

# Novice to Know-How Module Text

## Course 7: Providing Access to Preserved Digital Content

# Module 4: Resource Discovery

The development of this course was funded by The National Archives (UK) as part of the "Plugged In, Powered Up" digital capacity building strategy.

## 1. Introduction.

In this module, we will address resource discovery specifically as it relates to preserved digital content. The module will first explain what resource discovery means to users and organizations before discussing the kinds of processes and systems it involves, the types of metadata and documentation needed, resource discovery metadata standards, persistent identifiers and schemes, and basic steps to provide descriptive metadata and documentation about preserved digital content in a discovery platform so that it can be discoverable by users.

## 2. Resource Discovery is Finding Digital Content.

Resource discovery is about making sources of information findable to users and therefore, facilitates access. In archives, resource discovery has traditionally been provided through finding aids or catalogues. There are, however, some fundamental differences between digital and analogue resource discovery.

To ensure that our digital resources are accessible, organizations must provide suitable means of discovery to users so that they can find digital content of interest. Although different users will have different information needs, there is a common expectation that they will be able to find digital content of interest by searching or by coming across something while browsing collections.

Given the advancements and capabilities of web search engines, users may also have expectations that discovering and accessing digital content will be intuitive, straightforward, and seamless; that they will be able to visit the website, find an online discovery platform; search for specific content or browse collections; find something of interest; and locate a link or point of access to the requested content.

### 3. Resource Discovery Requires Processes and Systems.

Although users may have these expectations, it is not always easy to have a robust process in place for resource discovery within your organization.

The capabilities and expectations of your users must be considered next to your organization's current capacity and capabilities. For example, resource discovery can be limited by the functionality of an organization's access or storage environment.

A good starting point is to ensure that a few fundamental processes and systems are in place.

Effective resource discovery requires, at the least, that the organization has implemented a basic discovery and access mechanism. The DPC Rapid Assessment Model (RAM) has this as the requirement for reaching the 'basic' level of Discovery and Access. This means:

- Basic resource discovery exists for some digital content.
- Users can view or access digital content and metadata, either remotely or on-site.

### 4. What do we mean by discovery mechanism or platform?

By 'discovery mechanism', we simply mean a method and means for users to search and find information about the digital content in collections.

For this information about digital content, we give access to metadata to enable discovery. Metadata itself can be a mechanism for resource discovery because it provides the information that is used to communicate and connect users to digital content.

A search engine can be broadly understood as a discovery mechanism, but here we are specifically talking about catalogues, finding aids, or another 'discovery platform' that utilize metadata to offer users a means to search, browse, and find relevant content specific to your repository, archive, or collections.

### 5. Discovery Mechanisms: Analogue vs Digital.

It is likely that discovery mechanisms for analogue resources at your organization will already exist, for example catalogues and finding aids that can be accessible to users on-site or online.

But discovery mechanisms for digital resources can utilize data and metadata for the benefit of users. Digital content can be represented in multiple arrangements through the manipulation of data and metadata online, some of which may be more useful to users.

In addition to providing descriptions of digital content in traditional catalogues and finding aids, there are other methods and means for enabling resource discovery through other mechanisms, some of which we will highlight in the next slides.

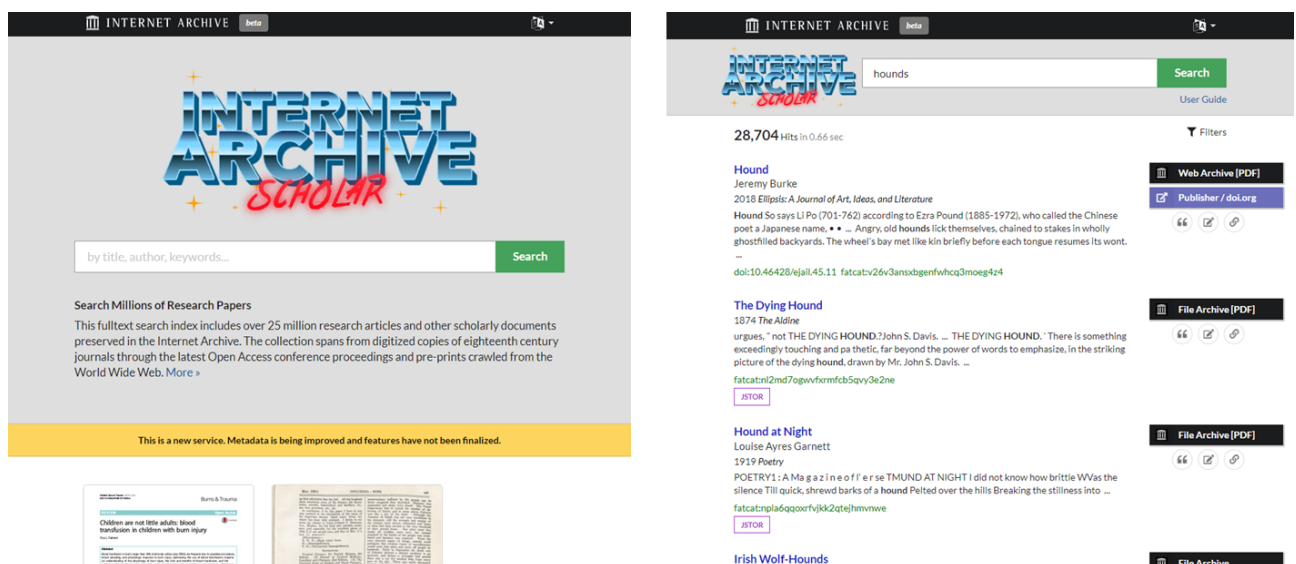
## 6. Digital content can provide new forms of resource discovery.

The nature of digital content and collections provide opportunities for new forms of resource discovery. Other systems and applications can utilize the metadata provided to offer different routes for discovery. Some examples include:

- Full-text searching and advanced search techniques, for example a search engine integrated with existing catalogue search tools.
- Advanced visualization, for example timelines for web archives.
- Geospatial discovery, for example map-based searches.
- Curated content, for example subject-themed browsing.

## 7. Full-text Search: Internet Archive Scholar.

Link to example: <https://scholar.archive.org/>.

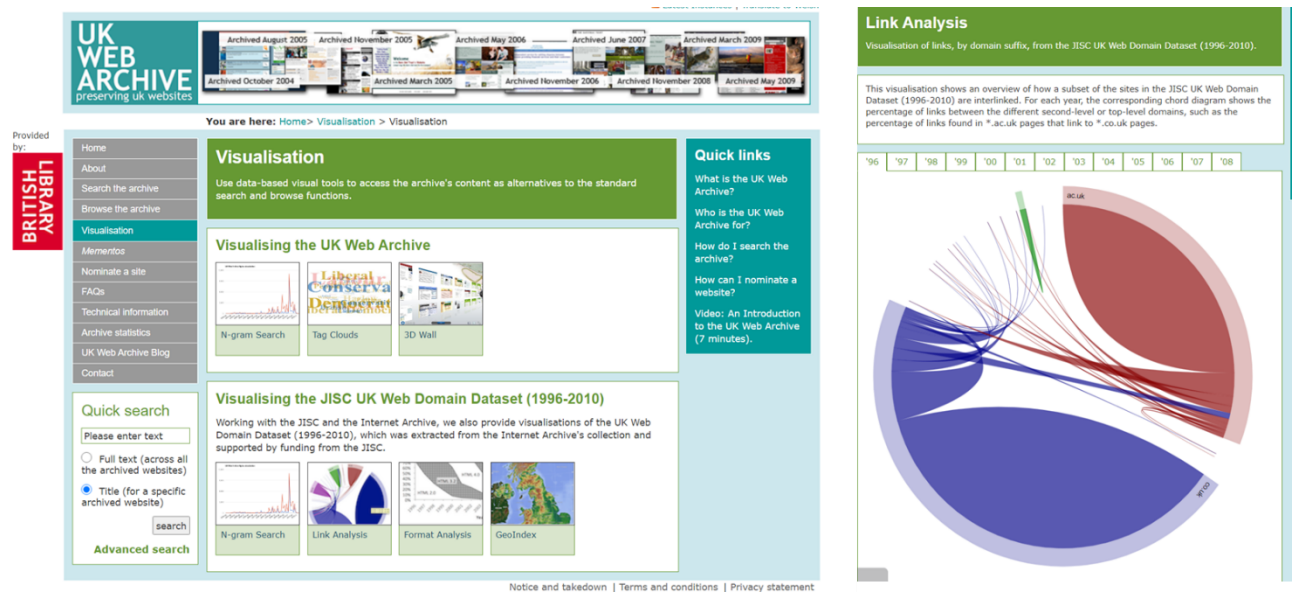


The Internet Archive Scholar search offers users a full-text search index of over 25 million research and scholarly resources preserved in the Internet Archive—both digitised and born-digital. The content includes digitized print materials, public web content, and general materials from their collections. It utilizes metadata from an open user-editable catalogue or scholarly work, and from text and data mining to enable its full-text searching.

## 8. Visualization: The UK Web Archive Visualising the JISC UK Web Domain Dataset (1996-2010).

Link to example:

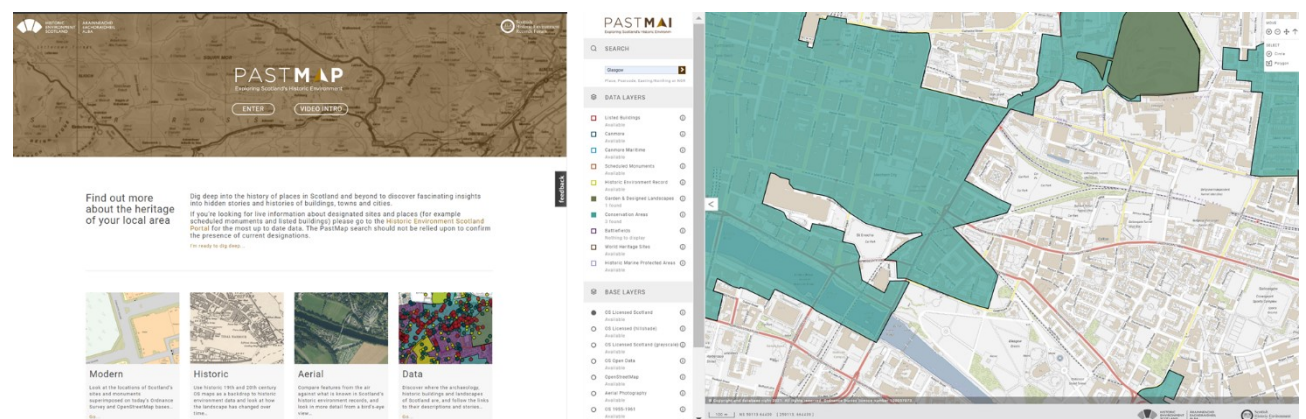
<https://www.webarchive.org.uk/wayback/archive/20181019114343/http://www.webarchive.org.uk/ukwa/visualisation>.



Another example illustrating visualization comes from a collaborative project done by the UK Web Archive, Joint Information Systems Committee (JISC) and the Internet Archive. The project took a specific JISC UK Web Domain dataset, containing websites crawled between 1996 and 2010 by the Internet Archive, and developed visualizations for presenting and searching the content. In particular, the Link Analysis visualization tool created during the project offers users an overview of how a subset of the sites in the Dataset were interlinked.

## 9. Geospatial Discovery: Historic Environment Scotland PastMap.

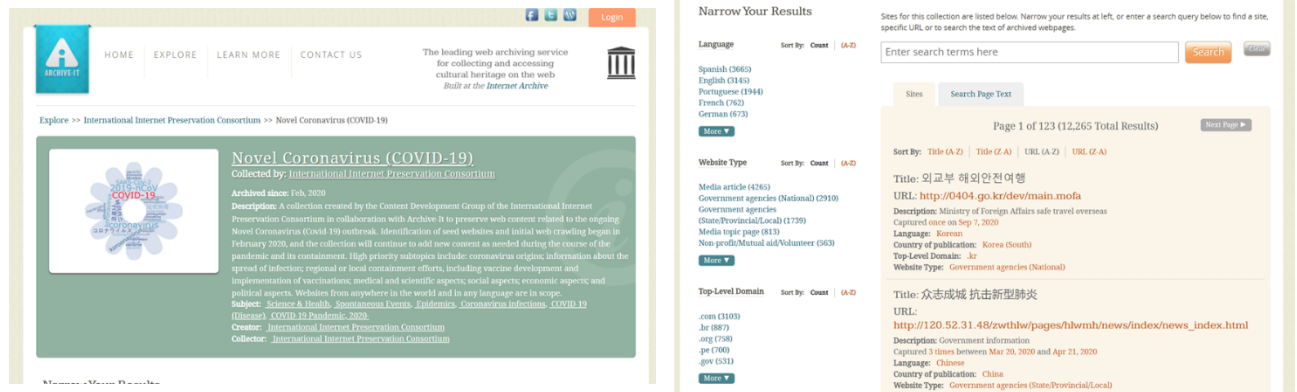
Link to example: <https://pastmap.org.uk/>.



The PastMap is an excellent example of discovery that utilizes geospatial data with other datasets held by Historic Environment Scotland and other partners. It provides a way for users to find and view historical and archaeological sites in Scotland. Its search allows users to browse maps or search for a specific site, look at different layers relating to the different datasets, and click on links to connect with resources which provide more details about the history and heritage of the area.

## 10. Curated: International Internet Preservation Consortium Novel Coronavirus.

Link to example: <https://archive-it.org/collections/13529>.



The screenshot shows the IIPC website interface. The left sidebar contains navigation links: HOME, EXPLORE, LEARN MORE, and CONTACT US. The main content area features a header for the 'Novel Coronavirus (COVID-19)' collection, collected by the International Internet Preservation Consortium. Below the header, there is a description of the collection, its purpose, and a list of websites included in the collection. The right sidebar shows search results for the collection, including a search bar, a list of websites, and a table of results with columns for Title, URL, and Description.

One more recent example of curated content for resource discovery is the Novel Coronavirus (COVID-19) collection created by the Content Development Group of the International Internet Preservation Consortium (IIPC) and Archive-It. It was created to preserve web content relating to the Coronavirus outbreak while enabling users to find COVID-related preserved content through keyword or faceted search.

## 11. No 'One Size Fits All' Approach.

The functionality of different discovery mechanisms can enhance or limit how seamless the experience is for users, and there is often no easy 'one size fits all' approach.

You should be familiar with existing systems and consider drawing on their functionality to provide a more familiar search and browsing experience for users at your organization.

At the same time, you should not let the functionality of existing mechanisms drive your decisions about how you should describe and arrange digital content.

In other words, discovery mechanisms for digital content are rarely developed in isolation. Think of existing processes and systems as an initial baseline and mode of access fit for current purpose. Over time, their functionality and effectiveness can be re-evaluated in light of what you find through user needs analysis.

## 12. The Role of Metadata in Resource Discovery.

Metadata and documentation of digital content in your collections are essential for undertaking digital preservation activities and for enabling and enhancing resource discovery. Metadata allows information about digital content to be read by machines and systems (interoperable) and also readable to users. Users will be unable to find and use preserved digital content without adequate metadata and documentation.

Metadata captured at the point of accessioning and ingest processes can be reused for resource discovery processes. The preservation metadata can be created, captured, added,

and incorporated into various discovery mechanisms like those mentioned in the previous slides.

But, before we dive into metadata for resource discovery, let us quickly recap metadata for digital preservation which was covered in the previous Novice to Know-How course modules. We will build on this to see how metadata and documentation collected for preservation activities can be used for resource discovery.

## 13. A Brief Recap on Preservation Metadata.

In earlier Novice to Know-How modules, we learned that preservation metadata refers to metadata used for preservation actions and is comprised of metadata that can fall into one or more categories.

- **Administrative metadata** about the management of a digital record (e.g. who created it, who is allowed to access it, or what rights issues are associated with it).
- **Structural metadata** about a digital object's internal structure (e.g. page, index).
- **Technical metadata** about the technical properties of a file or the hardware and software environments required to render or process the content (e.g. format, operating system).
- **Descriptive metadata** about the digital object and its content (e.g. creator, topic, subject, keywords), which make content easier to find in a search.

A two-level approach to capturing metadata was also described. Collection-level metadata can help users locate content that is part of a collection and understand the overall context of the collection or the files in it. File-level metadata helps users locate content in specific files, understand them and make use of them.

## 14. File Manifests and Digital Asset Registers.

The two-level approach to capturing metadata recommends recording the information in two documents: file manifests and digital asset registers.

At the file level, we can capture and generate preservation metadata using characterization tools like DROID. The resulting file manifests contain metadata that can also be helpful for resource discovery, for example file names, relevant dates, and formats. Additionally, the file manifests can provide information about the number of processed or unprocessed materials in collections to let users know what is available.

At the higher level, or collection level, we can create and capture preservation metadata in digital asset registers, which can also help resource discovery, for example collection name, creator, extent (the amount of content within the collection), owner, rights, access conditions, and provenance. This information provides contextual information for users to search or learn more about the content as well as whether there are any current access restrictions or permissions required to access the content.

In fact, some organizations have used a digital asset register as a basic finding aid in the absence of other discovery methods!

## 15. Build on what you have to advance resource discovery.

There should be an existing minimum standard of information required for preservation activities, with required preservation metadata created or captured in file manifests and digital asset registers.

Some of the preservation metadata that has been already collected during these activities can be used for resource discovery. It can be harvested and reused for catalogue records to provide information about the resources that can be discoverable to users through appropriate catalogues or other discovery mechanisms.

But it is also good to periodically revisit and review what you already have in place as the minimum standard of information to consider if any additional metadata or documentation is needed to enhance discovery for the benefit of your users.

## 16. Knowledge Check (Interactive Slide).

**Knowledge Check Question:** Which of the following is *not* true?

- Descriptive metadata can support users in identifying content of potential interest.
- Resource discovery encompasses a range of techniques for searching and browsing the content of the repository.
- Preservation metadata should only include technical metadata.
- Digital content can be described at both the file level and collection level.

The correct answer is "Preservation metadata should only include technical metadata". Preservation metadata can include other metadata types.

## 17. Tasks: Assess Current Inventory.

As mentioned, it is good to review what you already have to consider how much—if any—additional metadata or documentation should be created or captured to enhance resource discovery.

At a minimum, you should already have an inventory with enough preservation metadata to know what your content is, and this may include a combination of different metadata types in file manifests and digital asset registers.

You can use what you know about what you have (the existing file metadata) to assess what is needed for your discovery system to create, edit, or synchronize metadata appropriately.

As you carry out this assessment, it is helpful to consider the following questions:

- What types of digital content and objects comprise your collections?
- Are they institutional records, private or personal archives, special collections, research or project outputs?
- At what level is the content described?
- How much metadata has been captured or collected, and in which areas?

- Is that information provided in the current catalogue, finding aid, or discovery mechanism at your organization?

## 18. Tasks: Analyze User Needs.

Resource discovery is about making digital content findable to users. Therefore it is critical to conduct a user needs analysis to understand their expectations and basic requirements for discovering digital content in collections.

Gather relevant data regarding user groups when determining how much—if any—additional metadata is needed or whether new metadata standards for resource discovery should be adopted.

Some questions to ask during the user needs analysis might include:

- How do users currently search for resources?
- Can users easily find and discover digital content using the existing catalogue, finding aid, or discovery mechanism at your organization?
- What metadata do users need and why? (e.g. would technical metadata such as bit-depth for images, or compression rate for audio files be necessary)?
- What other requirements are identified by the user needs analysis?
- Do you expect user needs to change over time?

## 19. Tasks: Review Existing Standards.

If you have not already, you may want to establish a minimum standard of information required for resource discovery that uses one or more adopted metadata standards.

We highlighted two established digital preservation metadata standards in previous course modules: PREMIS (PREservation Metadata: Implementation Strategies) and Metadata Encoding and Transmission Standard (METS). However, neither PREMIS nor METS explicitly include metadata for resource discovery. So we may need to look to other existing resource discovery metadata standards for guidance.

As you review any existing standards, some helpful questions to ask include:

- What are the specific metadata standard or standards your organization currently adopts? Why were they selected?
- Do existing metadata elements and standards meet the needs of users? Are there any areas that are overlooked that are relevant to digital content?
- How much time and effort does it take to implement? Can the metadata be harvested from existing sources, are there tools available to automate the process?

## 20. Knowledge Check (Interactive Slide).

**Knowledge Check Question:** Before making a decision on whether additional metadata is needed for resource discovery, it is helpful to first (choose all that apply):

- Conduct or review findings of a recent user needs analysis to understand how those users typically search or browse for content of interest

- Measure or estimate the time and effort it takes to implement current methods of metadata creation or capture
- Look at your current inventory to see what types of digital content and objects comprise your collections
- Conduct interviews with external stakeholders to gather their thoughts and opinions on the best resource discovery metadata standard

The correct answers are “Conduct or review findings of a recent user needs analysis to understand how those users typically search or browse for content of interest”, “Measure or estimate the time and effort it takes to implement current methods of metadata creation or capture”, and “Look at your current inventory to see what types of digital content and objects comprise your collections”.

## 21. Standards for Resource Discovery.

There are a number of existing established standards for resource discovery you may consider adopting and implementing. These metadata standards, summarized in the next slides, will help guide decisions on which additional metadata and documentation you may need to capture, and how digital content is described and presented to users. The standards can be used in conjunction with PREMIS, METS, and other metadata standards.

Because there is no 'one size fits all' metadata standard for resource discovery, the best approach is to choose a standard based on what you know about your users' needs, existing discovery and access systems, and the types of digital content in your collections.

## 22. A Few Tips for Selecting a Metadata Standard.

Given that there is no 'one size fits all' solution, no one specific descriptive standard is recommended for everyone.

It is important to choose a standard with interoperability in mind. Interoperability enables metadata to be discovered and reused by computer systems on its way to reaching expanded audiences of users, which is why choosing standards for ensuring interoperability are important—to allow machines to process our descriptive metadata.

For this reason, it is also worth considering how the creation and capture of required metadata for a standard can be harvested from existing sources and if there are tools that can be used to incorporate or automate the process. Relevant questions to ask may include: Where will descriptive metadata reside? How will it be accessed? And will users be permitted to search the content itself?

## 23. Dublin Core.

One of the most widely used standards for discovery metadata is the Dublin Core Metadata Element set, which has 15 descriptive elements to facilitate simple resource discovery online.

It is a simple and generic metadata schema for resource description, designed to be cross-disciplinary, international, and genre-independent, so it can be implemented by different types of organizations.

**Pros.**

- It is easy to use and may be well-suited if you are first getting started and have limited resources, for example you can easily incorporate it into spreadsheets to create a Dublin Core XML file with minimal effort.

**Cons.**

- It is limited beyond its basic descriptive elements (e.g., rights, preservation), and despite its simplicity there can be problems with applications, for example distinguishing the 'creator' from authors, publishers, or other contributors.

## 24. ISAD(G).

One of the most widely used metadata standards for creating finding aids for archival collections is General International Standard Archival Description, or ISAD(G). It provides a list of 26 elements and rules for description, 6 of which are mandatory.

It describes a hierarchical approach to cataloguing archival objects, including a hierarchy of description and what information should be included at what level, offering detailed guidance for organizations with archives.

**Pros.**

- The guidance is detailed and comprehensive, aiming to reduce repetition and allow grouping of material from particular collections (fonds). It is well-suited if you are working with archives or digital content in defined collections.

**Cons.**

- It is designed with multi-level descriptions in mind, which may require time and effort to implement fully. There can also be semantic and interoperability problems when applied to systems designed without a hierarchical logic.

## 25. DACS.

Describing Archives: A Content Standard (DACS) defines 25 elements - a refinement of the 26 ISAD(G) elements - that are useful in creating systems for describing archival materials

DACS is broken down into two parts:

- Part One gives a set of output-neutral rules for describing archival material at different levels, which can be applied to different material types.
- Part Two gives guidelines for creating authority records in archives. These authority records can then be incorporated into the metadata of an archival record, or separate authority records can be created.

**Pros.**

- It is well-aligned with other existing archive standards of ISAD(G) and the library-focused cataloguing rules of Resource Description and Access (RDA), making it suitable for archive and library collections.

### Cons.

- It is also a standard for creating access tools such as catalogues or finding aids, making it easier to adopt and implement but may also limit which tools or systems can be used.

## 26. Implementing a Standard: Consistency is Key.

Once you select a standard, it is essential that it is properly and consistently implemented to ensure that the metadata in findings aids, catalogues, and discovery platforms will be useful to users and interpretable by many different systems.

This advice also generally applies to any process, system, or mechanism you put in place for resource discovery.

## 27. Persistent Identifiers.

Another means of improving resource discoverability are persistent identifiers (PIDs). PIDs provide a long-lasting reference to a resource.

All too often, links fail to take you to the referenced resource you expect. This can occur for technological reasons like server failure, but also human-created failures. For instance, organizations transferring journals to new publishers, or reorganizing their websites, can lead to broken links. Not only can this be frustrating for users but the consequences can be serious if the linked resource is essential for legal, medical, or scientific reasons.

Typically a persistent identifier has two components: a unique identifier; and a service that locates the resource over time even when its location changes. The first helps ensure a digital resource's provenance (it is what it purports to be), whilst the second will ensure that the identifier leads to the correct current location.

Persistent identifiers can also be used 'behind-the-scenes' within a repository to manage some of the challenges in cataloguing and describing, or providing intellectual control and access to born-digital materials.

## 28. Examples of Persistent Identifier Schemes.

Below are examples of PID schemes with the kinds of resources they are typically assigned.

A **Digital Object Identifier (DOI)** is typically assigned for journal articles, scholarly publications, and a range of data objects.

A **Handle** for Internet resources, with a central registry to resolve URLs to the current location.

An **Archival Resource Key (ARK)** for archival resources.

A **Persistent Uniform Resource Locator (PURL)** for redirecting web resources using standard HTTP status codes, acting as a permanent web address which contains the command to redirect to another page.

A **Uniform Resource Name (URN)** for mapping of namespaces into a single URN namespace, which may be applied independently of the availability of the identified resource.

These kinds of persistent identifiers assigned to digital resources not only provide a link to locate a resource over time but also connect to sets of metadata describing an item or link to the item itself to reliably identify and verify the content over time.

## 29. Resource Discovery in the Levels of Born-Digital Access.

So far, we have covered what resource discoverability means, what it entails, how metadata enables it, and introduced some of the approaches and standards for improving it at your organization.

It is now critical to talk about resource discovery in terms of reaching benchmarks through achievable actions.

Resource discovery interrelates with all five areas of action in the DLF Levels of Born-Digital Access, but we will look specifically at what is needed to reach Level One of Description—to 'provide required descriptive elements for a collection-level record and at least one descriptive note about the processed digital materials.'

## 30. Action 1: Description is available in an online searchable collection discovery platform. (slide 1 of 2).

The first action involves making metadata and documentation available to users online so that they can search and find digital content of interest without having to visit on site.

If your organization has an online catalogue or finding aid, which includes collection-level description of some of your digital content that is searchable for users, you have accomplished this action!

If you have not already, make sure that the online catalogue includes access arrangement and restrictions. Provide physical access and technical requirement notes about processed digital materials to describe how they can be accessed or, if they are inaccessible (due to being unprocessed or other restrictions), explaining how users can request access.

## 31. Action 1: Description is available in an online searchable collection discovery platform. (slide 2 of 2).

If you do not have any digital content described in an existing online catalogue or finding aid, you should create a collection-level record of digital content (where access rights permit) in the system using the standards you have chosen. You can also save time and effort by using relevant metadata extracted from file manifests and digital asset registers for creating the record.

If you do not have an online catalogue or finding aid, you will need to develop or acquire one. At this point, it can simply be a collection-level catalogue.

You might want to consider using freely available discovery platforms. For example, for archives based in the UK, The National Archives' Manage Your Collections application or Archives Hub system allows you to create, share, and manage descriptions of your collections.

At the very least, there should be a web page of collection-level descriptions with searchable text. For example, elements of your digital asset register can be made available online to provide a basic finding aid for users.

## 32. Action 2: Provide a count of unprocessed content in a collection. (slide 1 of 2).

The second action interrelates to the first but specifically requires you to include a count of unprocessed materials in collection-level descriptions to provide users with information about them.

This can be included in metadata elements like Extent, Scope and Content, or another specified in the metadata standard you are using.

If this is already provided to users in your online catalogue or finding aid, you have accomplished this action!

## 33. Action 2: Provide a count of unprocessed content in a collection. (slide 2 of 2).

If you have not provided this count of unprocessed materials, do so by recording it in collection-level notes.

As mentioned before, you can record it in metadata elements like Extent, Scope and Content, or another element that aligns and adheres to the metadata standards you've adopted.

Remember that you can save time and effort by revisiting what you have already collected about digital materials in file manifests and digital asset registers. The file manifest, in particular, allows you to generate and provide information about the number of files, formats, and size of processed and unprocessed collections to share with users.

## 34. Knowledge Check (Interactive Slide).

**Knowledge Check Question:** True or False, to reach Level One of Description, you must provide a faceted search to users.

The correct answer is 'False'. You do not have to provide a faceted search to reach Level One of Description in the Levels of Born-Digital Access.

## 35. From Level One to Level Two of Description.

Achieving Level One will set the stage for the next actions needed to achieve Level Two of Description in the Levels of Born-Digital Access—to 'provide required descriptive elements for a multi-level record and specified descriptive notes.'

This module will not go into detail on these actions, but they include:

- Describing the intellectual and physical arrangement of the materials.
- Providing the additional descriptive elements deemed necessary by established description procedures.
- Providing aggregate level description and intellectual arrangement throughout collections according to established processing standards.
- Sharing technical limitations with regard to unprocessed materials.

For those of you with established resource discovery systems, you can see how access to born-digital records can be supported through existing processes and systems, and by reuse of metadata you capture on ingest.

## 36. Module Summary.

This module explored what resource discovery means in the context of access; how it relates to users' expectations, requires systems and processes, how metadata itself can be a mechanism for enabling and enhancing access, and the importance of persistent identifiers.

While there is no 'one size fits all' approach for optimizing resource discovery, all organizations should provide collection-level descriptions of some of the digital content in the collections through an online discovery mechanism noting the count of unprocessed materials in the collection or collections. Doing this will achieve Level One of Description and set the stage for achieving Level Two.