

THE

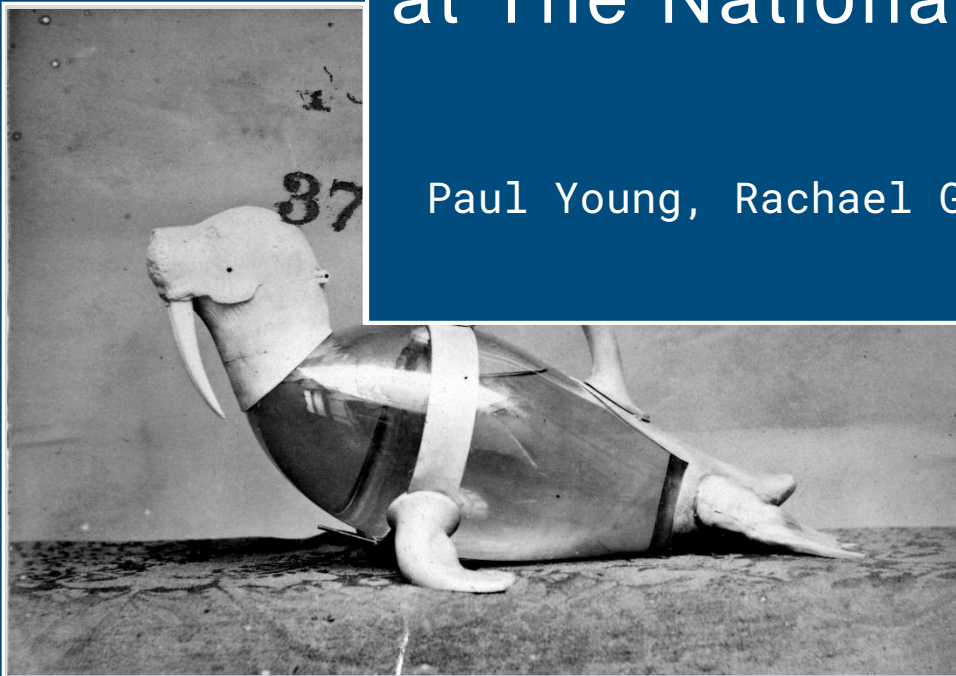
NATIONAL

ARCHIVES



Cloud Transfer: Lessons learned at The National Archives (UK)

Paul Young, Rachael Gardner, Balint Csollei



Agenda



Google transfer Proof of Concept (POC)



Cloud to Transfer Digital Records (TDR) Discovery



M365 export tests



GKIM Engagement

Google transfer Proof of Concept (POC)

- Long term plan to incorporate Cloud to Cloud transfer processes into Transfer Digital Records service (TDR), not just for Google but also Microsoft/SharePoint.
- Short term bodies using Google Workspace were closing and wanted to transfer records to us, requiring an interim approach.
- This allowed us to learn about cloud transfer and different possible approaches and challenges.
- As part of our research we looked at transfer of ownership inside Google Workspace. For TDR development the team has looked at RClone.
- For this POC the Google Drive API was used.



Metadata

<code>createdTime</code>	<code>string</code> The time at which the file was created (RFC 3339 date-time).
<code>modifiedTime</code>	<code>string</code> The last time the file was modified by anyone (RFC 3339 date-time). Note that setting <code>modifiedTime</code> also updates <code>modifiedByMeTime</code> for the user.
<code>modifiedByMeTime</code>	<code>string</code> Output only. The last time the file was modified by the user (RFC 3339 date-time).
<code>viewedByMeTime</code>	<code>string</code> The last time the file was viewed by the user (RFC 3339 date-time).
<code>sharedWithMeTime</code>	<code>string</code> Output only. The time at which the file was shared with the user, if applicable (RFC 3339 date-time).
<code>quotaBytesUsed</code>	<code>string</code> (<code>int64</code> format) Output only. The number of storage quota bytes used by the file. This includes the head revision as well as previous revisions with <code>keepForever</code> enabled.
<code>version</code>	<code>string</code> (<code>int64</code> format) Output only. A monotonically increasing version number for the file. This reflects every change made to the file on the server, even those not visible to the user.
<code>originalFilename</code>	<code>string</code> The original filename of the uploaded content if available, or else the original value of the name field. This is only available for files with binary content in Google Drive.
<code>ownedByMe</code>	<code>boolean</code> Output only. Whether the user owns the file. Not populated for items in shared drives.
<code>fullFileExtension</code>	<code>string</code>

Files



Photo by [Wesley Tingey](#)
on [Unsplash](#)

Approach - [nationalarchives/dp-research-googletransferPOC](#)

Metadata

- Identified relevant metadata held in Google which we would want to preserve (dates, Google IDs, Google Parent IDs, checksums (only non Google native formats), mime type)
- Department identified selected material which we then used the Google API to collect metadata to extract as a CSV and converted to TNA template.
- The department then added additional metadata about access conditions (FOI exemption codes, closure period)

Files

- Used the metadata CSV to make another call to the API, using the Google ID to identify appropriate files to download and the correct export format via mime type.
- Validated the downloaded files (except for Google Native formats) using the MD5 checksum.

Security and authentication



Photo by [Georg Bommeli](#) on [Unsplash](#)

- Need to be able to connect to multiple organisations.
- Has to be an acceptable approach for government departments to obtain approval from their IT.
- Has to be secure (will be handling sensitive information)
- Need appropriate permissions to run script, read only was initially chosen but proved not enough in some cases where download had been restricted

Native cloud formats

- Google formats such as Google Docs, Slides and Sheets are cloud native and cannot be exported in their original formats
- Traditional preservation techniques like checksum validation are not effective
- Not unique to Google but most widely known example
- Multiple export formats exist but no single export provides a complete version which preserves the content and presentation of the original format
- Current solution is to create multiple export formats, e.g. for Google Docs would export a Microsoft Docx and a PDF



Native cloud formats

- Previous discussions with Google about a more comprehensive solution. Interest from Google in looking at a solution but sway of digital preservation community will be most effective if we have a clear idea of requirements which we want from them and other companies.
- Raised in Bit List 2023.
- Risk highlighted by Google recent decision to retire Jamboard, meaning that by the end of next year Google will delete all Jamboards so would need to be exported or migrated before that date. Products can change or be retired with little notice.



THE

NATIONAL

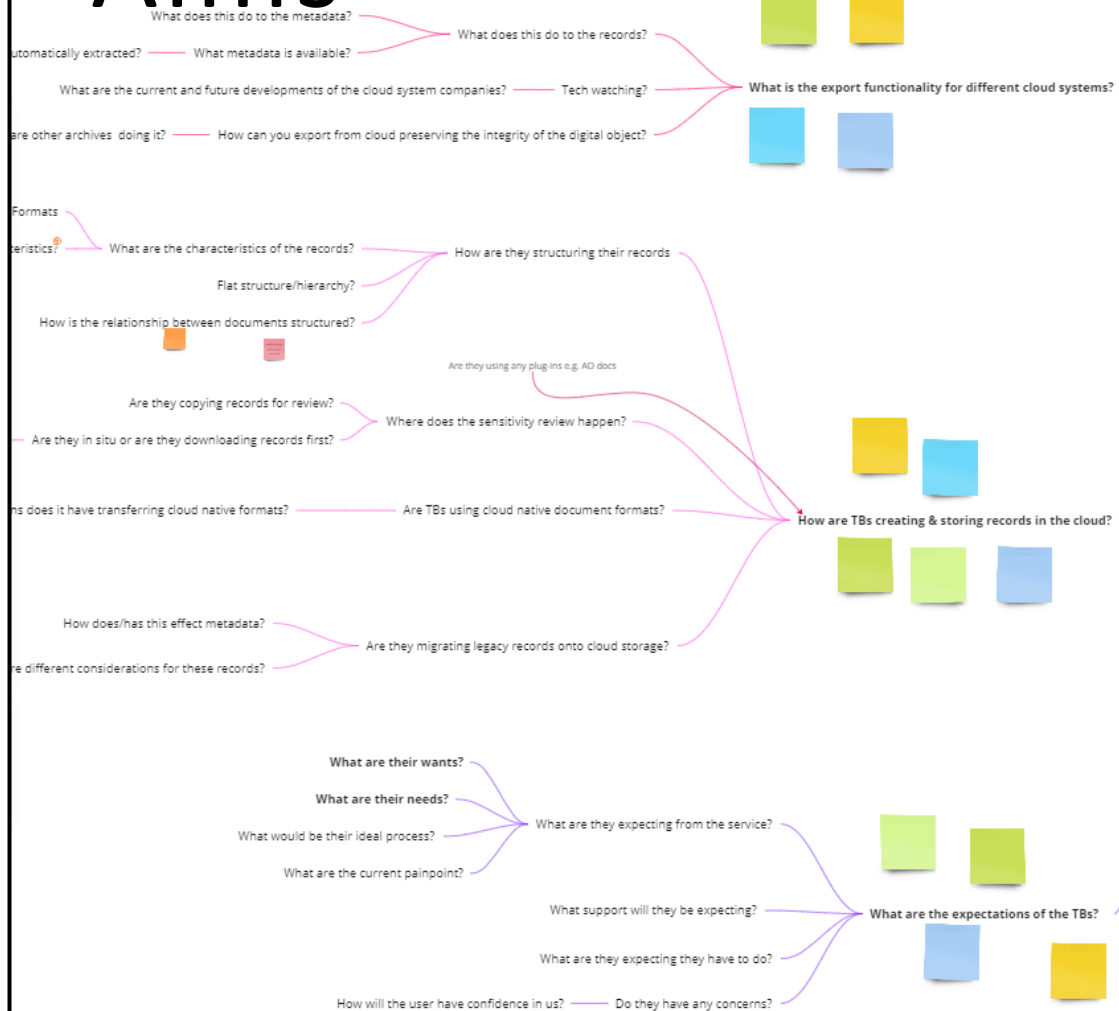
ARCHIVES

Cloud to TDR

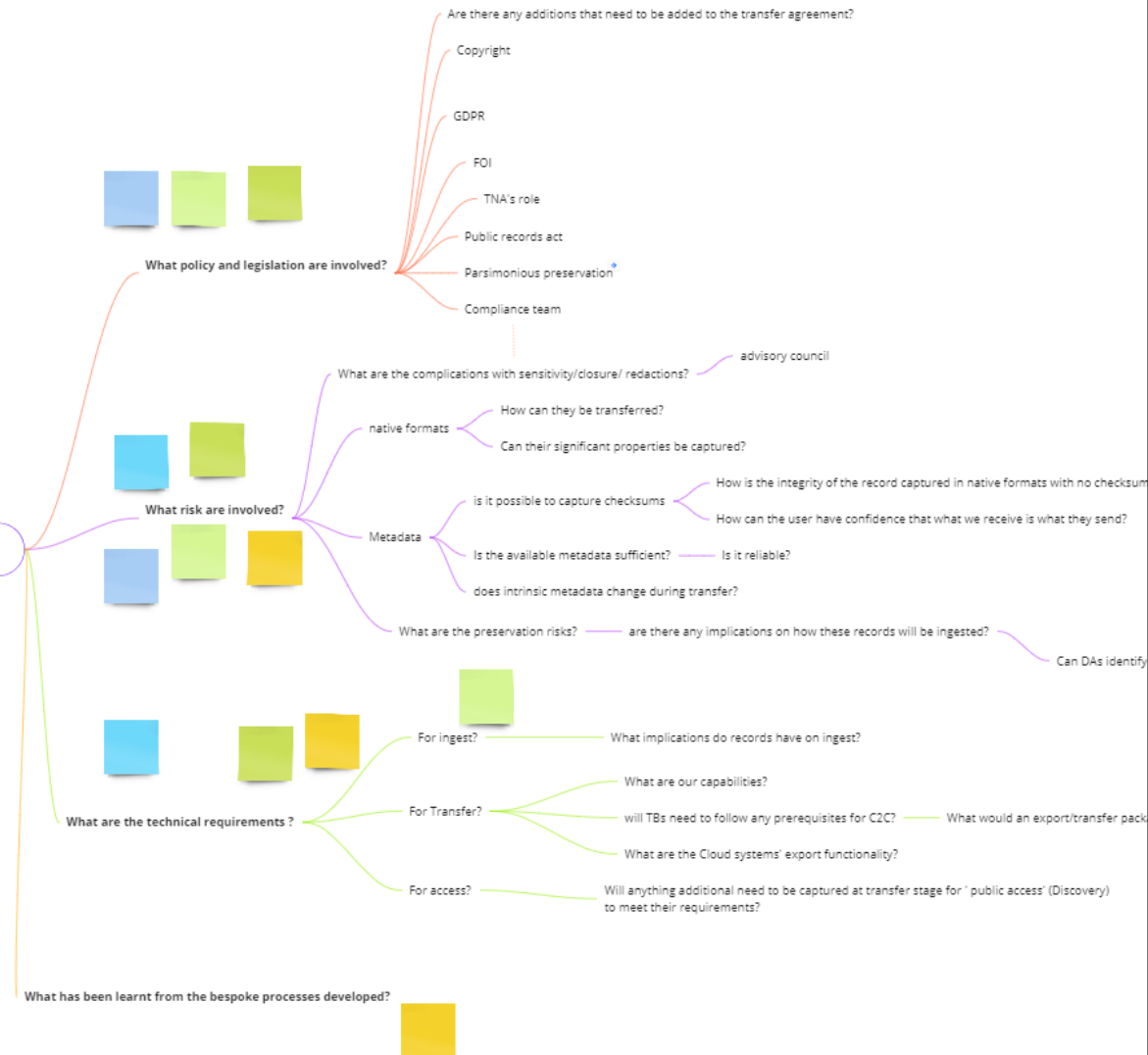
Discovery



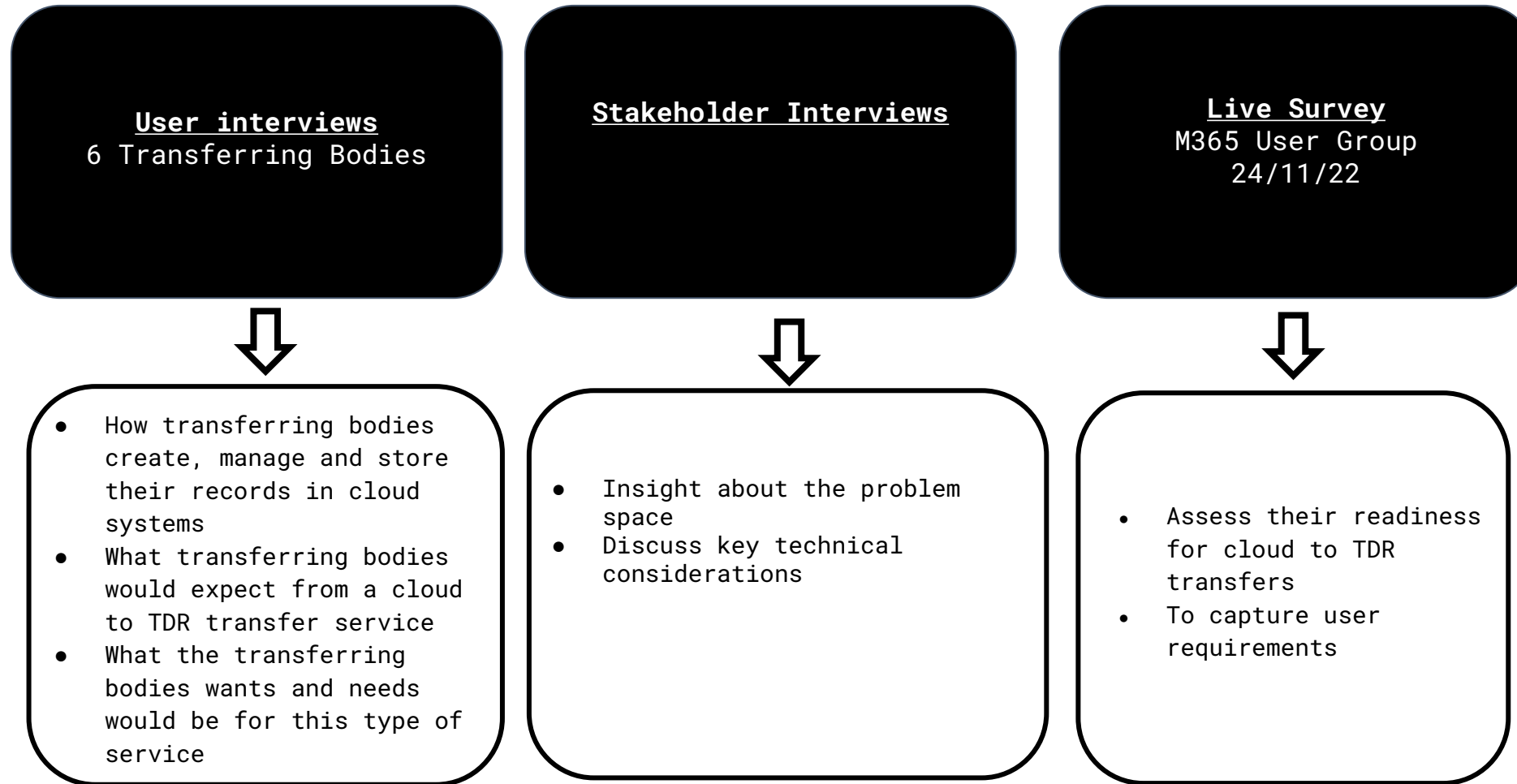
Aims



Questions



Methodology



User requirements

Needs

- **Cyber security** - the feature needs to smoothly integrate/ work with their internal IT security systems to allow for the movement of records from their cloud storage systems to TNA through TDR
- **Security accreditation** - the user needs to be assured of the security of the transfer process
- **Confidence** - in the integrity of the transfer and what they have transferred is what TDR has received
- **Transparency** - TDR to be transparent on the record journey
- **Accessible instructions and content** - accessible, clear and avoid overly technical language
- **Clarity on transfer requirements** - Users need clear guidance on what is required of them in terms of preparation for cloud to TDR transfer

Wants

- **Simple and easy** - the users want the cloud to TDR journey to be simple and easy, with intuitive UX that will allow for minimal human errors
- **Audit trail** - There is a definite want from the users for an audit trail of the records from transfer to discovery.
- **Transfer history** - There is a definite want from users for TDR to show a transfer history
- **Guidance on appraisal, sensitivity review and transfer preparation** - There is a strong user want for more support from TNA with the preparatory stages before actual transfer, as this is the biggest hurdle for users in their transfer experience

Technical considerations

Sharing and rights-based access

- General approach is export, and upload – seeing records as a package
- Potential for a different model

Metadata

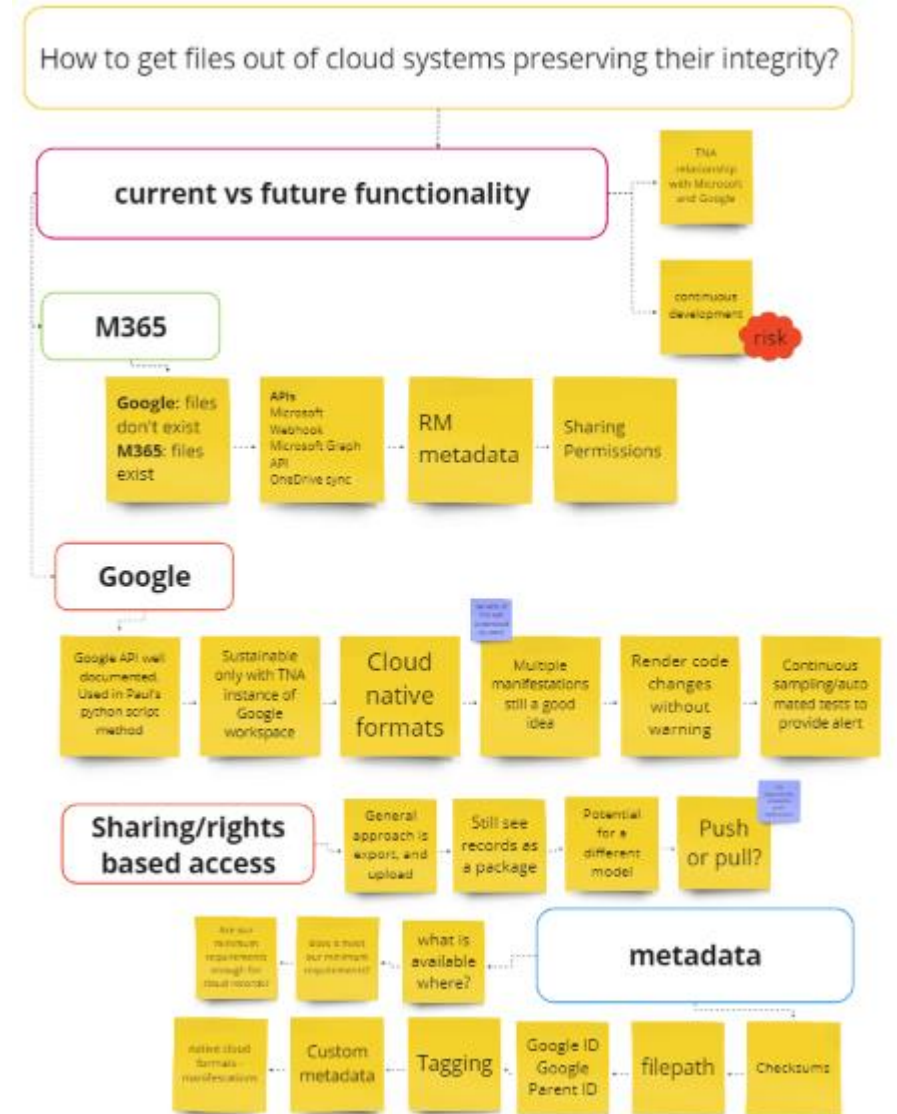
- What is available? Does it meet our minimum requirements?
- Are our minimum requirements enough for cloud records?

Risk

- Changing export functionality
- Continuous changes to google render code

Questions

- What if the filepath isn't enough to capture contextual metadata?
- Cloud/collaborative records – what/when is the record?
- Will there be a better solution for preserving google files in the future?



THE

NATIONAL

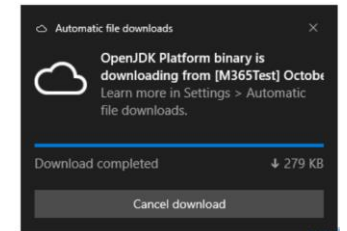
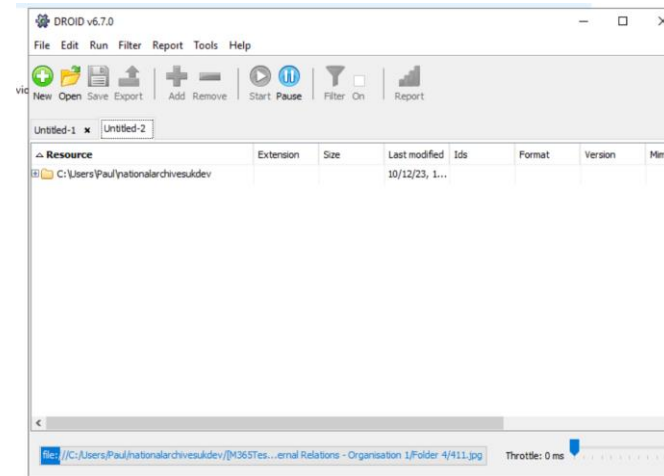
ARCHIVES

M365 export tests



TNA M365 export tests

- Internal TNA testing to look at export options:
 - ⑩ Exporting files and metadata using UI
 - ⑩ Used OneDrive Sync to test export and ability to run tools such as Droid over files held in SharePoint Online and OneDrive
 - ⑩ Tested extracting files and metadata using Powershell
 - ⑩ Two test environments to test sharing files across tenancies
 - ⑩ Environments offer future possibilities for testing



THE

NATIONAL

ARCHIVES

Engagement

Working with government to raise
standards in information management



Principles

Value

Authorities must manage their information in such a way that they can assess its current and future value.

Integrity

The authority must be able to rely upon and trust its information.

Accountability

The authority's information management must enable it to provide a clear and accurate account of its activity in accordance with its legal and other obligations.

Guidance

Code of Practice on the Management of Records issued under section 46 the Freedom of Information Act 2000

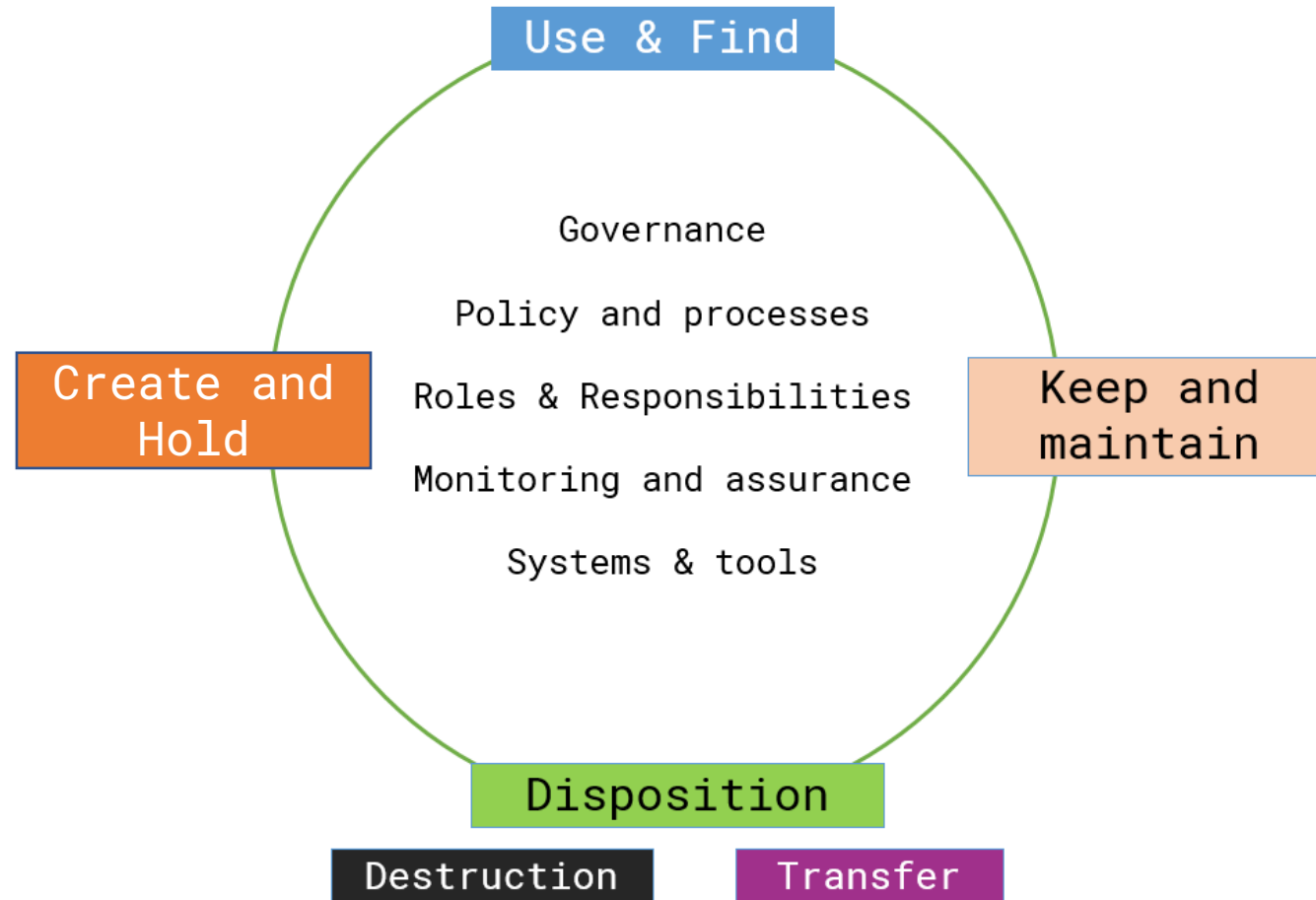
Code of Practice issued under section 46 of the Freedom of Information Act by the Secretary of State for Digital, Culture, Media and Sport providing guidance to public authorities on the keeping, management and destruction of records.

From: [Department for Culture, Media and Sport](#), [The National Archives](#), and [Department for Digital, Culture, Media & Sport](#)

Published 15 July 2021

 [Get emails about this page](#)

Applies to England, Northern Ireland and Wales



M365 & Google Workspace Maturity Models

- ❑ Putting the principles into practice
- ❑ Industry expert-defined approaches
- ❑ ~140 approaches/configurations in 5 areas and 31 questions
- ❑ Maturity levels benchmarked by M365 & Google Workspace User Groups
- ❑ Overall consensus on best practice, but areas of uncertainty remain
- ❑ Used by TNA's Information Management Assessment programme to self-assessing the effectiveness of their approach to information and records management in their M365 tenants

ID	Theme	Question	Improvement Needed	Ad
G1	Governance	If you use Microsoft 365, do you have a committee or board to oversee the use, architecture and configuration of your Microsoft 365 tenant?	We don't have a committee or a board, but our Microsoft 365 tenant has a clear ownership model. Issues and requests are difficult to raise and/or KIM/information governance needs are not always addressed.	We have a committee make decisions around 365 tenant is used, how raising requests and it isn't clearly defined.
			It is not clear who is responsible for overseeing our Microsoft 365 tenant.	We don't have a committee Microsoft 365 tenant has model. Issues and requests the tenant owners and adequately addressed.
PR1	Provisioning	If you use Teams and/or SharePoint sites, do you have a process that controls the creation of new sites?	Any user can freely create new Teams and/or SharePoint sites without any process or configuration	We have a semi-autonomous includes a number of new and configure new Teams sites.
			We have provided staff with a pre-designed and configured structure of Teams and/or SharePoint sites. Typically, they are not able to request more	We have a largely managed users to request the or and/or SharePoint sites.
			We use Microsoft 365, but do not currently use Teams or SharePoint	
PR2	Provisioning	If you use Teams and/or SharePoint, how consistently is functional metadata applied to content?	We don't provide organisation-wide functional metadata in our Teams and/or SharePoint sites.	We provide organisational metadata for staff to use and/or SharePoint sites reliably applied to content.
			No, we ask staff to manually configure each new Team and/or Site before any records are managed.	

M365 Maturity Model Working Group

- ❑ The model based on universal principles of archive and record management community.
- ❑ Microsoft 365 is rapidly evolving.

IRMS established the Microsoft 365 Maturity Model working group with:

- The National Archives, UK (TNA)
- National Records of Scotland (NRS)
- Public Record Office Victoria (PROV)

The working group will:

- Determine when Model updates are needed based on new M365 features
- Conduct benchmarking to keep best practices current



- You

Kian

-

Questions?



Photo by [José Ramos](#) on
[Unsplash](#)