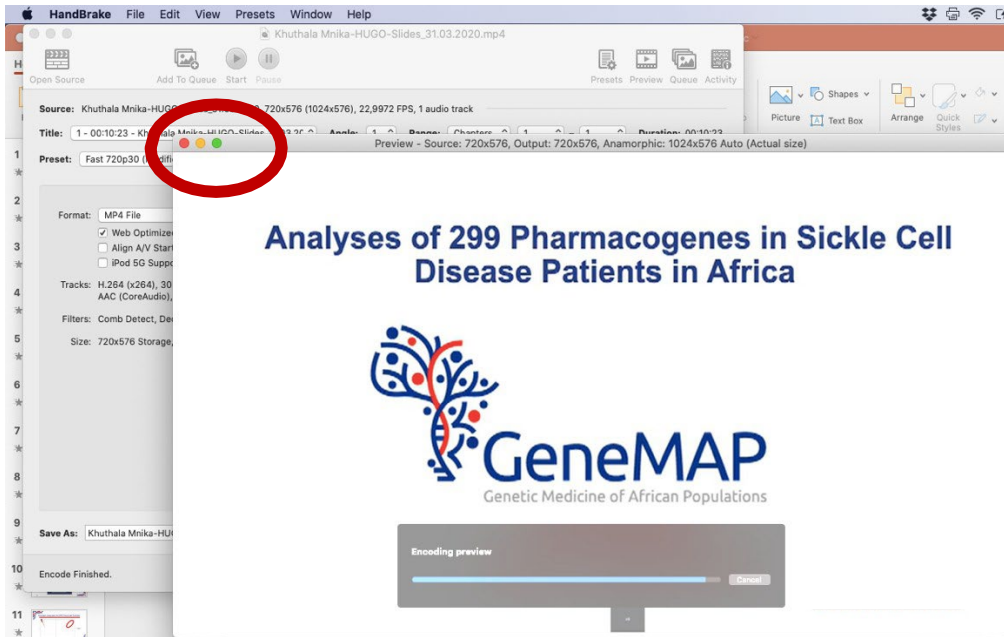
- After encoding is done, it will automatically save your video in your selected destination



- At the bottom corner you will see the status of the encoding. **Click** preview to check your video quality



11. When you are happy with the quality you can **close** the window



Well done! You have successfully recorded and resized your slide video

Version No.	Date	Internal Reviewer(s)	Author	Details of changes
1.0	21 April 2020	Mario Jonas	Samuel Adedey, Khuthala Mnika	None, first version
	15 March 2023		Mario Jonas	Added very brief intro, section headers and page numbers

Sign up for an Ilifu account.....	1
Getting your SSH key.....	1
For Linux.....	1
For Windows.....	2
Starting a collaborative RStudio session on Ilifu cluster.....	3

[Source: <https://docs.ilifu.ac.za/>]

Sign up for an Ilifu account

Although this is the first thing to do, signing up depends on the “Getting your SSH key” section below

https://docs.google.com/forms/d/e/1FAIpQLSfHtL7R_S2HHlecFeZn8sicABdvFO7NZC1TRJdU2S5GSL4zYQ/viewform

Getting your SSH key

When registering for an Ilifu account, you will be prompted for a public SSH key. Below are the instructions on how to obtain that public SSH key

For Linux

[Source: <https://docs.github.com/en/authentication/connecting-to-github-with-ssh/generating-a-new-ssh-key-and-adding-it-to-the-ssh-agent>]

1. Open Terminal.
2. Paste the text below, substituting in the email address you are using to register for an Ilifu account.

```
ssh-keygen -t ed25519 -C your_email
```

Note: If you are using a legacy system that doesn't support the Ed25519 algorithm, use:

```
ssh-keygen -t rsa -b 4096 -C your_email
```

This creates a new SSH key, using the provided email as a label.

3. Generating public/private ALGORITHM key pair.
 - a. When you're prompted to "Enter a file in which to save the key", you can press Enter to accept the default file location. Please note that if you created SSH keys previously, ssh-keygen may ask you to rewrite another

key, in which case we recommend creating a custom-named SSH key. To do so, type the default file location and replace `id_ssh_keyname` with your custom key name.

- b. Enter a file in which to save the key

```
(/home/YOU/.ssh/ALGORITHM): [Press enter]
```

- c. At the prompt, type a secure passphrase. For more information, see "Working with SSH key passphrases."

```
Enter passphrase (empty for no passphrase): [Type a  
passphrase]
```

```
Enter same passphrase again: [Type passphrase again]
```

For Windows

[Source: <https://www.purdue.edu/science/scienceit/ssh-keys-windows.html>]

1. Make sure you have OpenSSH installed
 - a. Open Windows start menu
 - b. Search for "Apps & Features"
 - c. Under "Apps & Features", click "Optional Features"
 - d. Confirm that "OpenSSH Client" is in the list of optional features
 - e. If not, click on "Add a feature", select "OpenSSH Client" and click "Install"
2. Create the SSH key in Windows 10 and 11

There are several ways of creating SSH keys in Windows. I will just focus on one method here.

- a. Open up the Start menu and type "*cmd*"
- b. If given several options, select "Command Prompt"
- c. In the command prompt, type the "*ssh-keygen*" command
- d. You will be prompted for the location to save the file to. The default location offered to you is generally ok since it saves it in your home directory
- e. *ssh-keygen* will generate the files mentioned above, as well as a key "fingerprint" and a randomart image. These will be rarely used but it can be saved into a file if you so wish.
- f. In the file explorer, navigate to your home folder to find your SSH folder. Windows has "hidden" files for configurations and various other purposes. The SSH config file is one such file.
 - i. In file explorer, click on the "View" menu item
 - ii. In the dialog box opening up, select "Hidden items" from the menu items on the far right
 - iii. Select the ".ssh" folder

- iv. You should see two files, “*id_rsa*” and “*id_rsa.pub*”. The latter is the public file that you would share with collaborators if you are asked. The contents is what you provide Ilifu during the registration process
 - g. More
3. More

Starting a collaborative RStudio session on Ilifu cluster

[Source: https://docs.ilifu.ac.za/#/tech_docs/software_environments?id=running-rstudio-server]

The instructions below are relatively simple, and the messages returned to you by the server is usually informative enough so that you can proceed to the next step.

1. Log into the Ilifu cluster with your user details
2. Once logged in, start an interactive job

```
USERNAME@slurm-login:~$ srun --nodes=1 --tasks=1 --mem=8g --time 08:00:00 --job-name="rstudio test" --pty bash
```

The command above says to “...start 1 computer node with 8G RAM to run for 8 hours and name the session, ‘rstudio test’...”

3. Add the RStudio module from the list of all modules/ programs available

```
USERNAME@slwrk-101:~$ module add R/RStudio1.2.5042-R4.0.0
```

4. Now start the Rstudio session

```
USERNAME@slwrk-101:~$ rstudio
```

If things went well thus far, you should receive the information below:

```
The environment variable RSTUDIO_PASSWORD was not set, so your password has been chosen for you. It is:  
lMoV0akqQu6V21DGulDUMyYZGpjmoL
```

```
Running rserver on port 43627
```

```
To connect to this server run this on your local machine:
```

```
ssh -A USERNAME@slwrk-101 -o "ProxyCommand=ssh  
USERNAME@slurm.ilifu.ac.za nc slwrk-101 22" -L8081:localhost:43627
```

```
Then visit http://localhost:8081 in your browser and use the username "USERNAME" to login with the password "lMoV0akqQu6V21DGulDUMyYZGpjmoL"
```

(You may need **to choose** a different port (other **than 8081**), so remember **to change** this **in both** the ssh **and** browser

5. Once RStudio is open in your browser, follow the instructions given you when you started the session above by providing your USERNAME and password to log in
6. At this point, you should have an Rstudio session running in your browser
 - a. This RStudio instance is running on your local machine
 - b. And has access to the processing power of the cluster that you logged into
 - c. As well as the data residing on the selfsame cluster infrastructure

Document version

Version No.	Date	Authorizer	External Reviewer	Internal Reviewer	Author	Details of changes
0.1	20 September 2023	Victoria Nembaware	Reuben Chianumba, Daniel Kandonga	Jack Morrice	Mario Jonas	n/a first version