



# Digital Asset Register Toolkit



Digital **Preservation** Coalition



# Contents

Introduction	1
Who is this Toolkit For?	1
What is a Digital Asset Register?	1
Why Build a Digital Asset Register?	2
What Makes a Good Digital Asset Register?	4
Step by Step Guide to Building a Digital Asset Register	6
1. Identify Your Motivations for Building a DAR	6
2. Setting the Scope	6
3. Make a Plan for Building Your DAR	7
4. Gathering and Adding Information	8
5. Using Your DAR	9
6. Updating Your DAR	10
7. Reviewing Your DAR	10
Template for Building a Digital Asset Register	12
General Tips for Developing Your DAR Template	12
Recommended Basic Template	13
Choosing a Format and Storage Location for Your DAR	18
More on Information Gathering	19
Existing Data Sources	20
Physical Storage Media	20
Characterization Tools	21
Survey	21
Interviews	22
Useful Resources	24
General Resources	24
Advocacy	24
Characterization Tools	25
Risk Management	25
Metadata	25
Appendices	26
Appendix One - Additional DAR Headings	26
Appendix Two - Example Interview Questions	30

This toolkit was developed by the Digital Preservation Coalition, with support and funding from The National Archives (UK). Much of the guidance included here was gathered from experienced digital preservation practitioners during a series of focus groups held in November 2023. Many thanks to all who participated in the focus groups, and for all of the excellent information and advice provided.

The contents of the toolkit are available for reuse under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International license](#).

© Digital Preservation Coalition, 2024

## Introduction

This Toolkit aims to offer practical guidance on how to build, maintain, and use a Digital Asset Register (DAR). This section introduces what a DAR is, why you may wish to build one, and who this toolkit is aimed at. Subsequent sections will detail what makes a good DAR, a step-by-step guide to developing one, an example template, further details about information gathering, and using and updating your DAR.

### Who is this Toolkit For?

This Toolkit is for anyone who would like to create a DAR to aid their digital preservation activities. It is targeted at practitioners who are working with digital resources and would like a method for gaining better intellectual control of their digital content. The Toolkit is primarily aimed at those developing a DAR for the first time, but it will also be useful for those wishing to review and enhance an existing one.

### What is a Digital Asset Register?

A DAR is a tool for gathering information about digital content to support an understanding of the digital content held by an organization, and to facilitate efficient management of its preservation. It will capture information such as a brief description of digital content, where content is stored, its size, what formats are included, and who holds intellectual property rights. A DAR might take the form of a spreadsheet, a database, or another format that is useful for practitioners.

A DAR provides an overarching view of digital content that helps practitioners to identify gaps, overlap, related risks, and other information useful to making preservation decisions and managing content over time. The information held in a DAR can also be useful when making the case for digital preservation, as it allows the generation of key statistics about the content.

A DAR, as a high-level summary used for managing content, exists at one end of the spectrum of the information you will need to help manage digital content. At the other end of the spectrum, you may look to maintain item level metadata in what can be referred to as “File Manifests”. These capture detailed metadata about individual files that can include the file’s name, location, size, and format.

## Why Build a Digital Asset Register?

Building a DAR can bring many benefits in relation to the management of digital content. By its nature, digital content is ephemeral. There is no easily identifiable “original” in the way there can be for physical, analogue records. Digital files contain binary information on storage media that is easy to replicate and requires software and hardware to be able to interpret it. To help manage this less tangible content, we need new ways to record information about what we have, where it is stored, and what we need to do to preserve it over time.

With these issues in mind, perhaps the primary benefit of a DAR is the ability to gain intellectual control of digital content held by your organization. Carrying out a survey to gather the information for your DAR will allow a unique opportunity to “get to know” your digital content. You will get a better “big picture” understanding of the landscape of your digital content, whilst also gathering useful data on content types, file formats, intellectual property rights, risks, and more.

It also allows you to capture a view of your digital content in a “non-systems centric” way. This means bringing together information by intellectual grouping, e.g. by accession, collection, or series, rather than capturing simple information such as how many files are held in a particular storage area, repository, or digital asset management system.

Building a DAR can also facilitate better management of digital content by:

- Capturing important information in one place that might otherwise be scattered across accession records, paper documentation, catalogues, or not captured at all.
- Acting as a finding aid for content not yet fully processed.
- Providing information needed to plan and prioritize ingest and processing of digital content.
- Facilitating retention decisions and processes.
- Establishing clear responsibilities for the management and preservation of digital content.

Building a DAR also brings a range of potential advocacy benefits, allowing engagement with a wide range of colleagues and providing a reason to discuss the importance of digital content and how it is managed. Due to its lack of physical presence, digital content can sometimes be “invisible” to those not working with them directly. The information and statistics that can be generated from a DAR can help to convey their extent. Ultimately, the resulting DAR can also help frame the digital content as an organizational asset, and the information it contains can be used to make the case for additional resources for digital preservation. Additionally, it provides a concrete reason to request the time and resources required to complete a full survey to assess the extent of your digital content. Without a defined and practical end

product like a DAR, it can be difficult to persuade decision makers of the importance of this process.

Finally, you may wish to build a DAR as it is a recognized good practice activity that is recommended by a number of accreditation and bench-marking processes. In the UK, a DAR is a useful tool to help support Archive Service Accreditation, providing information needed for the application, as well as being a demonstration of robust management practices. The development and use of a DAR is also mentioned in the DPC's Rapid Assessment Model (DPC RAM)<sup>1</sup>, and building your own will aid in leveling-up your digital preservation capabilities. These requirements can also, in turn, feed into advocacy activities, showing the importance of a DAR and how its development and use will reflect well on the organization's digital preservation activities.

---

<sup>1</sup> Digital Preservation Coalition, Rapid Assessment Model -  
<https://www.dpconline.org/digipres/implement-digipres/dpc-ram>

## What Makes a Good Digital Asset Register?

This section highlights some of the key pieces of advice covered later in this toolkit, you might use these to ensure you set off in the right direction for creating your Digital Asset Register (DAR). You may wish to summarize some of these points and other relevant requirements into a set of guiding principles to inform the creation of your DAR. This will allow you to refer back to these throughout the process to ensure you are meeting the aims you set out.

What makes a good DAR?	What makes a bad DAR?
A well-scoped DAR that captures information on a clearly defined set of digital content (e.g. all born-digital content held by the archive).	A poorly scoped DAR that captures patchy information about an indiscriminately selected set of digital content.
Minimum required information is captured about all digital content.	Detailed information is captured on only a few groups of digital content, with many others undocumented in the DAR.
Columns with headings whose purposes are clearly described and documented (this may be in an accompanying document to keep the DAR tidy).	Columns with vague headings that are not described and documented, or where it has been done poorly.
Information that is captured in well-structured formats and is processable where possible (e.g. numerical data or yes/no).	Long blocks of textual information of varying quality that is hard to interpret or process.
Information captured directly supports the DAR's proposed uses.	A large number of columns, many of which contain extraneous information that is never used.
Information captured does not significantly duplicate the information held in other systems.	There is significant duplication of information held in other systems.
There is a balance of effort between creating a DAR that is useful and the time required to complete entries.	New entries are excessively time consuming to complete.
Use of set lists of terms and/or drop-down lists for data entry where possible, to facilitate easy comparison of records.	No guidance or structure for data entry, leading to variable data quality.
Information is captured at a consistent level across the DAR (e.g. collection/fonds, series, or accession)	Information is captured at different levels across the DAR (e.g. at collection-level for some content, and at item-level for others).

<b>What makes a good DAR?</b>	<b>What makes a bad DAR?</b>
Use of a consistent method for capturing information about potentially measurable elements such as risk (e.g. a standard scoring metric for assessment)	Use of a range of different methods for capturing potentially measurable information, making it impossible to compare different records.
Clearly marking if entering information in each column is considered to be required or optional, to help ensure the most important columns are populated.	Not identifying if information is required or optional, thereby not making it clear which are the most important columns to complete.
A clearly established process for when the DAR will be updated (e.g. when new accessions are received or as part of a monthly data audit).	No standard update process, leading to a DAR that is rarely and/or inconsistently updated (e.g. new content are often not added or data entered is poor).
Clear roles and responsibilities for managing, updating, and reviewing the DAR.	No assigned responsibilities for managing, updating, and reviewing the DAR.
A plan is made and executed for the DAR to be reviewed at regular intervals, to ensure it remains useful and accurate.	No plan to update and review the DAR, leading to its continued use when it is out of date and not fit for purpose.



# Step by Step Guide to Building a Digital Asset Register

This section provides an overview of the recommended steps for designing, building, and maintaining a DAR. Some of the elements covered are examined in more detail in the “More on Information Gathering”, “Template for Building a Digital Asset Register”, and “Hints and Tips” Sections. Use this section to help plan how you will build your DAR.

## 1. Identify Your Motivations for Building a DAR

Having a clear understanding of why you are building a DAR, what it will be used for, and who will use it, helps to ensure you produce a register that is effective, functional, and straightforward to maintain. Take time to consider the following questions, and to document the answers, at the beginning of the process:

- What issues do you hope to address with your DAR?
- What opportunities do you hope to create?
- How will the DAR be used?
- Who will use the DAR?
- Are there any other interested parties to consider? (e.g. content creators or holders, managers or executives, IT colleagues)

You may wish to answer these questions on your own, or to invite other interested parties to contribute. After answering these questions, you may wish to use the information to develop a set of guiding principles to guide the development of your DAR.

## 2. Setting the Scope

Being clear about the scope of the content to be documented in the DAR is essential. Without clear scoping, critical content may not be included, or the DAR may be populated with lots of extraneous information. When setting the scope, it is important to consider what content will be included, as well as what information should be collected about that content. When first building your DAR, you may wish to err on the side of a broad but shallow register to ensure nothing important is missed. The scope can then be tightened, and more detailed information captured as the DAR is updated.

Answering the following questions, and documenting the answers, can help with setting the scope:

- What content should be included within the DAR? For example:
  - Born-digital content and/or digitized content?
  - Only digital content currently managed by the archive/library/digital preservation team, all digital content produced by the organization that will

- require preservation. (e.g. records in current use, web content, social media data, etc.), or some other defined grouping?
- Digital content generated by the organization and/or digital content received from external depositors?
- How will organizational culture affect your ability to capture information for the DAR?
  - Is there a clear mandate for preservation and/or is it a priority for the organization? If yes, it should be straightforward to engage colleagues in contributing to the gathering of information. No mandate and/or executive and managerial support can be a barrier to information gathering.
- What resources (time, tools, staffing, and skills) do you have available for undertaking this work? If they are limited, you may need to aim for a tightly scoped DAR.
- Are there types of content that need to be prioritized as they have a high-risk profile? For example, if they are critical organizationally or they are stored on unreliable/unstable storage media.
- Are there any “low-hanging fruit” that would be easy to capture information on first? For example, documenting all digitized material because it is relatively homogenous, or all born-digital material as it is stored in a single location.

As with identifying motivations (above), while you may be able to complete a scoping exercise on your own, it may be useful to include interested parties in the process.

### 3. Make a Plan for Building Your DAR

Once you have established the motivations and scope for your DAR, you should then develop a plan for how to achieve these aims in practice. At a minimum, your plan should cover the following:

- Clearly identified roles and responsibilities for designing the DAR, gathering information, ongoing management, and review.
  - It is important to carefully consider who should be involved and when. You should look to balance the advocacy benefits of getting other colleagues involved and the potential complications this may add to the process.
- The format that will be used for the DAR, and a template design. Ideally, this will also include documenting the various column headings/data elements and defining the information to be captured in each. More detailed guidance on this process is included in the section “Template for Building a Digital Asset Register” later in the toolkit.
  - Make sure to consider potential duplication of information that is held in other systems, and whether or not the inclusion of this information is necessary to facilitate the intended uses of the DAR. For example, some organizations find it useful to gather a small amount of provenance information within their DAR, to ensure it is easy to access and is recorded consistently across digital content. If,

however, you already have robust accessioning processes in place that record the information in an accessible register or system, then there is little point in duplicating this in the DAR.

- The Who, What, When, Where, and How of the information gathering phase:
  - Who will carry out the information gathering and who will they need to talk to/ask for information?
  - What information needs to be gathered?
  - When will the information gathering take place?
  - Where will the information be gathered? Will meeting space be required for interviews? Do you need access to survey and/or characterization tools? Will you gather information in a separate document before processing and entering it into the DAR?
  - How will you gather the information? Will you interview colleagues? Will you circulate a survey? Will you use characterization tools to gather information from various storage areas? Will you need to analyze data from existing systems, e.g. accession records or catalogue records?

It is also important to consider if you will need to gain approval for your plan. Having the plan approved by your line or program manager will provide a mandate for carrying-out the work, which may help to prioritize it. Having the plan approved by a senior manager may also help smooth the information gathering process by providing a directive that will motivate colleagues to participate.

## 4. Gathering and Adding Information

Your plan developed in the previous step should clearly describe your approach to gathering information for your DAR. This will likely be the most time-consuming and intensive step of building it. The following section of this toolkit will examine possible information gathering techniques in more detail, but they might include the following:

- Extracting information from existing resources, such as deposit agreements, accession records, catalogues, digital repository systems or other documentation. These may be held digitally, but in some cases, you may need to consult paper records.
- Interviews with colleagues to gather information on digital content they produce and/or manage.
- Circulating a survey to capture information on what digital content is produced and/or managed across the organization.
- Using characterization tools to survey digital content currently held on archive drives, shared drives, or other storage areas.
- Surveying physical media that has been included with deposited content but has not yet been transferred to the organization's storage.

You may wish to gather information in an intermediary document before processing it for inclusion in the digital asset register. This can be particularly helpful if your DAR is highly structured and utilizes techniques such as agreed lists of terms to be used.

When adding information, try to be consistent in the minimum level of detail included and also adhere to any guidance you have developed. When adding information to your DAR for the first time, you might find yourself updating and refining the initial guidance based on the lessons learned during the information gathering and the quality and completeness of the data that was collected. Continuous improvement of resources is always to be encouraged but do make sure to update the relevant documentation or guidance.

## 5. Using Your DAR

Once you have added all of the information to your DAR, you will be able to use it as a core resource to support your digital preservation activities. In particular, there are two key types of activities DARs are used to help with. These are:

- For planning and collections management. This can include:
  - Documenting where content is stored.
  - Helping to locate content when you need to provide access. This is generally for content prior to its ingest into a repository system. This process may also include checking if there are any reasons to deny or limit access, due to sensitive data or restrictions relating to intellectual property rights.
  - Planning for storage needs by tracking content growth over time.
  - Documenting risks to digital content, including identifying common risks which you should take action to address.
  - Planning and managing the ingest of content into a digital preservation repository. In particular, a DAR can be used for prioritizing content for ingest based on criteria such as risk profile and format types.
  - Managing retention and disposal.
- For advocacy and reporting purposes. This can include:
  - Producing figures and charts, for example showing progress over time (e.g. content growth, amount of content processed, or changes in the types of content being collected). This information can be useful for tasks such as annual reporting.
  - Making the extent of digital content evident and also aiding with transparency in relation to the amount of digital preservation work being undertaken.

Summary visualizations can be particularly effective in highlighting digital content that might otherwise be “invisible” to colleagues.

- Communicating with colleagues, in particular facilitating discussions with IT colleagues as you will be able to be clearer and more specific about requirements for storage, hardware, software, and other technical infrastructure.
- Supporting requests for prioritization of digital preservation activities and making the case for additional resources.

The information captured within your DAR may also be useful for activities such as applying for accreditation and measuring capacity and capabilities with tools such as maturity models. Finally, your DAR can be a great tool to help new members of staff joining your team become familiar with your organization’s digital content and the types of digital content they will be working with.

You may find this list of potential uses helpful when considering your motivations for creating a DAR and setting its scope.

## 6. Updating Your DAR

It is important that your DAR is kept up-to-date and does not lie dormant. Making a clear plan for updating it on a regular basis will ensure it remains accurate and relevant. How you approach updates will depend on a number of factors including your organization’s type and culture, when and how new digital content is created or received, and how much time and resources you have available to put towards the activity. There is no one correct way to approach updating your DAR, but you may consider a method similar to one of the following:

- Adding information at the point of acquiring and/or accessioning new digital content or deposit.
- At a regular time, each week or month, e.g. setting aside a regular afternoon slot for updating the information.
- As part of a regular content audit process, e.g. a quarterly survey of physical media.
- At regular points in the year if your organization follows a particular cycle of activity, e.g. at the end of each term in a university context or the end of an exhibition or tour if you work in the arts.

## 7. Reviewing Your DAR

It is important to review your DAR from time to time to ensure that it is still fit for purpose and that information is being added in a consistent and suitable manner. If multiple people are updating it, you may wish to identify a person or small group of people to have editorial

control of the register to oversee the quality of information captured, with checks carried out on a regular schedule.

Less frequent, but more in-depth reviews are also recommended to revisit the purposes of the DAR, its scope, and structure. These reviews may be triggered by significant changes in digital preservation practice or linked to a policy review or new planning exercise. For these reviews, you may wish to work through a set of similar steps as described in this toolkit for first building a DAR.

Finally, if you reach a point where all of your digital content has been ingested into a repository system, you may find that maintaining a DAR has become redundant. It is likely that all of the information that would be entered into and retrieved from the DAR is now managed by the repository or held in other systems such as those for accessioning and cataloguing. If this is achieved, you can make the decision to cease active use of the DAR. You should also consider at this point whether it would be helpful to retain the DAR as an institutional record and a back-up of information held elsewhere or, particularly if the DAR contains sensitive information, if it should be disposed of.

# Template for Building a Digital Asset Register

This section provides guidance on how to design the template for your DAR. This includes general tips on things to think about, recommendations for key headings to incorporate, and some suggestions about particular formats and tools you can use to practically implement your plan.

## General Tips for Developing Your DAR Template

The following are some general items to consider when you are developing the template for your DAR. Following these tips will help you to create a DAR that is easier to use and update.

- Use agreed lists of terms to be used and/or set formats for individual data elements where possible. This will help to ensure the capture of good quality data, as well as making it easier to manipulate. This can be implemented by having clear guidance on the terms to use under particular headings, or by using functionality like drop-down lists for data entry.
- Consider carefully who will be updating the DAR as well as who needs to use it. Some staff members may be more technically proficient than others and this may guide the format you choose for your DAR.
- You may also wish to add a layer of security to help control who can make edits to the DAR. This might be implemented using password protection for editing the document, or by placing it in storage where you can control access permissions.
- Remember to find a balance between trying to be comprehensive and the time taken to complete entries. Consider starting with a simpler set of essential headings and growing these over time.
  - Also consider if some columns might be marked as “required” or “optional”.
- Reduce duplication of effort by only including information that is captured in other systems (e.g. accession or catalogue records) where there is good reason. For example, the information in other systems is not readily accessible or is not consistently captured for digital content.
- Depending on how you manage digital content, you may wish to use separate tabs, sections or DARs for different types of content. For example, separating born-digital content and digitized content into different areas.
- Consider adding a dashboard that automatically generates key statistics about the digital content, and graphs and charts that provide visual representations of the stats. This may include total counts of number of files and the size of content, as well as breakdowns or these figures using factors such as year received or current processing status. This information is then readily available for advocacy purposes.

- Finally, if you have a digital preservation repository system in place, consider aligning the headings of your DAR with reports that can be produced by the repository to help streamline the information.

## Recommended Basic Template

This subsection provides an overview of the headings (or data elements) that we recommend you include within your DAR. These recommendations were derived from an analysis of 14 real-world DARs shared by practitioners. The most commonly used and fundamental headings are included here. Additional possible headings you may consider adding are included in Appendix One of this toolkit. Some of the headings below may duplicate information held in other systems. A note has been added to the description field for headings you may choose to omit to reduce duplication.

For each heading below, we have included suggestions as to how it might be formatted, as well as a description of the information you might record under each. If alternative names for the heading were used in any of the analyzed DARs, they have also been included. You may find one of these alternatives better matches the terminology used at your organization.

Accompanying this toolkit there is an Excel spreadsheet DAR template that includes each of these headings, a ready to use statistics dashboard, and guidance for completing and updating the template. This is free for you to download, edit, and use to help create your own DAR.



Heading Name	Alternative Name(s)	Suggested Format	Description
DAR ID	Accession Number, Reference Number, Reference Code, Catalogue Reference	Alphanumeric or numeric in a set format	A unique identifier given to the DAR entry. This may reuse an ID from another system, such as an accession number. It is important to use a consistent system of assigning IDs throughout the DAR.
Content Name	Digital Asset Name, Title, Name of Series, Name of Collection	Text	The name of the collection/series/grouping of content described in this DAR entry. The name entered here should be consistent with how the content is described in other systems and documentation.
Transfer Year	Accession Year	Numeric	The year the content was received by the archive/library/repository.
Accession Type	Deposit Type	Text using a set list of terms	The type of accession or transfer, e.g. gift, loan, purchase, internal records, transfer, etc. The list used should reflect the types of deposits your organization typically receives. This heading may be removed if you feel this information is sufficiently documented in another system or register.
Source	Depositor, Donor, Creator	Text	Name of the person, department, or organization that donated or transferred the content. This heading may be removed if you feel this information is sufficiently documented in another system or register.

Heading Name	Alternative Name(s)	Suggested Format	Description
Description		Text	A short text description of the content, which may include the subject(s) of the content, why it was created, what it was used for, etc. This heading may be removed if you feel this information is sufficiently documented in another system or register.
Data Owner	Responsible Staff Member	Text	Name of the member of staff (or the role name) who has primary responsibility for the management of this content. This information is particularly useful if the content has not yet been transferred to the archive/library.
Number of Files		Numeric	A total count of the number of files included in the grouping/collection of content being described.
Content Types		Text using a set list of terms	A summary list of the types of content included, e.g. text documents, images, databases, spreadsheets, CAD, etc.
Size		Numeric using a set unit	The total size of the grouping/collection of content. It is important to use a consistent data storage unit for all entries to allow calculations using the data, e.g. MB, GB, TB etc. Select the unit most useful to your context. For example, if you generally have smaller amounts of digital content, you may choose MB or GB.
Storage Location		Text using a set format	The location, normally as a file path, where the main copy of the content is stored.

Heading Name	Alternative Name(s)	Suggested Format	Description
Status		Text using a set list of terms	A description of the point the content has reached in your processing or preservation workflow, e.g. accessioned, quarantined, pre-ingest, ready for ingest, ingested, etc.
Sensitive Data	Data Protection	Text	Details of any sensitive, proprietary, private, or personal data within the content.
Rights	Intellectual Property Rights	Text	Notes on who owns IPR and other rights to the materials, including if there is any 3rd party material included, and any licenses that have been granted.
Risks		Text	Details of the key risks faced by the content, e.g. obsolete formats, limited metadata, unstable media. This may be a summary of a risk analysis exercise or information on results from using a risk analysis tool such as <a href="#">The National Archives (UK)'s DiAGRAM</a> .
Retention Period		Numeric, Text, or Date and using an agreed unit, set list of terms, or set format, respectively	Information on the period that the content should be retained and preserved for. This might be a number of years, until a specific date, or a named period such as "indefinitely".
Access Restrictions		Text	A description of any access restrictions in relation to the content, e.g. embargo periods or permissions needed for access.

Heading Name	Alternative Name(s)	Suggested Format	Description
Notes	Additional Information	Text	Any additional information about the content that has not been covered under other headings. If you find yourself frequently adding similar information here, this can indicate the need for a new heading to be added to the DAR.
Date Added		Date using a set format	The date the entry was added to the DAR.
Entry Created By		Text	Name of the person who added the entry to the DAR.
Date Updated		Date using a set format	The date the entry was last updated.
Updated By		Text	Name of the person who last updated the entry.

## Choosing a Format and Storage Location for Your DAR

In addition to selecting the headings and data elements for your DAR, you must also decide the format in which to maintain the information and where the DAR will be stored. When selecting the right format and location you will need to consider:

- What software options you have available and what functionality they offer that will help with building, managing, and using your DAR. For example:
  - Do you want to control data entry on certain fields by implementing pick lists?
  - Do you want to save your data in a flat table structure or develop a structure with more complex data relationships?
  - Does it need to have inbuilt security functionality to stop unauthorized edits?
- Who will be adding information to and/or using the DAR? What software do they have access to and feel comfortable using?
- Where is the DAR going to be stored? Is it possible to manage access permissions based on its location?
  - Your DAR will contain a lot of important data about your organization's digital content, and it is important to keep this secure. Make sure to consider security when selecting both the creation tool and storage location.

Excel spreadsheets are the most commonly used format for the creation of DARs as the software is relatively ubiquitous, it can support controlled data entry and validation, and you can also add data visualizations such as summary tables, graphs, and charts. These can even be added as a separate statistics dashboard. Free alternatives to Excel include Google Sheets and LibreOffice Sheets.

Spreadsheets are not, however, the only format options for your DAR. If you want to have a more complex data structure, as mentioned above, you may wish to select a database format for building your DAR. This will likely require a higher level of technical skill but does offer the advantage of developing a user interface for data entry and editing. Do keep in mind though, that these formats can be more difficult to maintain over time due to their complexity. Other, simpler, options that have been used by practitioners include a table in a Word or other word processing document, the Lists functionality in Teams, or using pages in Sharepoint.

Once you have selected the format and storage location for your DAR, take time to ensure it has been set up correctly and that functionality has been well tested before starting to input information in earnest.

## More on Information Gathering

It is important to carefully plan out how you will gather information for your DAR before beginning the process. This will help you to ensure you have considered all of the key factors:

- The scope of the content you are hoping to gather information on.
- The time, resources, and tools you have available to help with the information gathering, including the people who will be undertaking the task and the skills they have.
- The sources (people, documentation, and systems) you will gather information from.

It is essential that you are realistic about the amount of information you will be able to gather and that you ensure you have the focus time needed to dedicate to this task. Without a clear and realistic plan your DAR may become a half-finished resource that is never fit for purpose. Do remember that capturing something simple is better than nothing and you should not get caught up in worrying about it being perfect from the outset. Figure out the essential information fields and populate those first. These can be built out over time with more data elements as you have time and/or they are needed.

You may also consider a federated approach to capturing information and updating the DAR. In this approach you might be responsible for setting up the structure of it, and managing and monitoring its use, but responsibilities for information updates sit with data owners or creators. This approach can be helpful with advocacy and engagement or if you are capturing information for a large organization and/or with a wide scope. It does, however, also mean that you are beholden to the data owners, their willingness to complete the DAR, and the quality of data they produce. Before using such an approach, consider what level of buy-in there is for the process, whether colleagues can be relied upon to keep it up to date. Including DAR updates in process documentation, putting out regular reminders, and carrying out checks can all help with ensuring the necessary data is added. Some practitioners have also found environmental sustainability (not keeping unnecessary data) to be a good motivator to encourage participation. You may also need to consider what the consequences might be if colleagues do not keep the data up to date, in particular, the impact it might have on collections management capabilities.

We will now look in a little more detail at some of the key data sources and methods for capturing data for your DAR.

## Existing Data Sources

No matter the scope of your DAR, you will almost certainly look to gather information from the range of existing systems and documentation maintained by your organization. Data sources may include:

- Accession/acquisition records and paperwork
- Documentation provided by depositors
- Records in cataloguing systems
- Reports from Digital Asset Management or Digital Preservation Repository systems

Before beginning information gathering it can be useful to create a list of these relevant systems and documentation maintained by your organization, the types of information they may contain, and the format of that information. This can then be used to plan how to prioritize gathering information and what you will look for where. For example:

- Accession records and acquisition paperwork may be the best place to find information on ownership and intellectual property rights;
- Descriptions of the digital content can potentially be harvested from cataloguing systems; and,
- A repository report might include details of the number of files included and their total size.

## Physical Storage Media

Depending on the processing status of the content held by your organization, you may need to carry-out a survey of unprocessed external/legacy storage media. The extent of this survey will depend on the time and resources you have available. Some issues to consider are:

- Is there a list, or can a list be generated, of all of the accessions/digital content containing such storage media?
- Where is the media stored? Has it all been collected together and is easily accessible, or is it stored amongst physical content, and therefore might be difficult to locate?
- Is there any documentation that lists what is on the media?
- What types of media are there? Will you require technical help, new hardware, or to consider outsourcing data access and copying?

You will also need to consider carefully if you will simply be viewing the content to gather information, or do you have the time and resources to undertake copying of the content onto more secure storage ready for processing? This may significantly increase the time required

for the task, but overall, it will be a more efficient use of resources, be less risky than multiple uses of the media, and mean the content will likely be on more stable and secure storage.

Ideally, surveying and processing digital content held on physical media would be a priority due to the high level of risk associated with these storage formats. But in reality, you may find that carrying out such a survey is not feasible as part of your initial information gathering, due to time available, the complexity of the media, and how it is stored. If this is the case, it is recommended that you consider addressing physical storage media as a possible next step or follow-on project, to ensure your DAR is comprehensive and that you have mitigated the high level of risk to content stored on this type of media.

## Characterization Tools

Characterization is the process of identification and description of what a file is and its defining technical characteristics, such as file format, size, and the software used to create it. There are a number of characterization tools that can be used to generate this information for a group of files held in one or more folders. These are, therefore, particularly useful for capturing summary information such as “Number of Files” and “Total Size” for your DAR, without having to manually count and calculate everything.

A characterization tool can generate this information for small amounts of digital content in seconds, although longer processing times might be required for larger amounts of digital content. Indeed, some tools may struggle with particularly large-scale digital content. In this case, you will need a computer with sufficient processing power and/or you may need to analyze the content in sections.

You can find more information on the range of characterization tools that are available via [the COPTR tools registry](#) and some of the commonly used tools are included in the [Useful Resources section](#) below. Some tools are accessed via a user interface like the software and apps you will be familiar with using, while others can only be run using the Command Line interface. The free training course [Novice to Know-How](#) also includes more detailed information on understanding digital files and characterization, as well as providing demos of a tool called [DROID](#). It is possible to run DROID from a user interface as well as from the Command Line. [The DROID tool demos](#) are also included in the Digital Preservation Handbook.

## Survey

If your DAR's scope is relatively wide, for example if you are aiming to include details of digital content at your organization still in active or semi-active use, you may find it helpful to capture information for your DAR by circulating a survey. The benefits of using a survey are



that you can gather information for a large group of people quickly, you may capture details of content that you were unaware of, and, if the survey is well-designed, the data captured can be in easily processable formats that might translate directly to DAR fields. The downsides of using surveys are that you must rely on the goodwill of others to ensure the right people complete the survey, and the data may vary in quality.

If you do plan to use a staff survey to gather information, you should consider the following:

- Does your organization subscribe and provide access to a specialist survey tool, such as Qualtrics, SurveyMonkey, or TypeForm? These can provide useful functionality for creating, structuring, and analyzing the data from your survey. If these are not available, you may consider using a free service like Google Forms.
- Make sure to carefully consider the questions you add to your survey. Aim for clear, jargon-free questions and an emphasis on easily processable data where possible. For example, use answers with check boxes, pick lists, or data validation to ensure data is entered in a consistent way.
- Limit the number of questions that allow text-based answers. The information gathered is more likely to be of variable quality and will take significantly longer to process.
- Consider how you will communicate about the survey. Will you distribute it via email, an intranet site, and/or a newsletter? Are there particular members of staff you should send targeted messages to? How many reminders should you send? Would offering accompanying information sessions help to increase engagement? Do you need to work with managers to have them encourage staff to complete the survey?

Finally, it is important not to underestimate the amount of time you may need to spend analyzing the data captured from a survey before it can be included in your DAR. As mentioned above, text-based answers can take a significant amount of time to process, and even more straight-forward quantitative data will likely require some quality checking to make sure it is fit for purpose.

## Interviews

Interviews are another potential information gathering approach that can be helpful for building your DAR. They can be particularly useful when information gathering is being carried out by an individual or small team, but the content is managed and/or created by a larger number of people. Interviews will provide an opportunity for an in-depth dive into the knowledge an individual may have amassed, but perhaps not documented, about content they manage or have created. For example, they are very good at unearthing details about content hidden away in a forgotten corner of the staff shared storage, that might otherwise

have been missed. Interviews are, however, very labor intensive to plan, stage, and analyze, so it is important to carefully consider if you have the time and resources available to include them as part of your information gathering process.

If you do plan to include interviews, then it is useful to consider the following:

- What questions do you need answered? You will find that you are able to cover fewer questions than you might think in a session.
- Are your questions clear and jargon-free? Is there any specialist terminology that you need to include and will have to explain?
- What type of interview format will work better? Structured, semi-structured, or unstructured?
- What information do you need to provide to the interviewee(s) ahead of time? A participant information sheet with a description of what will happen during and after the interview is useful. This can include venue information, if you wish to record the interview, and the list of questions.
- Is there a suitable venue available for the interview(s)? This can be in-person or virtual. It should be somewhere that both you and the interviewee feel comfortable.
- Make sure you are well prepared for the interview; with all of the documents you require and any note-taking tools or recording equipment ready.

An example set of interview questions to help with building a DAR are included in this toolkit as [Appendix Two](#). You can use this list as a starting point for developing a question set that is suitable for your own needs, matching the DAR template you have designed. Also, as with surveys, make sure you have adequate time available to analyze the outcomes of the interviews. Extracting the information you need and formatting for inclusion in your DAR might take a significant amount of time.

## Useful Resources

### General Resources

**Novice to Know-How: Digital Preservation for Beginners training** – A training course offering a practical introduction to digital preservation.

<https://www.dpconline.org/digipres/prof-development/n2kh-online-training>

**The Digital Preservation Handbook** – An authoritative and practical guide to digital preservation.

<https://www.dpconline.org/handbook>

**Awesome Digital Preservation** – A community-owned list of resources for digital preservation.

<https://github.com/digipres/awesome-digital-preservation#get-started>

**DPC RAM** – A broadly-scoped maturity model for digital preservation that supports benchmarking of current capabilities and future planning.

<https://www.dpconline.org/digipres/implement-digipres/dpc-ram>

**NDSA Levels of Digital Preservation** – A maturity model focusing on the more process-orientated elements of digital preservation.

<https://ndsa.org/publications/levels-of-digital-preservation/>

**DPC Topical Notes Series** – A set of 2-page introductions to key digital preservation issues written in plain English for use in advocacy and when talking to non-specialists.

<https://www.dpconline.org/digipres/discover-good-practice/tech-watch-reports#topical-notes-series>

**The National Archives (UK) Digital Preservation Workflows** – Guidance on how to establish digital preservation processes at your organization.

<https://www.nationalarchives.gov.uk/archives-sector/advice-and-guidance/managing-your-collection/preserving-digital-collections/digital-preservation-workflows/>

### Advocacy

**Executive Guide on Digital Preservation** – A guide containing messages and motivators that can be used to communicate with senior executives, legislators, and budget holders, as well as decision and policy makers, about digital preservation.

<https://www.dpconline.org/digipres/implement-digipres/dpeg-home>

**Business Case Toolkit** – A toolkit to help with preparing a business case for securing resources.

<https://www.dpconline.org/digipres/implement-digipres/business-case-toolkit>

**The Bit List** – A list of “digitally endangered species” that can aid with advocacy.

<https://www.dpconline.org/digipres/champion-digital-preservation/bit-list>

## Characterization Tools

**DROID** - <https://www.nationalarchives.gov.uk/information-management/manage-information/preserving-digital-records/droid/>

**FITS** - <https://projects.iq.harvard.edu/fits>

**Siegfried** - <https://www.itforarchivists.com/siegfried/>

**JHOVE** - <https://jhove.openpreservation.org/>

**Data Accessioner** - <https://github.com/digitalpowrr/DataAccessioner>

## Risk Management

**DiAGRAM** – A tool to help with managing risks to your digital content.

<https://diagram.nationalarchives.gov.uk/>

**SPOT** – A model for understanding technological risks.

<https://www.dlib.org/dlib/september12/vermaaten/09vermaaten.html>

**DRAMBORA** – A model and tool to allow practitioners to audit digital preservation risks.

<https://www.dcc.ac.uk/tools/drambora>

## Metadata

**Dublin Core** - <https://www.dublincore.org/>

**PREMIS** - <https://www.loc.gov/standards/premis/>

## Appendices

### Appendix One - Additional DAR Headings

This appendix includes a list of additional headings you may consider adding to your DAR Template, depending on your needs and digital preservation working practices. As with the core recommended headings, these have been drawn from a collection of in-use DARs.

Heading Name	Alternative Name(s)	Suggested Format	Description
Importance		Text using a set list of terms	A rating of the importance of the content, e.g. vital, high usage, key asset, etc.
Related Collection(s)	Associated Series/Collections, Parent Collection	Text	The name of related content groupings, whether digital or analog. May include a link to catalogue records or accession numbers.
Asset Type		Text using a set list of terms	The type of digital content, e.g. born-digital, digitized, mixed, etc.
Accruing		Text	A brief description of additional content expected to be received in the future.
File Formats		Text using a set format	A list of the file formats included within a group of digital content, e.g. JPG, DOCX, DXF, XLSX.
Encryption		Text	A summary of any issues relating to encryption. Details for individual files can be stored in file manifests.

Heading Name	Alternative Name(s)	Suggested Format	Description
Storage Type	On Prem/Cloud	Text using a set list of terms	The type of storage used for the main copy of the content. This might include options such as: Archive Drive, On Premise Storage, Data Tape, Cloud
Back-Up	Copy Locations	Text	Details of additional copies of the content and where they are stored. This can include basic information about back-up procedures.
Media Formats		Text using a set list of terms	A list that specifically identifies any physical media types that content is stored on, e.g. external hard drive, USB drive, CD-RW.
Physical Media Location		Text	Information on where physical media containing digital content is stored. This information could be captured in the Storage Location field, but you may also wish to separate out physical locations to make this information easily accessible.
Unreadable Media		Text	Details of any unreadable physical media that is included in the digital content. This may be due to issues such as a lack of equipment or errors when attempting to access the media.
Storage Notes		Text	Notes on the storage of the content that do not fit under any other heading.
Priority		Text using a set list of terms	An indicator of the priority level of the content for processing and/or preservation. This might be a simple Yes/No or use levels such as High/Medium/Low.

Heading Name	Alternative Name(s)	Suggested Format	Description
Accompanying Metadata		Text	Details of any metadata provided with the content and where it is stored.
At Risk		Yes/No	Is the content currently at risk?
Risk Score		Numeric	A numerical representation of the risk level of the content to facilitate sorting and prioritization by risk level.
Preservation Planning		Text	A summary of preservation plans that have been developed for the content. This may include a link or file location for specific plans.
Preservation Actions	Migrations	Text	A summary of preservation actions that have been carried out on or for the content. This might include format migrations or emulators that are available.
3rd Party Preservation		Text	Details of any preservation responsibilities that have been deprecated to a third party. For example, where preservation is carried out by a partner organization, a national data repository, or a 3rd party service.
Meets Minimum Metadata		Yes/No	Is the digital content accompanied by the minimally required metadata as established by your organization's internal standards?
Virus Checked		Yes/No	Has the content been virus checked since it was received.
Catalogued		Yes/No	Has the content been catalogued?
Appraised		Yes/No	Has the content been appraised?

Heading Name	Alternative Name(s)	Suggested Format	Description
Suitable for a Volunteer		Yes/No	Is the content suitable to be processed and/or catalogued etc. by a volunteer?
In Repository		Yes/No	Has the content been ingested into the organization's digital repository?
Disposal		Numeric or Text	A number or description of files that have been deleted after appraisal or a retention period has expired.
Disposal Date		Date using a set format	The date that content was deleted.
Access Level		Text using a set list of terms	A description of the level of access that is currently provided, e.g. Open access, not accessible, limited content accessible, embargoed
Access Copies		Text	Details of any access copies that are available and/or a pointer to their storage location



## Appendix Two - Example Interview Questions

The following is a set of example questions that might be used for an interview to gather information about the digital content that should be recorded in your organization's DAR. They can be used as a starting point for developing your own question set: adding, removing, or tailoring the questions as needed to reflect the information you need to gather, how your interviews will be carried out, and organizational culture and working practices. This list is meant to be relatively comprehensive, so it is unlikely you would aim to cover all of these in a single interview session. Additionally, you may wish to consider the terminology used (are there particular terms used in your organization? Will you need to explain terms?)

1. How does your role relate to the creation, acquisition, and/or management of digital content?
2. Is there any particular digital content of digital content that you create, acquire, and/or manage?
  - a. What is their purpose/significance?
  - b. Is it important that they be preserved?
3. Who owns the digital content? (e.g. the organization, a partner organization, a donor, etc.)
  - a. Are there any agreements in place in relation to ownership, use, and/or intellectual property rights?
4. Who has ultimate responsibility for managing the digital content?
5. What types of content are included? (e.g. file formats)
6. How much digital content is there? (e.g. in MB or GB)
  - a. Is the digital content expected to grow? Can you estimate by how much and when?
  - b. Are there multiple versions of particular content?
7. What tools, equipment, and/or software are used in the creation, acquisition, and/or management of the digital content?
8. Do you maintain, or know of, any documentation in relation to the digital content? (e.g. metadata, technical documentation, catalogue records, a data management plan.)
9. Where is the digital content stored?
  - a. Is it backed-up regularly? How?
  - b. Are there security measures in place? (e.g. permissions needed to access, password protection)
10. Does the digital content contain any sensitive data? (e.g. personal data, proprietary information about the organization that should not be shared)
11. Are there any particular risks to the digital content?
12. How long do we need to preserve and retain the digital content for?
13. Has there ever been any data loss?

- a. Could the digital content be recreated if it were lost?
14. Is access provided to the digital content? (e.g. to other colleagues, research partners, members of the public)
- a. Who are the current users of the content?
  - b. Are there any other groups who might use the content in the future?
  - c. Are there any ethical issues in relation to providing access? (e.g. the content includes information first nations/Indigenous peoples or vulnerable groups)
15. Are there any other people you think I/we should talk to as part of this information gathering process?
16. Is there anything else you would like to add?