Monash University Research Data Archive (MURDA) Service Overview

Enquiries: murda@monash.edu

Acknowledgements

Monash eResearch Centre

- Stephen Dart
- Steve Quenette
- Adrian Tritschler

Monash eSolutions (Records Management)

- Sandra Ennor
- Catherine Nicholls

Monash University Library

- Neil Dickson
- David Groenewegen

Table of Contents

Overview

Introduction

The MURDA service is broken down into seven core functions, each of which has an accompanying process as part of the overall workflow. These functions encompass the various stages of the data management lifecycle and relate to:

1. The identification of research collections that should be archived,
2. The analysis of research collections,
3. The ingestion of data into the Archive,
4. The sentencing of ingested collections,
5. The preservation of collections with long-term or permanent retention periods,
6. The provision of access to ingested collections, and
7. The deletion of archived collections.

Figure 1 – MURDA Core Functions

Identification

When the MURDA was first created, the lack of data provenance was identified as a key issue. As a result, there was a need for a consistent and unambiguous process for identifying potential candidates. To that end, a specific definition for collections with unknown provenance (orphaned collections) was created along with four specific criteria:

1. There is no known owner and there are no active users,
2. The owner and/or users are known but have left Monash,
3. The groups (either local or network) that control access to the collection (if any) are empty or only contain staff members that have left Monash,
4. The collection contains data that has not been changed for at least twelve months and the owner and/or users cannot be contacted.

Orphaned collections usually lack an easily identifiable owner and older collections, often those from decommissioned or end-of-life services, may lack useful metadata of any kind. The lack of good metadata makes assessing the value of data considerably more difficult and, without useful metadata and the context that it provides, informed decisions cannot be made. In such cases, collections must be identified and analysed in order to allow stakeholders to determine whether or not they should be retained. Rather than remain in their original location thereby reducing the available capacity (often disk space) for other users, placing such data within an archive is a better option. Storing data in the MURDA not only allows it to be curated appropriately, but also places it in a centrally-managed location with metadata records and frees up storage resources that can be used for other purposes.

The MURDA process for identifying potentially orphaned research collections is not exhaustive, because different storage services have different tools and logging capabilities depending upon their configuration. As a result, some services lack useful system-level metadata which is most often the case when research groups possess internal storage solutions such as NAS (Network Attached Storage) boxes, often for historical reasons. Such devices are often administrated by a single individual and are therefore not always properly maintained or supported in the long term. The lack of technical metadata may therefore necessitate the acquisition of information from non-technical sources. Within Monash, non-technical sources generally include eSolutions, professional networks, email records, help desk tickets in various systems and organisational knowledge. If a collection's data owner or a contact person can be identified, then an informed decision can be made and assessing the value, status and retention period of a collection is a much easier task. If no information about a collection or its owner can be acquired through the MURDA's identification process, the collection must be examined and a thorough analysis of the collection is required. This is outlined in detail as part of the MURDA analysis process.

Analysis

Once an orphaned collection has been identified and the decision has been made to store the data within the MURDA, the collection must be analysed. The analysis process provides information about what the data is (even if the information is limited and very basic), which is essential in order to enable informed decision making by stakeholders. The process is also required for data retention and sentencing as well as for the creation of a metadata record. The MURDA process for analysing data collections is relatively straightforward, but necessitates a working knowledge of many different data types, file formats and research workflows across disciplines. Because the information gathered may be ambiguous, the

results of an analysis may be abstruse and this should be considered when sentencing data. Further, as the analysis process is currently a manual one, it does not scale well and substantial resources may be required for large collections.

Although most of the MURDA's processes are currently manual, ideally the various functions would be automated as much as possible. This has proven to be problematic however due to the lack of effective tools. It should be noted that while there are applications that can scan through filesystems and generate useful statistics, it is difficult to determine exactly what kind of data a collection contains without domain-specific context. A collection of PDFs for example, could be anything from a collection of research papers used by a PhD candidate for their literature review, to scanned patient records and medical histories used in a clinical trial. The statistics produced by commercial products cannot provide such context and this is further complicated by organisational and legislative privacy policies that prohibit the analysis of file content except under specific circumstances. Due to this, a manual assessment of orphaned collections is required in most cases in order to collect useful information. A recent proposal has been developed that aims to address the current need for manual analyses, as well as combine and automate other MURDA processes. This is a long-term project which will take some time to investigate and implement.

Ingestion

The primary purpose of the first two processes is to identify and analyse orphaned collections. These two processes are not always necessary however, because if the data owner is known then they can confirm that the MURDA is an appropriate option for the data and can provide detailed information about the collection's content and retention periods, etc. As such, the ingestion phase and the remaining MURDA processes are applied to all collections, not only those that are orphaned.

Once a MURDA candidate has been identified and information has been acquired from either the data owner/contact or from a collection analysis, the information is used to create a metadata record. Once created, any existing storage allocations, export lists and/or secure access groups are updated. The data is then bundled up if appropriate and stored in the MURDA following an established naming convention. The data is verified as part of the process and the original copy of the data is deleted. The naming convention used by the MURDA is mainly historical in nature, as it predates more contemporary services like the Data Dashboard which now hold collection metadata and which did not exist when the Archive was created. This meant that embedding metadata in the file/directory name was a necessity due to the lack of a centralised system and local resources.

Sentencing

Sentencing collections stored within the MURDA is one of the most important aspects of the service. "Sentencing" is a term primarily used in records management and is officially defined as the process of matching information held by the organisation to a specific class of a records authority. In practice, it is a data assessment process within the MURDA that uses PROV (Public Record Office Victoria) research data categories to determine the period for which a collection should be retained. MURDA sentencing relies upon the expertise of staff in the Records Management team, who have the authority to apply sentences (retention periods) and, more importantly, can authorise the disposal (deletion) of collections where appropriate. It is important to note that the aim of the sentencing process is not to dispose of data, but to ensure that stakeholders have an opportunity to assess data for disposal once a

collection has reached the end of its retention period. At that time, the data can be retained for a longer period or deleted as appropriate.

Identifying, analysing and ingesting collections into the MURDA is naturally important, however it is the sentencing process that differentiates the MURDA from more traditional digital archives. While storage archives generally succeed in storing old data, they tend to only increase in size over time because the data inside them is rarely deleted. The operators of such archives assume stewardship, however without domain knowledge and due to the presence of orphaned data, stewards generally take the view that data within the archive "might be needed" or should be retained "just in case". This perspective is mirrored by many researchers when they consider their own data, which results in even more data being archived.

The end result is that proper digital curation within storage archives is often minimal, which leads to increased storage costs. Storage, regardless of the media used, costs money and even cheaper tape-based storage has operational and running costs that need to be considered. Therefore, implementing a sentencing regime allows organisations to save money by only storing data that is useful, or that they are obliged to retain for compliance or legislative reasons.

Sentencing data is not only financially beneficial but can also help to lower the level of risk. Retaining data that should have been deleted has the potential to increase risk, especially if the data is of a sensitive nature or should have been deleted in line with third-party or ethics agreements, etc. While there is always a risk when deleting data that "may be needed", there are also governance frameworks that specifically outline the duration for which specific data types should be retained (excepting agreements or special arrangements). These frameworks are very useful not only as guidelines, but because they can protect organisations in case the decision to delete is challenged. For the MURDA, one of the most useful applications of the PROV categories within the sentencing process has been peace of mind and the effective management of risk.

While orphaned data is common, even data collections with clear provenance can be problematic. Researchers are often uncertain which data they should and should not retain, and are often concerned that their data may be required if their work is audited. By including staff that have the ability to formally sentence data for retention or disposal through the application of a governance framework (the Records Management team in eSolutions), the onus of responsibility moves away from the researcher to specialised staff. Those staff members then take responsibility and ownership of the decision to retain or delete data where appropriate, which provides the researcher with support in case of challenges or disputes, as well assurance that their data is being managed appropriately.

One of the most problematic aspects of the sentencing process is assessing the value of a collection. Determining the value of data is one of the most challenging steps, even for collections with good metadata, because the value of a collection is rarely clear without detailed information about the data and discussions with the owner. There is a level of risk when determining the value of a collection without knowing the specific compliance framework(s) and data agreement(s) to which the collection may be subject, therefore when sentencing data it is often wise to err on the side of caution and select a longer sentence if in doubt.

Generally, without any other information (as is often the case with orphaned collections) the value of data is indicated by the amount and quality of a collection's metadata. In an ideal scenario, the determination of a collection's value would be the result of liaising directly with the data owner, assessing the available metadata and considering any data encumbrances. If the owner of a collection cannot be identified, then metadata is usually acquired through the MURDA analysis phase but the results of the analysis may be ambiguous. For example, if an orphaned collection contains a single directory called "iTunes Library", then an initial assessment would suggest that the collection is likely to be of little value to the organisation. Due to the complexity of research data and the myriad ways in which researchers conduct their work however, this may be an erroneous assumption. Is the data a backup of a user's personal iTunes library, or does it contain data produced by a researcher as part of a specific project that is stored alongside the tool (iTunes) that they used? Could it be both? Due to the prevalence of such scenarios (the above is a real example), it is incumbent upon those conducting data sentencing to prudently analyse data for its possible value and level of risk. This is true even for collections that may initially seem to have little value and are believed to be low-risk.

Preservation

As with any service that has a mandate to retain data for legislative compliance, the MURDA has a process in place to support the preservation of collections with long-term or permanent data retention periods. Research data is often highly complex and exists in many formats, both open and proprietary. The latter is of particular concern with regard to digital preservation, due to the possible lack of support in the longer term. File formats tend to evolve and change over time, and while open formats can be reused and modified by the community, proprietary formats are usually only maintained and supported by commercial organisations as long as it is financially beneficial to do so. This issue may be compounded by the inability of researchers to access their data in one of these formats without the use of specialised applications and tools. These tools are often subject to restrictive and expensive licences, which presents financial barriers and may make the ongoing use of old data in unsupported formats unsustainable. As with the file formats themselves, the long-term maintenance and support of such tools is also uncertain. In some cases, often when specific instruments and technologies are in use, data is not only generated in a proprietary format which necessitates the use of specialised tools, but the tools themselves cannot be used without the presence of physical hardware such as dongles. As a result, digital preservation can pose a significant challenge when such data needs to be archived.

The existing MURDA preservation process is quite straightforward because Monash University is currently developing a governance framework for digital preservation. When the framework is finalised, it will be used to update and refine the MURDA preservation process. Generally, MURDA collections are assessed for preservation based upon their retention periods. If long-term or permanent retention is required, then factors such as the data format(s), type(s), specialised tools and other factors such as compliance are considered, and the collection is updated and re-packaged as appropriate. Currently, MURDA collections that need to be retained for at least ten years are subject to the preservation process, but the need for preservation is considered when any collection is ingested into the Archive.

Access

In order to ensure that collections are stored securely, the MURDA was designed to be a closed environment. As a result, end-users and other unauthorised individuals cannot

access MURDA metadata records, or the collections to which they pertain, without authorisation from the relevant data owner or MURDA administrators. The primary reason for this is that it is not always possible to identify exactly what kind of data an orphaned collection may contain, or the security classification of that data. As a result, once a collection is archived and placed in the MURDA, the stewardship of the data changes and access is restricted to not only ensure that there is no unauthorised access, but to assist in complying with any data security requirements. The inability of end-users to access MURDA metadata or collections directly also helps to ensure that the data and records have not been altered and have retained their integrity (excepting bit-rot, digital preservation concerns or other issues).

The MURDA access process is a formal one that requires approval from the data owner. If the data owner is unknown or cannot be contacted, then the relevant Head of School or equivalent is consulted. Once an access request has been approved, a copy of the requested data is made available through an appropriate medium depending upon the sensitivity of the collection. It is also important that MURDA administrators can audit access requests to collections upon request, therefore all access requests are documented regardless of whether or not access to the data has been granted.

Deletion

Although the primary purpose of the MURDA is to store data and retain it in the long term, it is important to note that there is no benefit in retaining data that has been determined to have no value or that is stored elsewhere (except in specific situations). One of the primary benefits of deletion is the reduction of storage costs, something that should be considered even when cheaper tape-based storage is used. It is also worth noting that the removal of unneeded data reduces administrative overhead and makes large archives more manageable. In addition, retaining collections for longer than their specified retention periods may actually increase the level of risk to the organisation, which is especially true for collections that are subject to data agreements and those provided by external parties. Medical data is one area that requires special consideration in this regard in order to ensure that Monash adheres to state and federal legislation. Furthermore, the need to comply with international frameworks such as the GDPR also indicates the need to delete data where appropriate, wherever it resides. As a result, compliance with legal and academic frameworks should be considered at all stages of the data management lifecycle. It should also be noted that researchers are not always aware of the compliance frameworks that concern their data, therefore when a collection needs to be archived an excellent opportunity presents itself to explore such considerations with the owner prior to ingesting the collection.

With regard to the deletion process, it is worth noting that it is not an automatic one and that each collection is assessed independently during a formal review. Data is only removed when a sentenced collection has passed its retention period, has been reviewed in consultation with the data owner/stakeholder and deletion has been authorised. It is through negotiations with the data owner or equivalent that a decision is made regarding the future of a collection, which can result in a longer retention period or the deletion of the data. In the former case, the justification for a longer retention period is considered by MURDA administrators and Records Management staff and a new retention period is negotiated. If the rationale behind the retention adjustment does not warrant the full duration of the requested extension, short-term increases are implemented in consultation with the

stakeholders. Once all parties have agreed to the new retention period, the EDRMS (Electronic Data and Records Management System) and MURDA metadata records are updated and the collection is retained until the next review.

If a collection is no longer required, then formal approval to delete the data is requested which is then documented alongside the reason for deletion. The data is removed from the system after which the stakeholders are informed and the various metadata records are updated. Due to the need to track and audit the Archive, the metadata records for all MURDA collections are retained, even if the collections have been deleted. This provides an historical view of the service and also helps to ensure that all transactions have been properly documented.

Orphaned Research Collection Identification Process

Process

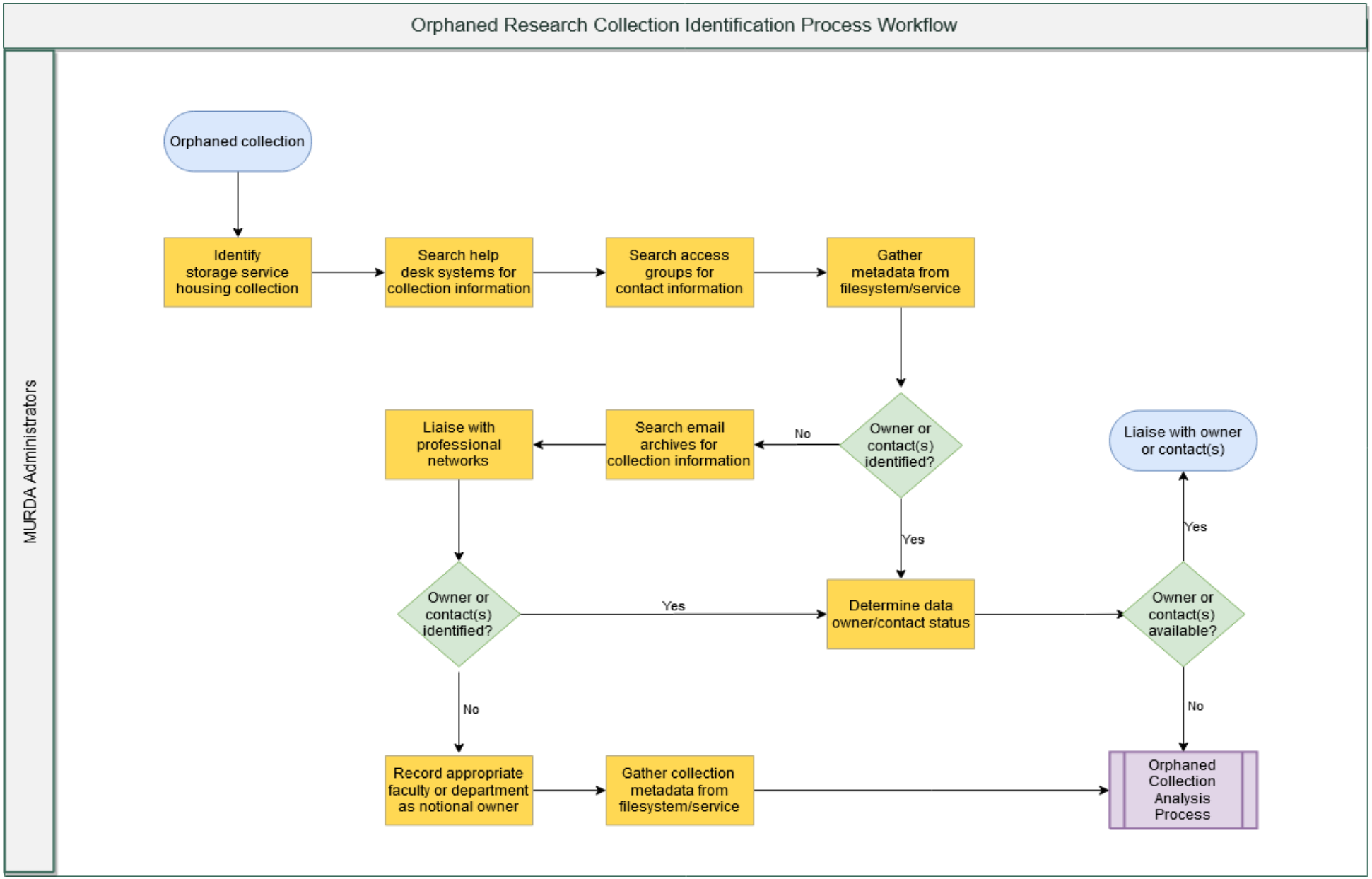
A collection is considered to be orphaned if it fulfils one or more of the following criteria:

1. There is no known owner and there are no active users,
2. The owner and/or users are known but have left Monash,
3. The groups (either local or network) that control access to the collection (if any) are empty or only contain staff members that have left Monash,
4. The collection contains data that has not been changed for at least twelve months and the owner and/or users cannot be contacted.

In order to determine whether or not a collection fulfils any of the above criteria, the following questions and actions are considered:

1. Which storage service/infrastructure houses the collection?
2. Can any documented engagements or reported issues be located within any of the relevant help desk systems?
3. Can any references to the collection be located by consulting email archives?
4. Do staff have any contact people within their professional networks that may know something about the collection?
5. Does the collection have any associated groups?
    a. If so, acquire a membership list and confirm that the users (if any) are still at Monash.
6. If the service's filesystem is directly accessible, is there any directory or file ownership information (POSIX UIDs or GIDs, etc) that can be linked to specific individuals or Monash IDs?
7. If the storage is mounted on another service (e.g. a Nectar VM), who was the last person to log into the machine and are there any regular users?
8. Is any server ownership or hosting information available through eSolutions online tools?
9. If the collection is exported to another machine/server (typically via NFS but other protocols such as Samba/CIFS may also be used in some cases), are there any requests for server access or any hostname information?
    a. If a hostname is included in the relevant export, can any machine/server ownership or hosting information be located through the eSolutions online tools?
10. Are there any logs relating to the service in question?
11. When was the last file modified?
12. Are there any other indications of use or access?
13. If the data owner or contact has been identified, they should be contacted and an assessment should be made in order to determine if the collection is a MURDA candidate.
    a. If the MURDA is an appropriate location for the data, staff should liaise with the data owner/contact and arrange to ingest the collection.
    b. If the MURDA is not an appropriate location, staff should liaise with the data owner/contact, update the collection metadata and look into migrating/deleting collection if appropriate.
14. If the data owner or contact cannot be identified, the collection is considered orphaned and the data should be ingested into the MURDA.

Orphaned Research Collection Identification Workflow Diagram



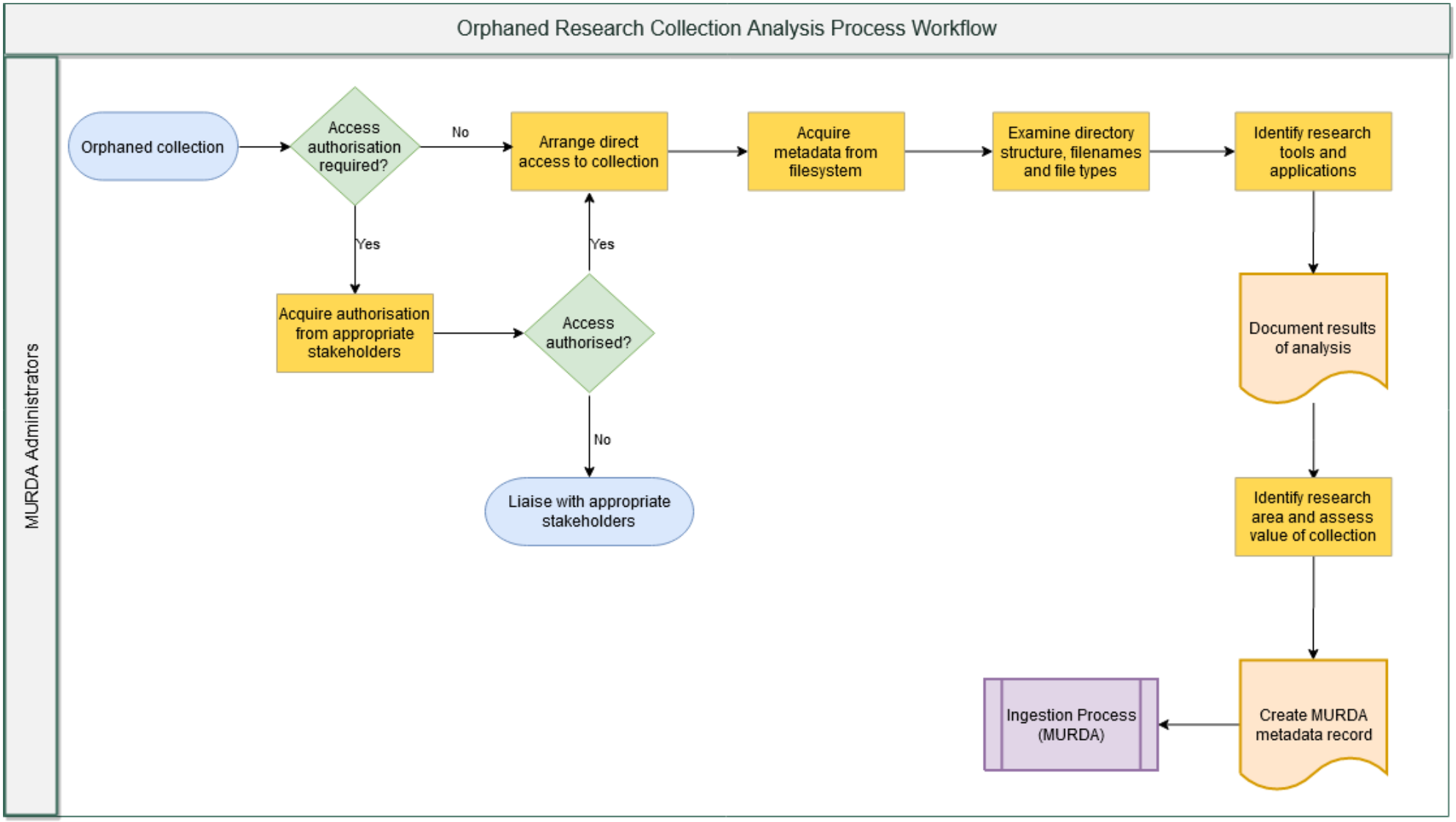
Orphaned Research Collection Analysis Process

Process

The objective of the analysis phase is to acquire information about the data (sizes and dates, etc) using system tools and then examine the files and directories within an orphaned collection in order to gain some understanding of the research area. This information, among others, is then used to create a metadata record. The general approach used when analysing collections is to:

1. Request authorisation to access the orphaned collection/storage service if appropriate.
2. Arrange direct access to the orphaned collection by adding staff to the relevant secure access groups or IP/hostname export list, etc or gain access to the underlying storage service's filesystem.
3. Using standard filesystem commands appropriate to the operating system in question (dir, ls, du, find, etc), generate information about the collection if not already known. Information should include:
   - The size of the collection,
   - The date upon which the last file was modified,
   - The date upon which the last file was accessed, and
   - The full path to the location on the filesystem where the collection is stored.
4. Examine the directory structure and look for key terms that provide insights and context into the research area/focus. For example, a directory called "thesis" would indicate PhD-related data.
5. Examine the data and look for applications and research tools. The presence of specific programs and applications can help to indicate the sort of analysis and processing to which the data was subjected, which can provide useful context. For example, a directory called "phaser" that contains .PY files (Python scripts) may indicate that PhaseR (a tool used in genomics) has been used, which suggests a specific research focus. There are many programs and applications called "phaser" however, so such information should be considered within the context of the information gathered through the rest of the collection analysis.
6. Examine the filetypes and look for proprietary file extensions and common filetypes used by specific disciplines. For example, SHP files which are commonly used by tools like ArcGIS or DICOM files which are likely to have been produced by imaging instruments, often in medicine.
7. Use the information acquired to determine the most likely research area and discipline.
8. If possible, use the information acquired to determine the likely value of the collection.
9. Create a MURDA metadata record.
10. Arrange for the collection to be ingested into the MURDA.

Orphaned Research Collection Analysis Workflow Diagram



Ingestion Process

Process

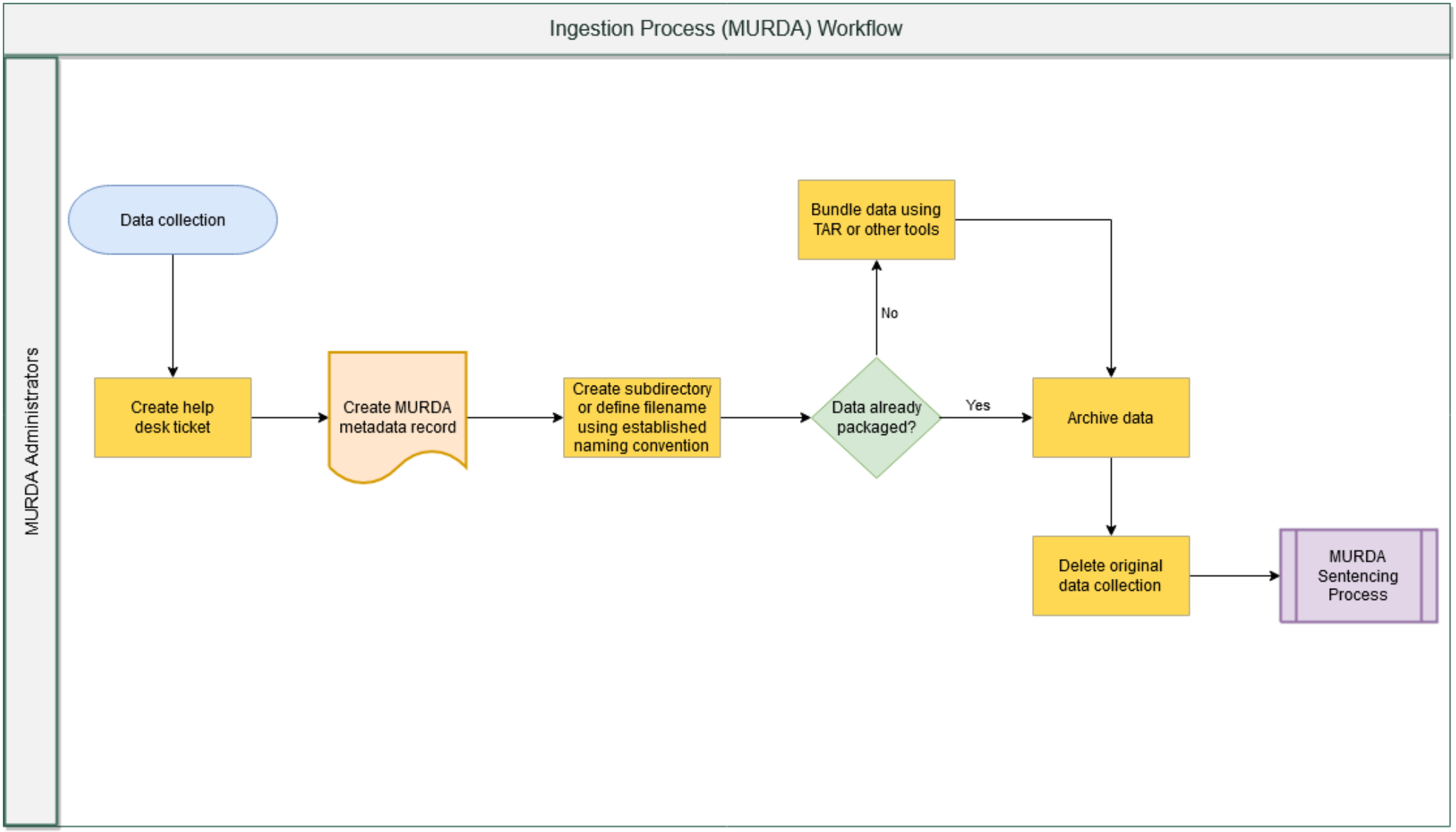
The specific steps for ingesting data into the MURDA are as follows:

1. Create a help desk ticket using the following naming convention:
    a. “RDSM Allocation Decommissioning - <$Collection_name>” (used for RDSM (eResearch) collections), or
    b. “MURDA Archive - <$Collection_name>” (used for non-RDSM collections).
    c. Populate the ticket with the details of the original request (if submitted via email or another external medium) as well as any other relevant information (notes from phone calls and emails, etc).
2. If there is an existing eSolutions Service Desk job that contains most of the information about the archive request, then populate the internal help desk ticket with the basic information. Also ensure that there is a clear reference to the Service Desk job's Request or Incident number in the format #REQ###### or #INC######.
3. Create a child object (MURDA metadata record) under the MURDA RDSM allocation in the Data Dashboard and populate it with the following information:
    a. MURDA-### (Sequentially generated by the system, e.g. MURDA-491).
    b. Collection Title (Should use the MURDA naming convention, <$Service_Name>-<$Directory_or_Collection_Name>-Archive, e.g. “RDSM SmithLab Archive”).
    c. Source (The original location that once housed the collection. Should be the full path detailing the original storage location (if relevant), or the name of the service if the path is not relevant, e.g. “Bridges” or “server1.monash.edu:/fs1/collection”).
    d. Location (The path of the collection within the MURDA allocation. Should use the MURDA naming convention, <$Directory_or_Collection_Name>-<$Data_owner (First_nameLast_name or “Orphaned” if unknown)>-<$Date in ISO date format (YYYYMMDD)>)-<$Help_desk_number>, e.g. “SmithLab-JohnSmith-20181103-FD4903”).
    e. Size (GB) (Produced by the filesystem where appropriate).
    f. Data Owner (If there is no owner, specify "MeRC" or the relevant faculty if known).
    g. Last updated (The date that a file was last changed or “mtime”. This is not the same as the last accessed date (“atime”) which should be considered when sentencing the data).
    h. Retention Period (Years) (Determined by the sentencing process).
    i. Description (Should be as detailed as possible).
4. If the collection has already been bundled up or compressed appropriately by the data owner/equivalent, for example the data is comprised of ZIP, 7Z or TAR.GZ files, etc, then the data can be migrated immediately. It should be noted that there is no clear benefit in compressing the data a second time and, in addition, retaining the original structure and filetype(s) helps to ensure that the data is not modified which is also useful for non-repudiation and digital preservation purposes.
5. If the collection is not bundled up or compressed, the data should be bundled up into one or more TAR files, each of which should be 500GB or smaller. For large collections, a script is used that not only bundles up data into TAR files, but also splits up larger collections into 500GB tarballs and performs an integrity check (checksum).
6. For collections under 500GB in size, a single TAR file following the MURDA naming convention should be created in the appropriate area of the MURDA (e.g. /RDSM_Archives/SmithLab-JohnSmith-20181103-FD4903.tar for RDSM allocations). For collections above 500GB in size, multiple TAR files will be produced therefore a

directory should be created in the appropriate area of the MURDA using the MURDA naming convention (e.g. /RDSM_Archives/SmithLab-JohnSmith-20181103-FD4903/).

7. For Virtual Machine (VM) images such as OVA files and similar data, TAR files are not required as the native format is sufficient.
8. While the TAR process is running (or at any time), document all information concerning the collection in the help desk job. Include useful email correspondence, especially if context around the collection and its owners and contents is scarce.
9. Once data is being ingested, or earlier, contact the Records Management team in order to provide them with the details of the new MURDA collection and also confirm the duration that the data needs to be retained as well as its sentence.
10. If the original collection is stored on a MeRC-administrated system such as RDSM, it can be deleted immediately once ingested because approval should have been provided by the data owner during the MURDA discussion process (or are orphaned collections with no owner). For non-MeRC/RSS services, contact the relevant stakeholders via email or via the eSolutions Service Desk job as appropriate and confirm that the collection has been archived. The relevant stakeholders will then arrange for the original copy to be deleted and will update the external help desk ticket (if any).
11. Update, finalise and close any internal help desk tickets.

Ingestion Workflow Diagram



Sentencing Process

Process

The accuracy of the sentencing process is directly related to the amount and quality of metadata available for a given collection. If the data owner is known, then staff can liaise directly with them in order to determine the appropriate retention period as well as identify any encumbrances or data agreements, etc. If a collection lacks an owner and/or little information is available however, sentencing decisions and retention periods are usually more conservative in order to reduce the level of risk. The specific steps involved in sentencing a collection are:

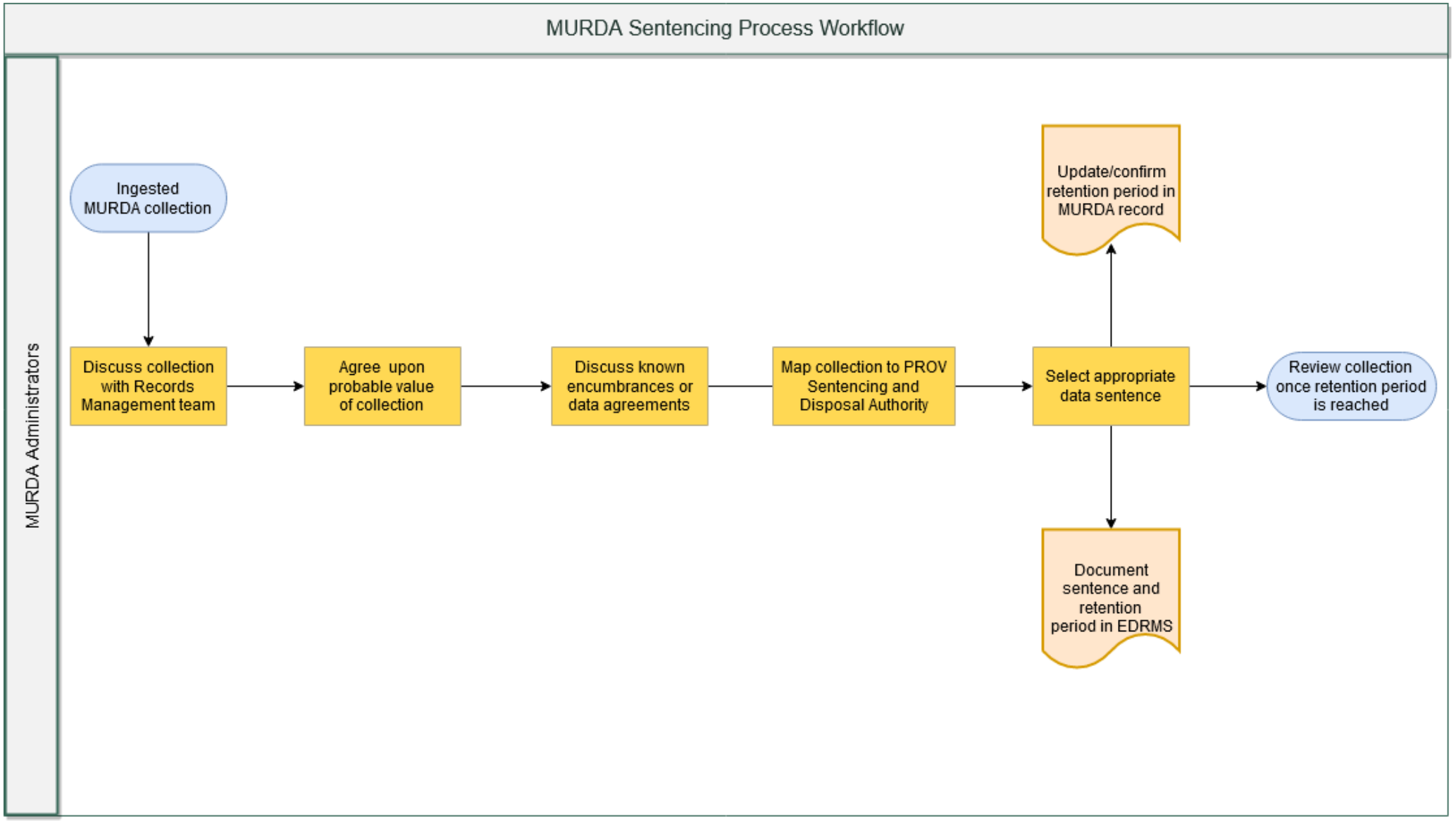
1. Discuss the MURDA collection with the Records Management team.
2. Agree on the probable value of the collection using information provided by the data owner/contact (if known) or through the collection analysis.
3. Discuss any known encumbrances or agreements which may relate to the collection and its retention period/disposal (i.e. deletion). Encumbrances relating to specific fields of research should also be considered where appropriate.
4. Map the collection to a specific Public Record Office Victoria (PROV) Retention and Disposal authority sentence.
5. Consider any relevant encumbrances as well as other factors, and then discuss and agree on the appropriate sentence for the collection.
6. Update the MURDA record (if appropriate) and ensure that the nominated collection's retention period is accurate.
7. Arrange for the MURDA record's retention period and sentence to be added to the EDRMS for review and tracking purposes.
8. Review the collection once the retention period has elapsed.

Public Record Office Victoria - Sentences

MONASH University

PROS 16/07 - 3.3.2	Data and datasets created as part of research activities within the institution, which are of regulatory or community significance.	PERMANENT	**Permanent Archive**
PROS 16/07 - 3.3.3	Data and datasets created from clinical trials as part of research activities within the institution. Excludes data and datasets included in class 3.3.2	TEMPORARY	**Destroy 15 years after completion of research activity**
PROS 16/07 - 3.3.4	Data and datasets created as part of research activities within the institution which involve minors. Excludes data and datasets included in class 3.3.2	TEMPORARY	**Destroy 15 years after child reaches the age of 18**
PROS 16/07 - 3.3.5	Data and datasets created as part of research activities within the institution. Does NOT include data created for specific research activities for which additional regulatory requirements apply, including: clinical trials, gene therapy and research involving children. Excludes data and datasets included in class 3.3.2	TEMPORARY	**Destroy 5 years after completion of research activity**

Sentencing Workflow Diagram



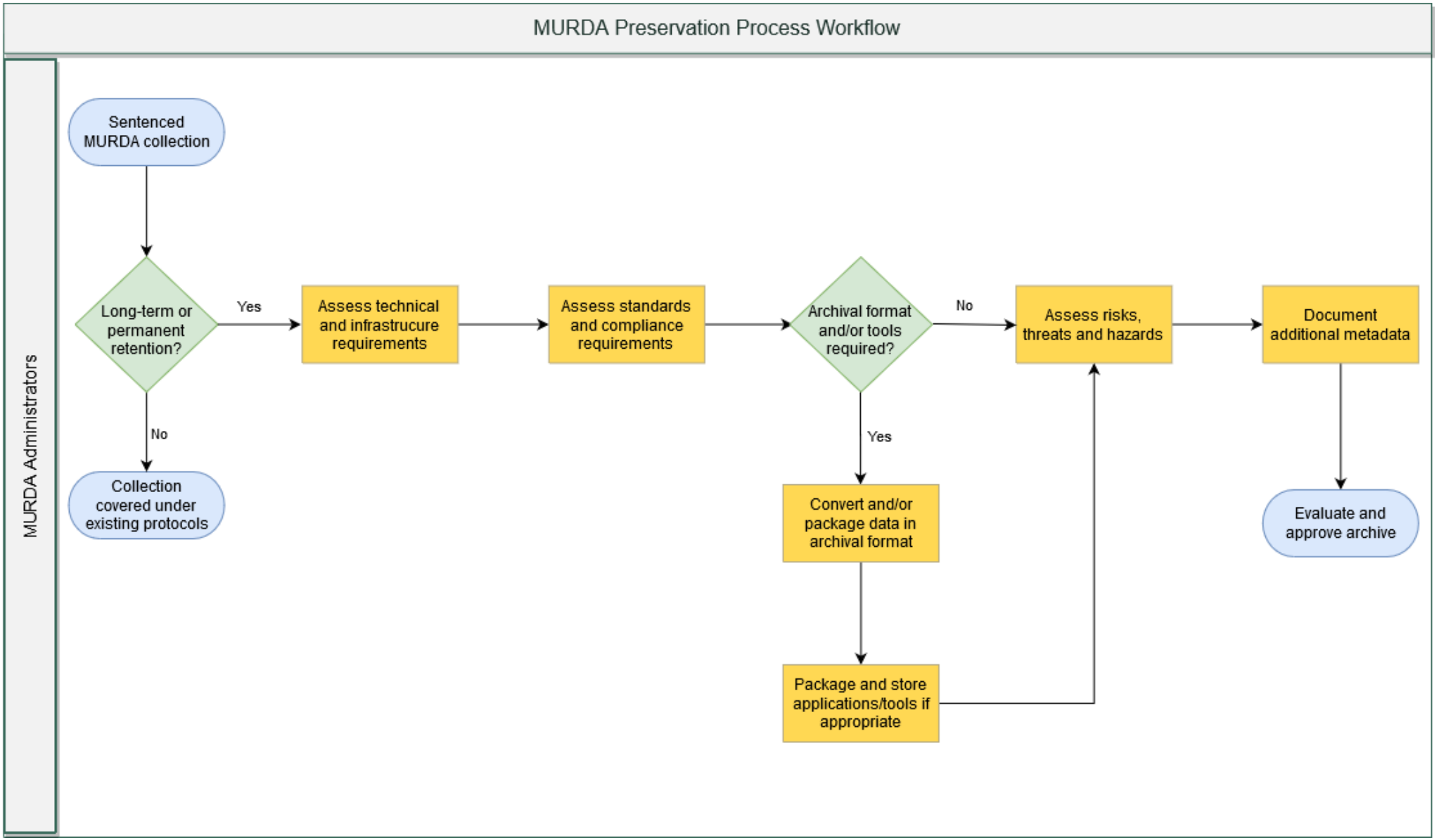
Preservation Process

Process

The preservation process is only applied to MURDA collections that need to be retained for at least ten years, but possible preservation complications are considered for all ingested collections. For collections with long-term retention requirements, the following actions should be taken, ideally in consultation with the data owner (if known):

1. Examine the collection and its metadata record in order to determine the technical requirements. This should include how the data should be stored, whether or not it needs to be packaged with specific software, whether or not it should be stored in its native format or should be converted and re-ingested in the Archive, etc.
2. Assess any relevant compliance requirements (e.g. reporting, auditing and disclosure, etc) and any relevant standards. These may include specific security requirements such as encryption, etc.
3. If any specific tools or data conversion is required, the collection should be extracted from the Archive (if appropriate) and repackaged before being re-ingested into the MURDA.
4. The technical and security risks as well as possible hazards should be considered and the relevant experts should be consulted where appropriate.
5. If any changes were made or any new requirements have been identified, the MURDA and EDRMS records should be updated.
6. If appropriate, the modified archive should then be approved by the relevant stakeholders.

Preservation Workflow Diagram



Access Request Process

Process

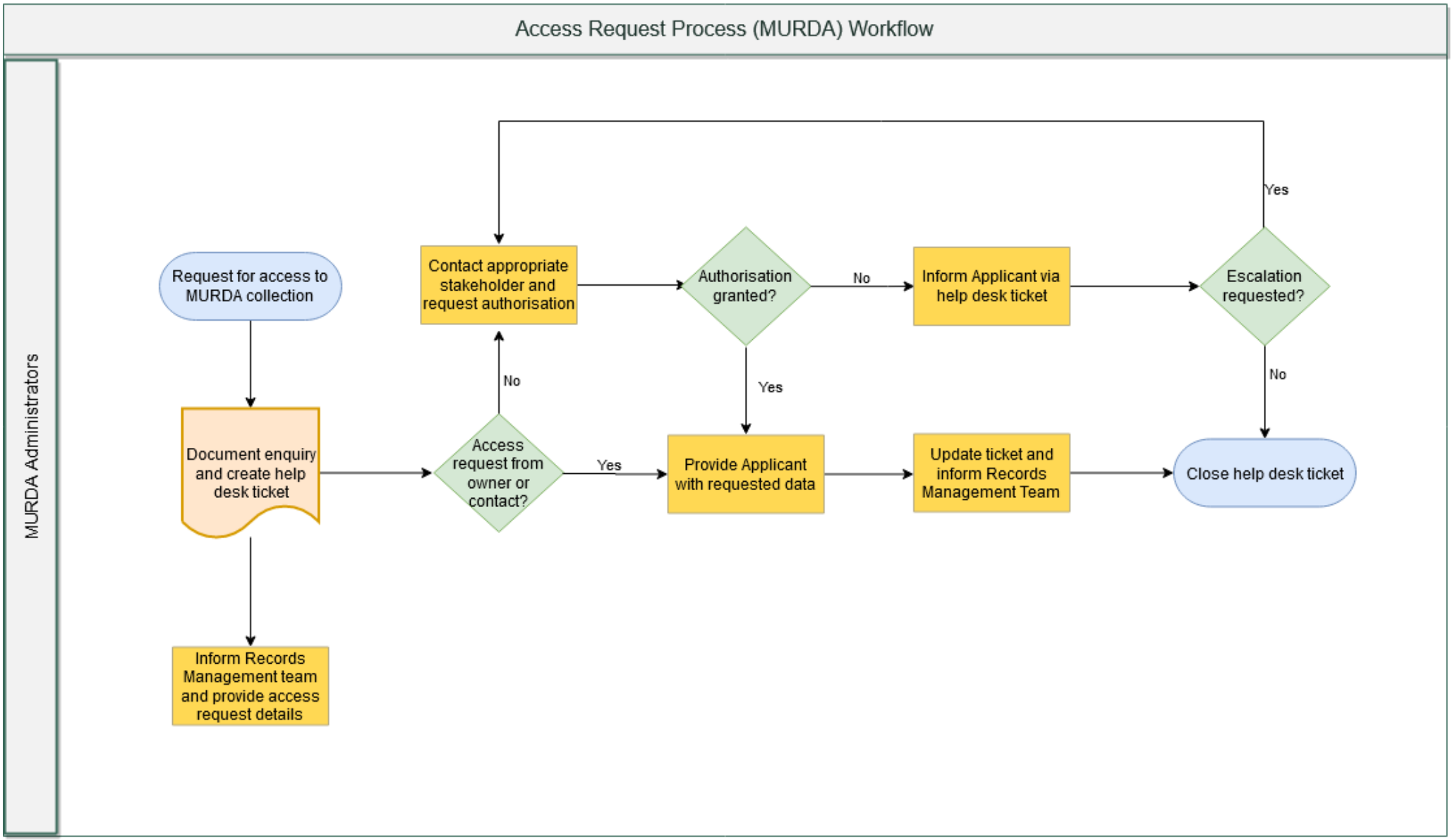
If access to a MURDA collection is requested, a process has been established to ensure that only authorised individuals are provided with access to the data and, in addition, that the enquiry is tracked and recorded. The process is currently being refined to ensure that access requests to individual collections can be more efficiently audited as required:

1. When an Applicant requests access to a specific collection, the request should be recorded in a help desk ticket and its progress should be documented by MURDA administrators. Tickets should be created using the naming convention “MURDA Access Request - <$Name_of_collection>”.
2. The MURDA administrators should contact the Records Management team (eSolutions) and provide them with the Applicant’s details, the details of the access request and the help desk ticket number. The request, as well as its outcome, should be appended to the relevant MURDA collection record in the EDRMS.
3. If the collection’s metadata record lists the Applicant as the data owner or equivalent, then the following actions should be undertaken and documented in the help desk ticket:
    a. The Applicant should be provided with a copy of the whole collection, or a portion thereof, via an appropriate medium depending upon their needs and the technical requirements. Options include but are not limited to CloudStor, RDSM storage, eSolutions storage or Shared Drive (Google) depending upon the sensitivity of the data and other relevant factors.
4. If the collection’s metadata record does not list the Applicant as the data owner or equivalent:
    a. The data owner/equivalent of the relevant collection should be contacted and asked to approve access. If the access request is approved:
        i. The Applicant should be provided with a copy of the whole collection, or a portion thereof, via an appropriate medium depending upon their needs and the technical requirements. Options include but are not limited to CloudStor, RDSM storage, eSolutions storage or Shared Drive (Google) depending upon the sensitivity of the data and other relevant factors.
    b. If the access request is rejected, the Applicant should be informed via the help desk ticket. The Applicant may, at their discretion, choose to challenge the decision and request escalation, which is dependent upon the specific use-case and rationale. If appropriate, the Applicant can liaise directly with the data owner/equivalent and/or the access request can be escalated to a more senior member of the relevant faculty or organisational unit (e.g. an Associate Dean, etc).
5. If there is no known data owner, then the relevant Head of School (or equivalent) is contacted in order to approve the access request. This process should also be followed if the Applicant is no longer a Monash staff member. If access is approved:
    a. The Applicant should be provided with a copy of the whole collection, or a portion thereof, via an appropriate medium depending upon their needs and the technical requirements. Options include but are not limited to CloudStor, RDSM storage, eSolutions storage or Shared Drive (Google) depending upon the sensitivity of the data and other relevant factors.
    b. If the access request is rejected, the Applicant should be informed via the help desk ticket. The Applicant may, at their discretion, choose to challenge

the decision and can request escalation, which is dependent upon the specific use-case and rationale. If appropriate, the access request can be escalated to a more senior member of the relevant faculty or organisational unit (e.g. an Associate Dean, etc).

6. The outcome of the access request should be documented in the help desk ticket and the EDRMS record.
7. Once the access request has been resolved, the help desk ticket should be closed.

Access Request Workflow Diagram



Collection Deletion Process

Process

The deletion process provides a framework that allows MURDA collections to be removed while still retaining a record of the collection. Retaining metadata records of deleted collections and the rationale behind their removal not only ensures that the University is protected if the decision to deletion is challenged, but also allows the Archive to be audited and curated effectively. The process commences when a daily collection report is produced by the EDRMS and is examined by Records Management staff. If the report indicates that one or more collections have passed their retention periods, the following steps are followed:

1. The Records Management team contacts the MURDA administrators and informs them that one or more collections need to be reviewed. A list of collections is provided.
2. A help desk ticket is created for each collection in order to document the details and progress of each review.
3. The nominated collection is reviewed in consultation with the data owner or equivalent in order to determine whether or not the collection is still required.
4. If the data owner/equivalent indicates that the collection should be retained for a longer period:
    a. The rationale behind the extension is documented in the help desk ticket,
    b. MURDA administrators and representatives from the Records Management team liaise with key stakeholders in order to negotiate a new retention period, the length of which is dependent upon the rationale justifying the extension,
    c. All parties agree upon a new retention period which is formally approved by MURDA administrators and representatives from the Records Management,
    d. The MURDA metadata and EDRMS records are updated, and
    e. The help desk ticket is closed.
5. If the data owner/equivalent indicates that the collection is no longer required, then the collection is identified as a candidate for deletion and:
    a. MURDA administrators formally request authorisation to delete the collection via email,
    b. The help desk ticket is updated with the authorisation as well as the rationale behind the removal of the data,
    c. The data is deleted,
    d. The stakeholders are informed,
    e. The MURDA metadata and EDRMS records are updated, and
    f. The help desk ticket is closed.

Collection Deletion Workflow Diagram