Consensus driven digital preservation

An organisational workflow

| **Jay Gattuso** |  |  |
| --- | --- | --- |
| *Author affiliation*  *Country*  *first.author@email.edu*  *ORCID* |  |  |
|  |  |  |

**Abstract – This paper offers an end-to-end workflow that indicates all activity steps, user roles, and significant milestones that can be followed to successfully deliver technical interventions to digital content - especially in an organisational context that has multiple stakeholders, interested parties, and overlapping policy items. .**

**Keywords – Workflow, Preservation Action, Process**

**Conference Topics – Resilience; Exchange**

# Organisational context

# The National Library of New Zealand (NLNZ) has been actively preserving digital content as a “business as usual” aspect of its operation since 2010[i]. In this time, we have learnt lots about how to ingest, characterise, assess and carefully store digital content in a digital preservation repository. We have started to work through how an organisation with a national heritage scope and mandate actively cares for digital content, not least starting to understand how we can identify risk, and what “care” really means for collection items.

A modern national library delivers many functions, most of which are not pertinent to this discussion. It is useful to briefly describe those structural entities at NLNZ that are directly relevant to the overall collection and care of digital content:

Preservation, Research and Consultancy – This is the digital preservation team, nine staff members including analysts, developers, technical specialists, and management roles. This team is responsible for the maintenance/development of the digital preservation system, and the delivery of digital preservation for the organisation.

Collection Development and Description – This is a portion of the published content processing team, approx. 20 colleagues including librarians, cataloguers, and technical specialists. This team is responsible for (amongst other things) the published collection scope, legal deposit acquisition and cataloguing functions.

Research Collections (digital) - This is a portion of the unpublished content processing team, approx. 10 colleagues including archivists, librarians and technical specialists. This team is responsible for (amongst other things) acquisition, technical processing and ingest of digital unpublished or special collections.

Curatorial Services - This is the curatorial service team, 12 colleagues responsible for the overall stewardship of all the collections in the care of NLNZ. This team covers all collection types, with a strong (and understandable) historical weighting towards traditional/physical media.

Conservation Services -This is the conservation team, 11 colleagues responsible for the planning, assessing and delivery of conservation for NLNZ. The team is predominantly focused on traditional / physical content.

For brevity we have not described any of the digitisation services, nor are we making a distinction between born digital, and turned digital content. For this model there is little technical difference between collection items of either type, and a central plank of our philosophy is that there must be a common treatment process for all content that requires preservation.

# Model Overview

To create a reusable and commonly understandable model, these organisational entities have been mapped on to a simple functional model. The functional model is used as the basis of the whole processing and decision-making workflow, it’s essential to make sure we have clearly defined these functions.

There are 3 main types of roles/functions in this process; “intellectual”; “technical”; and “manager”.

The manager class is self-explanatory. They are responsible for oversight, escalation, blocker resolution and sign-off.

The intellectual role is predominantly concerned with intellectualisation of the object, the majority of which is found in the curatorial activity, and other collection development activities.

The technical role is concerned with technical understanding of the digital object and is mainly held by the preservation team.

The archivist/librarian role sits somewhere in the middle of these two classes.

Another way of describing the difference between these two roles is that for any given digital content, the technical role is responsible for the accurate interpretation of the digital object - understanding the digital container/data/structures/standards, technical delivery and technical risk comprehension. The intellectual role is responsible for the accurate interpretation of the digital content and establishing its place/purpose in the collection[i].

The model proposed in this paper describes the formation of a project team responsible for delivering the successful treatment, the shape of that team, and the iterative steps required to understand, propose and mitigate any preservation risk. The model starts at a high level of problem assessment, thorough controlled steps to agreed detailed intervention. A key feature is the mechanism through which the project shares intellectual and technical understanding to identify the best outcomes for at risk content.

Specifically, the model is designed to prevent each role specialist from needing to operate outside of their specialism, guided by their professional code of ethics[ii], and with sufficient organisational buy-in for the work being attempted.

This approach is contradictory to contemporary archival practice, specifically the accepted archival / organisational processes that necessitate the archiving function being the least granular (as possible) view of the records in the collection. The broader organisation might point out that there isn’t enough resource to consider each record/file/item individually.

This is achieved through clear role boundaries, and a structured problem-solving process that is highly iterative, collaborative, agile, and supports incremental progress.

This paper seeks to address the need of the preservation function to have as granular as possible input about the purpose of a given file alongside the organisations need to work swiftly and efficiently through large records sets. This proposed blended approach must be given organisation support in the formative stages of establishing a sustainable and effective way of working.

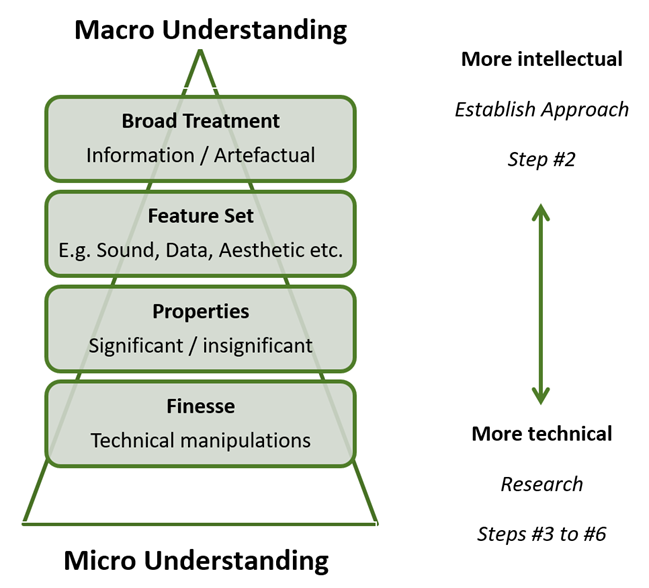


Figure 1 - The intellectual / technical spectrum

## Introducing The Model

Using the RASCI model[i], this paper proposes a process that will result in the successful completion of digital preservation treatments. Each process step is listed and followed by the RASCI roles for each step.

The process requires a key project team consisting of preservation, archival/librarian and curatorial expertise. Their respective manager roles are also required for support and escalation.

This process is built around the InSPECT framework**[ii]**, the NAA performance model**[iii]**, and our own experience at NLNZ.

The process itself has five phases:

· **Establish Approach** – where the project is initiated, and a direction of travel established.

· **Research** – where the technical and intellectual challenges are explored.

· **Finalise** – where the resulting treatment methodology is described.

· **Do** – where the treatment is applied to the collection items.

· **Ingest** – where the new items are integrated into the collection.

The following section is two version of the consensus driven model - the first is with the lens as the described roles at their simplest abstraction, the second is how those roles would be applied to the NLNZ context.

| Phase | Step # | Step | Roles | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Responsible | Accountable | Supportive | Consulted | Informed |
| Begin | **0** | **Risk flagged against object** | System, Technical | Manager System,  Manager Technical | Technical | - | - |
|  |
| Establish Approach | **1** | **Project start - risk assessment** | Technical | Manager Technical | Technical | - | Management All |
| *"Yes this is a genuine risk"*  *"Yes these objects require attention"* |
| **2** | **Treatment approach agreed** | Intellectual | Manager Intellectual | Technical | - | Management All |
| *"We need to preserve the spreadsheet"*  *"This object is essentially just a list of names and dates"* |
| Research\* | **3** | **Research options, make viable examples** | Technical | Manager Technical | - | - | Intellectual |
| *"We could use this tool or that tool"*  *"This script converts this format into something else"* |
| **4** | **A/B comparison of options / treatment products** | Technical | Manager Technical | Intellectual | Intellectual | - |
| *"These sound files seem clearer that those"*  *"I like how the paragraphs are positioned on the page"* |
| **5** | **Describe decisions required and issues with approaches** | Technical | Manager Technical | Intellectual | - | Curatorial |
| *"There are umlauts as macrons throughout the files, what do we do about that?"*  *"Do we need to respect the incorrect original page numbers? "* |
| **6** | **Make decisions on options and identified issues with migration method(s)** | Intellectual | Manager Curatorial | - | Technical | Management All |
| *"I would like to see the date codes appearing like option #4"*  *"I want you to use this particular font when there is no record of the original"* |
| Finalise | **7** | **Finalise treatment, prepare documentation** | Technical | Manager Technical | Intellectual | - | Intellectual |
| *"This is how we will configure this tool"*  *"These scripts will produce properly structured objects as agreed"* |
| **8** | **Confirm treatment**  *"I agree the examples I've seen achieve the criteria we set out"*  *"I would like you to proceed with option #2"* | Intellectual | Manager Intellectual | - | Technical | Management All |
|
| Do | **9** | **Complete treatment** | Technical | Manager Technical | - | - | Intellectual |
| **10** | **Quality/Sanity check** | Intellectual | Manager Intellectual | Technical |  | Management All |
| *"I am happy the new objects fit the criteria we agreed"*  *"I agree that these objects are of a good enough quality to go into the collection"* |
| Ingest | **11** | **Prepare changes for system** | Technical | Manager Technical | - | - | - |
| End | **12** | **System updated / Project closed** | System | Manager Technical  Manager System | - | - | Management All |

\* The research phase (steps #3 to #6) is highly iterative. This phase will involve large amounts of discussion amongst the project team. Fast failure is desirable. The project team should be encouraged to explore and test ideas within the allotted

| Phase | Step # | Step | Roles | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Responsible | Accountable | Supportive | Consulted | Informed |
| Begin | **0** | **Risk flagged against object** | System, Digital Preservation (Tech) | Manager Digital Preservation | Digital Preservation (Tech), Digital Archivist | - | - |
|  |
| Establish Approach | **1** | **Project start - risk assessment** | Digital Preservation (Tech) | Manager Digital Preservation | Digital Preservation (Policy) | - | Management Team - All |
| *"Yes this is a genuine risk"*  *"Yes these objects require attention"* |
| **2** | **Treatment approach agreed** | Curatorial | Manager Curatorial | Digital Preservation (Tech), Digital Archivist | - | Management Team - All |
| *"We need to preserve the spreadsheet"*  *"This object is essentially just a list of names and dates"* |
| Research\* | **3** | **Research options, make viable examples** | Digital Preservation (Tech) | Manager Digital Preservation | - | - | Digital Archivist, Curatorial |
| *"We could use this tool or that tool"*  *"This script converts this format into something else"* |
| **4** | **A/B comparison of options / treatment products** | Digital Preservation (Tech) | Manager Digital Preservation | Digital Preservation (Policy) | Digital Archivist, Curatorial | - |
| *"These sound files seem clearer that those"*  *"I like how the paragraphs are positioned on the page"* |
| **5** | **Describe decisions required and issues with approaches** | Digital Preservation (Tech) | Manager Digital Preservation | Digital Archivist | - | Curatorial |
| *"There are umlauts as macrons throughout the files, what do we do about that?"*  *"Do we need to respect the incorrect original page numbers? "* |
| **6** | **Make decisions on options and identified issues with migration method(s)** | Curatorial | Manager Curatorial | - | Digital Preservation (Tech), Digital Archivist | Management Team - All |
| *"I would like to see the date codes appearing like option #4"*  *"I want you to use this particular font when there is no record of the original"* |
| Finalise | **7** | **Finalise treatment, prepare documentation** | Digital Preservation (Tech) | Manager Digital Preservation | Digital Archivist, Digital Preservation (Policy) | - | Curatorial |
| *"This is how we will configure this tool"*  *"These scripts will produce properly structured objects as agreed"* |
| **8** | **Confirm treatment** | Curatorial | Manager Curatorial | - | Digital Preservation (Tech), Digital Archivist | Management Team - All |
| *"I agree the examples I've seen achieve the criteria we set out"*  *"I would like you to proceed with option #2"* |
| Do | **9** | **Complete treatment** | Digital Preservation (Tech) | Manager Digital Preservation | - | - | Digital Archivist, Curatorial |
| **10** | **Quality/Sanity check** | Curatorial | Manager Curatorial | Digital Preservation (Tech), Digital Archivist |  | Management Team - All |
| *"I am happy the new objects fit the criteria we agreed"*  *"I agree that these objects are of a good enough quality to go into the collection"* |
| Ingest | **11** | **Prepare changes for system** | Digital Preservation (Tech), Digital Preservation (Policy) | Manager Digital Preservation | - | - | - |
| End | **12** | **System updated / Project closed** | System | Manager Digital Preservation | - | - | Management Team - All |

\* The research phase (steps #3 to #6) is highly iterative. This phase will involve large amounts of discussion amongst the project team. Fast failure is desirable. The project team should be encouraged to explore and test ideas within the allotted

## The Model

## Step 0: Risk flagged against object

This paper assumes that a process has triggered an object to be flagged for attention, and plots the steps required thereafter for successful preservation treatment.

## Step 1: Project start - risk assessment

The first step is a validating mechanism that takes the risk flag (from wherever it is generated) and ensures that there is group buy-in to proceed with researching treatment.

The project team is formed, consisting of Digital Preservation Technical, Digital Preservation Policy, Digital Archivist, Curator roles and their supporting managers and agree to work on the content.

At this point additional content may be added if the project team agrees there is viable overlap of technical and intellectual concerns and characteristics.

For example, the process might have been initiated because some specifically formatted/structured files in a collection have a specific risk identified. The project team might be aware of other similarly formatted files in other collections that could be included in the treatment corpus.

Once the corpus is agreed, the corpus should remain closed to new files for the remainder of the project. This is to ensure the project scope can be agreed early in the process, and the success criteria are a consistent benchmark throughout.

## Step 2: Treatment approach agreed

Agreeing the treatment approach is crucial to delivering successful treatments.

Agreeing the treatment approach involves the intellectual role indicating the “intellectual” success criteria. That is, “what are the high-level requirements for preserving this content successfully?” It is essential that this step is owned by the intellectual role. Through being the “voice” of the collection, ultimately, they are responsible for the collections, and all the objects that compromise them.

Any work that results in new expressions of original objects must have a strong link to its original form, and the intellectual role must act as the gatekeeper, protecting the object from any unnecessary or inappropriate treatment.

This step is not a fulsome inquiry into the technical feature set of the objects. That occurs in the research phase (steps #3 to #6).

At minimum, there must be an indication of the broad treatment approach (information/artefactual – see fig. one) to reduce the likelihood of inappropriate treatments being attempted, and ultimately resulting in wasted effort.

This minimal information allows the technical role to focus its effort on preserving the information bound into a file or preserving the file itself[i].

It is useful at this time for the project team, led by the intellectual role to spend some time indicating the “purpose” of the object(s) in the collection where possible, i.e. why is this object in the collection. Whilst not mandatory, any time spent understanding and describing the broad features of the objects under treatment from a “performance” perspective is time well spent, helping both the preservation and archival roles to form a more complete understanding of the objects under treatment, and ultimately resulting in more suitable treatments and outcomes.

This “purpose in the collection” statement might be a somewhat complex proposition, so to try and illuminate the process we can work through a basic example.

When establishing a new collection, a shockwave flash file is ingested. During the ingest process, these files are flagged as being of a type that has some known difficulties in providing access. The shockwave flash file, when opened, displays some images to a soundtrack.

To be able to mitigate these risks, it’s essential the technical role understands the purpose of the file in the context of the collection.

Through discussion with the intellectual role, the minimum steer preservation required is the broad approach, that is, and simply put, is the importance of the object as an informational component that can be described such as sound, some text, an image, or is the object of importance the file itself. Making decisions at this level significantly reduces the risk of the technical role proposing a treatment that would not be considered successful later.

Having established the broad approach, it’s desirable to understand more about the given file. For simple formats this trivial, we would expect a sound file to remain a sound file.

In our example file we might agree that the shockwave flash file is informational (we don’t really need to maintain this object as a shockwave flash file) and the feature set we agree is important is a sound track and moving image component that are synced, and some controls to allow the person viewing the file to start/stop/fast forward/rewind/restart the content.

The technical role can then take these higher-level features and work towards developing a treatment process that ensures these important features persist over time.

The product of this step is some recorded notes of what a successful treatment should address, not at a technical level, but esoterically intellectually.

During this step, the project team should also be cognisant of the amount of effort indicated by the intended approach. It may be that one approach is more resource intensive than another and discounted for that reason when the impact of the treatment is considered.

This resource weighting is difficult, and all project members should have resourcing (time, effort, equipment etc.) in their minds when assessing the higher-level treatment criteria and agreeing potential success criteria. The technical role is likely to have the best estimation of required effort; the intellectual role is likely to have the best understanding of the relative importance of the objects in the collection.

## Step 3: Research options, make viable examples

Having agreed on a broad approach, the next phase is options analysis. The following four steps are highly iterative, and the project team should work in an agile-like way, with regular contact, to ensure that blockers are quickly addressed, and fast failure is allowed (both at this stage, fast failure is expected and encouraged at all stages). Through good group working practices the project team will maintain a sense of momentum, and more importantly, collective buy-in of the process and expected products.

This is the main “work” step in which the technical role explores the properties of the objects and works towards finding a viable future state for the objects.

This phase may also result in finessing work being required. Finessing work is loosely defined as technical work that falls out of the treatment steps as requested by the intellectual role but is not immediately obvious from the initial technical manipulation. For an example of this see the discussion on WordStar fonts in our 2014 iPres paper[ii].

In this step the technical role is responsible for translating the agreed approach into viable treatments.

## Step 4: A/B comparison of options / treatment products

The project team will collectively discuss / critique / explore the viable options. Ideally this is undertaken iteratively, with broad preservation concerns being addressed early, cascading into more refined solutions and treatment components. This is how the project team will work from macro understanding to micro understanding.

## Step 5: Describe decisions required and any issues with approaches

Previous experience in attempting this work suggests that as we work towards a viable treatment method, issues will be discovered, both new and previously noted. As the research work is undertaken by the technical role, it is essential that concerns that are not transparently resolvable by the technical role are answered or addressed by the intellectual role. This is another area where the ultimate responsibility of the collection resides at all times with the intellectual role[iii]. The technical role is required to provide the intellectual role with enough information to support them in making critical decisions about the future of the objects.

For examples of the sort of issues that need some direction see the MCW paper from 2015[iv].

## Step 6: Make decisions on options and identified issues with migration method(s)

The primary risk at this phase (and in fact for the entire process) is that decisions are not made as required to progress the treatment approach. The project team should be mindful of this risk and work swiftly and carefully to maintain project momentum.

Having described / documented any key decisions, the project team must work through any decisions and unresolved issues to find potential answers. The technical role (per the previous step) is largely responsible for identifying and describing the issues. The intellectual role is responsible for the objects and is the final decision maker for any issues.

This highly iterative research phase is complete when there are no more options or issues that require attention.

If the intellectual role is unable to decide on a preferred option, the project team either focuses on the existing options looking for refined resolution or returns to steps #3/#4/#5. The intellectual and technical roles should work closely and carefully to ensure the issues are well understood and key decisions are clearly described and documented.

If the project team is unable to progress the options, the project team will escalate the concerns to the manager roles. The manager roles ensure the intellectual and technical roles progress the project through to successful resolution – perhaps by mediation disagreement, offering further steps or options, or at worst case pausing/ending the project if no agreement for options can be found.

## Step 7: Finalise treatment, prepare documentation

Having resolved any issues, and agreed on the general approach to the treatment, the technical role completes any treatment plan required by the project team. This should clearly describe the treatment processes, including any key issues/resolving decisions (per step #6).

## Step 8: Confirm Treatment

The project team must agree that whilst the technical issues/options may reside with the technical role, the intellectual role must be responsible for the final decision and confirming the treatment.

The outcome of this stage is an agreed treatment plan (from step #7). This will describe the final product and will form some basis for the final checks at project close.

## Step 9: Complete Treatment

The technical role is now responsible for completing the treatment plan as agreed with the project team. Regular stand-ups may still be needed if the treatment is complex, however in the main the technical work should be well described by this time.

It may be that new issues are found at this stage, and the project team must be comfortable returning to the options assessment phase if there is no remedy found.

## Step 10: Quality/Sanity check

Once the treatment plan has completed, the intellectual role uses the agreed success criteria to confirm that the treatment has resulted in successful outcomes, and any newly created objects are “sufficient surrogates” for the original objects.

## Step 11: Prepare changes for system

Having passed the quality checks, the new objects are packaged for ingest back into the system.

## Step 12: System updated / Project closed

Once completed, the project is closed

REFERENCES

[i] Gattuso, J. 2017, Preservation actions: Where we started and, where do we go from here? <https://natlib.govt.nz/files/digital-preservation/Gattuso-july-2017-preservation-actions-pre-print.pdf> - Last accessed Dec 2021

[i]<https://en.wikipedia.org/wiki/Responsibility_assignment_matrix#RASCI> - last accessed Dec 2021

[ii] [https://web.archive.org/web/20141017095627/http://www.significantproperties.org.uk:80/methodology.html](https://web.archive.org/web/20141017095627/http:/www.significantproperties.org.uk:80/methodology.html) (2009) - last accessed Dec 2021

[iii]<https://www.ingentaconnect.com/contentone/ist/ac/2005/00002005/00000001/art00006?crawler=true> (2005) - last accessed Dec 2021

[i] You want to preserve what? Webb et Webb, Colin, David Pearson and Paul Koerbin. “'Oh, you wanted us to preserve that?!' Statements of Preservation Intent for the National Library of Australia's Digital Collections.” *D Lib Mag.* 19 (2013)

[ii] McKinney, P., 2018. [A Draft Code of Ethics for Digital Preservation (pdf, 745KB)](https://natlib.govt.nz/files/digital-preservation/Draft-Code-of-Ethics-for-Digital-Preservation.pdf)

[i]<https://digitalpreservation.natlib.govt.nz/assets/NDHA/Publications/2015-/Gattuso-july-2017-preservation-actions-pre-print.pdf> (2017) - last accessed Dec 2021

[ii]<https://digitalpreservation.natlib.govt.nz/assets/NDHA/Publications/2014/WordStar-ipres2014-4.pdf> (2014) - last accessed Dec 2021

[iii] Curatorial policy?

[iv]Internal NLNZ document - “.MCW potential migration: Initial position on early questions”<https://dia.cohesion.net.nz/Sites/CMD/COPR/_layouts/15/DocIdRedir.aspx?ID=KUWHQ4YV2Q5A-1162214148-569> (2016) - last accessed Dec 2021