**Pre-submitted Questions**

#1: GRAVES PAPERS CURATION (Librarian-1’s info)

Why does the Financial Cataloguing Form (FGA 049) have less information in it than the Section IV Financial Matters spreadsheet?

#2: GRAVES PAPERS CATALOGUING (Librarian-2’s info)

What Quality Assurance processes are used to test MARC and MARC XML records created from the Section IV Financial Matters spreadsheet?

What Quality Control processes are used to ensure end-to-end quality of tasks, documentation and data (e.g. audit processes)?

#3: OAIS REFERENCE MODEL

Does the OASI Model provide enough detail to guide digital curation actives in WIT (creation) and for DRI (submission)?

How does the Model assist you as a member of the multi-disciplinary team engaged in designing an archival system for the Cooley Collection?

On a scale of 1-5, how useful is the OAIS Model in your role (1: very poor / 5: very well)

<https://libguides.bodleian.ox.ac.uk/digitalpreservation/oaismodel>

#4: DCC DIGITAL CURATION LIFECYCLE MODEL

Does the Lifecycle Model provide enough detail to guide digital curation actives in WIT (creation) and for DRI (submission)?

How does the Model assist you as a member of the multi-disciplinary team engaged in designing an archival system for the Cooley Collection?

On a scale of 1-5, how useful is the DCC Model in your role (1: very poor / 5: very well)

<https://libguides.bodleian.ox.ac.uk/digitalpreservation/digitalcurationlifecycle>

**=====================**

**Interviewee** Librarian-2

**Attendees** Researcher-1, Researcher-2 and Researcher-3

**Date** 16-Apr-2021

**Note(s)** Recording not preferred. No digital recording taken (video or transcript)

Notes written by Researcher-3

**Intro**

I was doing this curator role, but am not doing it now… primarily working as a Librarian now

Not being doing digital curation on a regular basis

Interview will do the best for the QA/QC questions

Some of the questions I cannot answer (unknown)

The Cooley Collection is potentially ground-breaking for WIT

Work of Researcher-1 and Researcher-3 has forced the Librarian’s to look at their processes and practices they use and look at putting together more robust practices

MARC is very much a library channel to allow mapping of fields to systems

MARC is not 100% sufficient to meet the demands of the data / digital object

When writing this up this is the feedback from the Deputy Librarian

**Question 1: Graves papers curation**

Financial Cataloguing Form (FGA 049) has less info than the Section IV Financial Matters spreadsheet

Some of the info on the spreadsheet might be auto generated vs. manually created

You don’t always have to catalogue everything to the nth degree

Papers can be part of a wider schema, a generic template to ‘catch all’, as the other staff in the library were doing this cataloguing [not Librarian-1 and/or Librarian-2]

Each artefact associated to a page within the template

Original info from the archive plus external information additional sources added to the excel

The WORD doc is the info from the resource and the EXCEL is the culmination of all info available

EXCEL used to provide the richest amount of data available

TEMPLATES help show mandatory info (based on bibliographic cataloguing) but it has other info also

Prompting curators to capture mandatory, have

Resources aren’t always curated to the nth degree because the value of the record and the priority of the object is prioritised based on interest in the topic/item

This explains why the scale and level of cataloguing of the record differs between items &/or collections

Flaws in records is about lack of consistency and wide variations of data – that’s the QC and QA

Before cataloguing need to look at the level of detail wanted to be captured and level of the granularity determined for the resource/collect

The value placed on the item/collection is based on the tacit knowledge of the curator

* Preservation and conservation to ensure archive is futureproofed / manage that data long-term
* User focus, usability and access to the resources (usability and accessibility)
* HARD TO DETERMIN what is of value now vs. in the future – use professional best guess
* CONSISTENCY is the foundation of QC and QA in cataloguing
  + If you don’t have processes/polices that causes problems with accessibility or completed the project and/or crate records that are not fit for purpose

Augment the info by adding extra to the Excel

The Excel is also a QC element because it supports the formatting of Data

**Question 2: QA processes to test MARC**

Use MARC Edit for converting excel / tabular data into MARC XLS files

As a tool it allows librarians to QC the data based on fields linking

If there is a flaw in how data is structured, the problem is flagged

**TOOLS USED**

Open Refine

Text pad also used to search/find/replace to help find inconsistencies in the data

Use artificial constructs in how to catalogue records so that a Title Entry is consistent across all artefacts of a certain type

Users/Researchers can use the construct to give them consistency that the Addressee first, Sender second, Date third on the list, etc.

Also supports the format of the date…. Do you keep long-form date, what happens if short date is referenced elsewhere in the letter and or in other records (copies) on other sites?

Looking to keep a standardised format for consistency

Up until now WIT has been using MARC

DRI will trigger a move to DUBLIN CORE

**MAGDALENE**

Looking to bring the records from EXCEL converted to MARC to migrate them to DUBLIN CORE

Triggering them to think about how to fit square pegs into round holes

If starting with the best system for archives they wouldn’t start with MARC

The links from the Oxford Library are where WIT would like to be, not there yet

How do we catalogue a non-digital item currently?

Use Anglo-American Cataloguing rules

N. America / Canada/ England agreement about how bibliographic records are collated

Now in Google key-search word, how tech lets us find info

Tech has led to a large amount of recall but low-level precision

The AAC rule leads to a high level of precision and low level of recall

Many archives are sat in boxes in archives, not on the web and not known about/accessible

Multiple copies of the same book, have different bar-cods have

Digital book and physical book catalogued separately then connected by series, author or URL

Catch-all field of notes allows all info to be captured to put all notes into that single field

**ANSWERS TO PREVIOUS QUESTIONS**

What intangible knowledge is lost during cataloguing?

Experience and level of detail of the curator is key to the richness

Templates help to give cataloguers milestones and consistency to record creation

A lot of these skills have to be learnt, not book learnt, and is built from practice

There is no concrete process/approach as each person and institution is different

The mediation of the prob is the creation of policies and processes in institutions

QC curation work/metadata authoring

Producers create records

Peer review records by senior curator

Authors need to meet the conventions of the cataloguing requirements

Physical label added to the book to give the unique identification (adds to the bar code of the item)

Photos

MARC and DUBLIN CORE allow for the specifics of the cultural heritage of the item

Document the format of the object

Information can be harvested afterwards

If this is not done at the start, the user cannot later understand or find the record

Part of the QC / QA process

What is the QC / QA processes?

No universal standard

Down to each individual institution

More mature institutions have the resources to create processes

Smaller organisations don’t have the staff or budgets

If working in a university Cooley Collection wouldn’t be available until all of the policies, processes for the collection management had been defined

That is best practice, but without the resources to do that, the archive would be locked away and unusable

Too ridged a QC processes could prevent the archive being made available i.e. there could be sensitive material in the archive and / or risk of loss of material if an item is taken (no record of it existing)

Legal restrictions on access e.g. CSO have 100 years limit on accessing data

Red Kettle donation has a 100-year embargo on financial docs in their donation to WIT library

Tacit knowledge is about knowing how to manage access in the absence of specific policies/process

QA / QC process helps ensure the security of the archives

**Question 3 – OAIS and 4 – DCC models**

Many of the digital models in existence… Also have the 3-legged stool model

All models have the same goal of preserving the record and its information

WIT are not using standards in the same way as OAIS and/or DCC Models

It is a serious flaw in their process

WIT are doing be-spoke digital preservation versus adhering to any digital standard

Local storage, externa drive (USB) and Cloud

One of the key tenets of digital preservation has secure and allows access from a futureproof perspective

The Digital Object Identifier (DOI) is a key element to the digital objects and preservation

The DOI gives records their unique ID to allow researchers to include that information in their research based on specific items

Waterford institute of Technology (WIT) existing structures don’t offer that structure (e.g. DOI capture)

Digital Repository of Ireland (DRI): does have structures and has got WIT thinking about these requirements

Digital Repository of Ireland (DRI) governing:

Structure of objects

Format of objects (e.g. TIFF files)

WIT will have to reconvert JPEG to TIFF to match the preservation strategy to DRI…. The migration plan of WIT to futureproof their own work

BASED ON WHAT I HAVE SEEN / KNOW TO BE FACT (considerations for Researcher-3 to think about)

What model would I recommend based on what I see now?

Point in time where WIT is diverging from a linear path into a more digital space

Knowing the resources available in WIT, which model is something that can be used practically?

Current work practices, much better recording of data than the preservation of data

Pathway that I am at now (Researcher-3)

Evaluation of current practices

Look at what options are available, what would I advocate as a process (QA and QC)

These are the standards available

These are the things WIT are doing already

This is where there are gaps in the process

This is what I would advocate for future work

**EUROPEANA**

Open a call with DRI with

Convert the MARC record into an XML file for ingestion into DRI repository

Doesn’t go live immediately

Top Level is most important decision for WIT to define e.g. Collections: Cooley, Magdalene

Once things are uploaded, they can be reviewed, edited/checked before going live

Many different stakeholders involved in the creation of archives.

DRI is highly structed which has huge benefits

Once ingested into DRI – and published - EURPEANA will be able to harvest the data for their searches

KEY Considerations

Copyright

Digital realm has a whole lot more considerations but also offers a lot more control

Moving records to the DRI helps with futureproofing (Dublin Core + XML format) the records are interoperable

If WIT leaves DRI at any time, the records can be downloaded and placed elsewhere

EUROPEANA, structed entries allows users to see all entries versus the local silo (aggregated information)

Bad preservation practice is taking down resources after they have been published e.g. cite link break

**SUMMATION**

Librarian-2: No problem clarifying questions raised in this session. Draw a line in the sand for new questions… doesn’t want to promise time and commit to time that he cannot give.

Without the expertise of the Lab and the students there wouldn’t be the knowledge/skills in any one person

The INSYTE-COOLEY lab is a Research and Learning tool

**SIDE NOTES**

Researcher-1: DRI is the connecting step to get WIT on track

Librarian-2: they are swamped with trying to support students accessing the library… busier than ever

Researcher-3: Confirming that something is unknown is as useful as knowing what is known and done [e.g. QA/QC processes, archive models]

Being the institution that follows other institutions in digital archiving is not a bad thing, can learn from their experiences and mistakes and create a better archive, first time around, for the Cooley Collection