**Documentation Workflow Checklist**

# Pre-Project

* Decide on the software tool to use. You can either use your own Collections Management Systems (CMS), an excel spreadsheet, or a third-party tool. Each have pros and cons.
* Using your own CMS

+ Data quality tools e.g. lookups and validation

+ Support multiple simultaneous users

+ More guided data entry than a spreadsheet

* Access may be tightly controlled, and this can add to the training/onboarding costs for the project
* Not being able to adopt a ‘bring your own device’ model, or not having a web interface can limit the number of project users
* It can be difficult to find the necessary fields to record transient, project specific information
* Higher training threshold with many functions unnecessary for given project
* Other commonly used tools such as Microsoft Excel

+ Very familiar

+ Can add lookups and validation

+ Cloud-based versions can support multiple simultaneous users

* No way to guide user beyond column-by-column approach
* Risk of losing row integrity (careless deletes can cause unintended shifts in the data)
* Difficult to integrate images/image viewing
* Using third-party no/low-code solutions such as AppSheet (a Google product) or Powerapps (from Microsoft)

+ Can add lookups and validation

+ Multi-user support easy to implement

+ Can build guided data entry interface

+ Image handling easy to implement

+ Knowledge threshold to build powerful tools much lower

+ Can use choice of familiar software as underlying data sources (e.g. Excel, GoogleDrive, etc.)

* IT policies may preclude their use due to data access restrictions
* Costs often based on number and frequency of use which can be hard to predict in volunteer projects

**Initiating The Project**

 Create a data dictionary.

* This will allow you to record and manage the output from the initiation stages.
* The results of the following points should be recorded in this document which will become the specification from which to build the data capture solution.

 Consult with stakeholders.

* Engaging with stakeholders will allow you to collaboratively establish what information the project aims to gather.
* The goals should be realistic and achievable. Setting unrealistic goals can result in asymmetrical records, with records made early in the project being very detailed and later records being very sparse as time runs tight and data is dropped from the remit to meet deadlines.

 Establish the relationship between the physical archive and the data.

* For example, are the items boxed or aggregated in some way and is there information on/about these entities that should also be captured.
  + Note this is different to the total number of *items* you wish to process, rather the number of *types* of items.

 Decide any additional information required by the project.

* For example, it may be useful to add a set of fields to record an item's status in the workflow, as well as corresponding date and note fields. A simple status of pending/in progress/complete/problem can help track progress and identify issues that are holding up the process.

 Ensure who created records, and when, is captured.

* Recording this information is important when there are issues in the captured data and it’s necessary to identify who captured this data and when. It can suggest where more training is needed or identify deficiencies in the data capture protocol.

 Identify who will be the authority for each piece of data.

* This person will be responsible for fixing discrepancies in the captured data, standardising and/or normalising where necessary.

 Decide on how to approach gathering/cleaning of data.

* Common approaches to data which includes the names of people, organisations or places, are:
  + Capture verbatim by users and standardise as a separate task before repatriating the data to the host repository.
  + Provide a controlled list to select from (accepting that continual additions may be needed to this list).

 Map the data requirements to the relevant fields in the repository.

* As part of this process you will also identify information that will not return to the repository.

 Enhance data dictionary.

* This is done by recording data types, identifying relationships between data, default values, picklists and opportunities for validation.
* Where possible, utilise the same validation/term-lists as the repository to ensure data are aligned.

 Consider which information is mandatory.

* Discussion with those most familiar with the materials to be captured will help understand the completeness of the information and where it may be impossible to always record a value.

 Consider whether you will record the absence of information.

* Explicitly recording that information is not present can reduce confusion as to whether information was missed or absent, and prevent having to re-check work.

# BUILDING FIRST DRAFT, ITERATING AND IMPROVING

 Understand the physical workflow of the data.

* This is crucial when developing the data capture workflow so that the user has the right fields in front of them at the right time.
* Getting this right will both reduce handling of items and reduce the handling time.

 Establish the priority order for your data.

* If there are multiple entities that need data captured (e.g. aggregating entities like boxes or folders), which will come first, second, third, etc.
* Where there are aggregating entities in the physical arrangement, it is helpful to capture their details first, which then allows items they organise to be associated and, if necessary, data can be propagated.

 Choose a value that is unique to each record.

* Alternatively you can add a field that will generate a new unique key for each record.
* Adding a new key value that is non-meaningful in any way other than being unique can be very helpful. Whilst *theoretically* unique, object numbers often aren’t and having a separate identifier avoids the problems of overlapping records this would otherwise cause. Though if object numbers are supposed to be unique, it is then advisable to build in a method for flagging where this is not the case. This value often does not need to be used beyond the project so does not necessarily need to be repatriated.

 Look for opportunities to automate repeatable information.

* Examples of automatable categories are: usernames, record create date, record modification date.

 Consider the visibility of fields and if they are conditional.

* If a certain field or fields are only necessary when a given value is selected by a user in another field, it can be helpful to hide unnecessary fields until this condition is met, reducing user confusion/compulsion to enter data in every visible field.

 Ensure fields holding mandatory data are clearly indicated.

* It is important to ensure that all fields for mandatory data can be fully completed.

 Include links to useful documents for end users.

* This could include ‘How to’ documents on using the data capture tool, but also shared notes or links to sources of information pertinent to the project. This helps collaborative working.

 Pass to the end users for feedback.

* Having a tool that is both flexible and powerful allows the data capture solution to be very closely tailored to the specific needs of the project, but even more so, the tool can be tailored to the specific set of users.

 Add elements that were not specified or changed.

* Despite best efforts at anticipating what will be needed, once the data collection begins there will likely need to be additions or changes where

things work in a different way to anticipated (e.g. materials of a markedly different size, or with obscured elements, folded sheets/overlaps, may need to be imaged differently, and thus must be processed outside of the general order).

* If new concepts are added, ensure that records created prior to their inclusion are aligned to the new requirements.

 Ensure regular communication with users.

* A running and responsive dialogue between users and those responsible for the data capture tool is important. Without it, users who experience difficulties with the tool may lose motivation with the project or start using fields and recording information in non-standard ways, which will cause issues when closing the project and repatriating the data.

# Closing Down The Project

 Check for completeness of all records.

* One of the advantages of using a no/low-code solution with a spreadsheet as the underlying data source is that this checking can be done in the spreadsheet, where it is easy to see missing data column by column.
* Using a workflow status tracking system can be useful here, to quickly identify records that have not been marked as complete, but all records should be checked.

 Standardise data where necessary.

* If there are data that need to be standardised before repatriating to the eventual repository (e.g. people or places), these should be passed to the data owner.