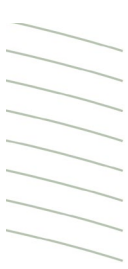# Technology, Metadata and Mannix Library Special Collections: Descriptive Tools and Considerations

by Huw Sandaver

Catholic Theological College, University of Divinity

This paper discusses the varying metadata and technical standards used with Mannix Library's special collections, specifically descriptive standards, the MARC standard, the Dublin Core Metadata Element Set in use with the University of Divinity's CONTENTdm repository and tools such as Mirador, Omeka and the services they interact with. The paper merges and expands on a virtual conference presentation for ANZTLA Virtual Mini-Conference 2020 and several posts for a pre-conference cataloguing workshop.

The polysemic nature of the historical collections in libraries has presented unique opportunities. Library special collections present a unique opportunity. Considerable work has been undertaken over the last several years to evaluate Mannix Library's rare books and special collections, which consists of roughly ten thousand items spread over several different categories. The criteria around what is considered special is fairly fluid, however, most importantly, age, scarcity and provenance is taken into consideration.

The Rare Books and Manuscripts Section of the Association of College and Research Libraries mentions several professional competencies[1] around metadata and technologies, the most relevant being that librarians working with special collections should:

- Maintain awareness of issues, standards, trends, and current best practices regarding the full range of information technologies
- Understand the use of methodologies, including digital imaging and recording technologies, for creating reproductions of primary source materials
- Be able to engage and collaborate with scholars to develop innovative teaching strategies and scholarly tools utilizing information technologies and special collections materials
- Be familiar with the use of digital asset management systems and metadata for providing access to digitized primary source materials
- Develop and maintain knowledge of standards, rules, best practices, and tools used for organizing and describing special collections' materials in library catalogues, archival finding aids, databases, and Web sites
- Understand integrated library systems, bibliographic utilities, and federated, Web-based platforms for sharing information about collections
- Develop and maintain knowledge of descriptive practices in related fields, such as archival and museum communities, and appropriately adopt and apply such practices to the description of special collections materials.

Information Technologies

The most obvious information technology librarians working with special collections interact with is their library management system. Mannix Library uses WorldShare Management Services from OCLC. The metadata module within this system enforces validation of MARC before a record can be saved, therefore inherently adhering to best practices for metadata creation. Information technology changes rapidly and regular reading of journals such as *Information Technology and Libraries, Library Technology Reports* and *Code4Lib Journal* allows for current insight into best practices regarding library information technologies.

At Mannix Library various IT platforms and services are in place to assist with special collections:

Ubuntu Linux Server

Mannix hosts its own website (https://mannix.org.au) on a cloud-based Ubuntu (https://ubuntu.com) Linux server. This allows for not only the hosting of websites, but also to create

---

[1] 'RBMS Competencies for Special Collections Professionals: Approved by the ACRL Board of Directors, July 2008'.
College & Research Libraries News 69, no. 10 (2008): 622–29. https://doi.org/10.5860/crln.69.10.8083.

scripts that specialised tools including ImageMagick (https://imagemagick.org/index.php), an IIIF image server (https://iipimage.sourceforge.io/documentation/server/), Omeka (https://omeka.org/), Pandoc (https://pandoc.org), Tesseract (https://github.com/tesseract-ocr/tesseract), MediaWiki (https://www.mediawiki.org/wiki/MediaWiki).

OAI-PMH though Omeka and CONTENTdm

The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) (https://www.openarchives.org/pmh/) allows interoperability between different repositories, effectively allowing metadata from one system to display in another. Currently the Omeka site for Mannix Library's Archbishop Goold Special Collection (https://gooldlibrary.omeka.net/) is being harvested by the Atla Digital Library (https://dl.atla.com/collections/university-of-divinity) and the OCLC hosted CONTENTdm (https://divinity.contentdm.oclc.org/) is currently being harvested by WorldCat (https://www.worldcat.org/). JSTOR also harvests the material (https://www.jstor.org/site/university-of-divinity). The protocol allows for greater engagement with collections by capturing the collection and displaying it in a larger aggregated database, thereby minimising how many different systems a scholar has to search through. The harvesting systems may also provide greater functionality than the host platform. In the case of the Atla Digital Library and WorldCat both of the systems are enabling shareability of the collection through social media and other services that the native platforms aren't providing for. Both WorldCat and Atla Digital Library offer a citation in different formats not provided for on the host platforms. Choice of platform that includes the protocol is therefore a high value and low maintenance way of increasing usage and enhancing functionality.
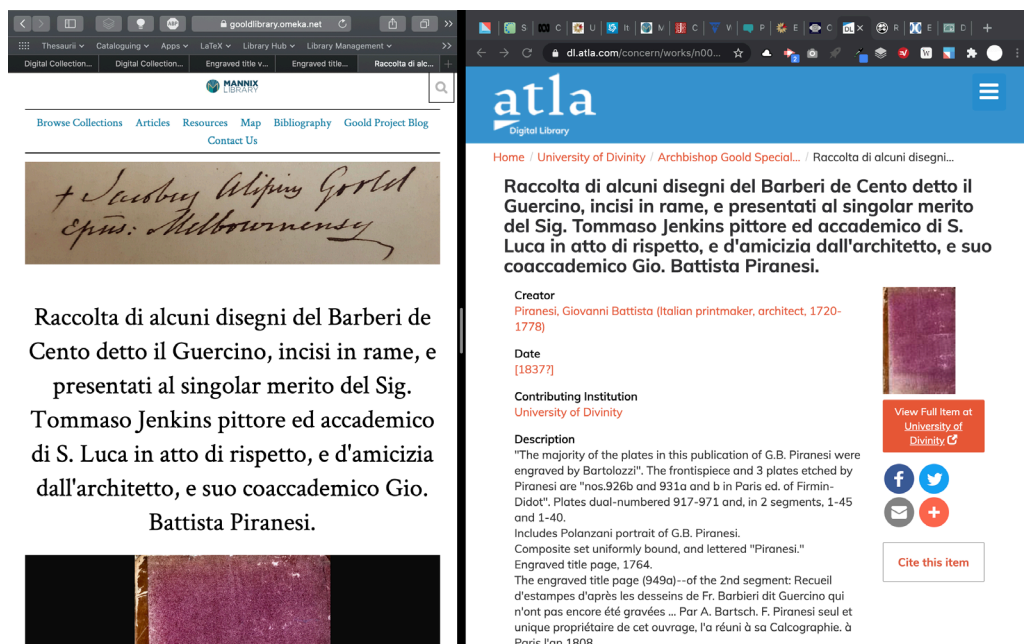

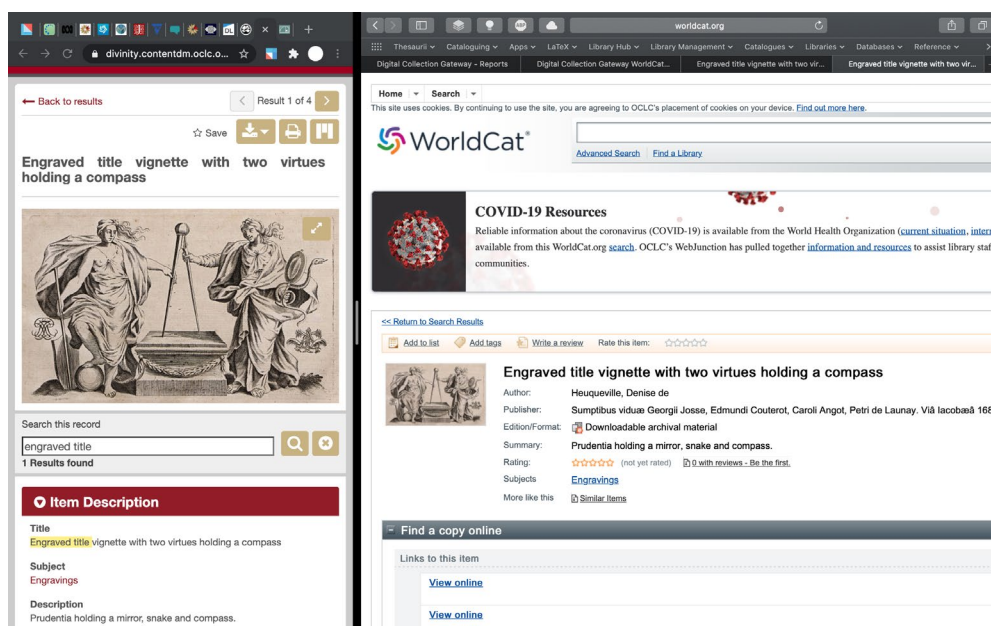Figure 1. Omeka and Atla Digital Library showing the same item through OAI-PMH.

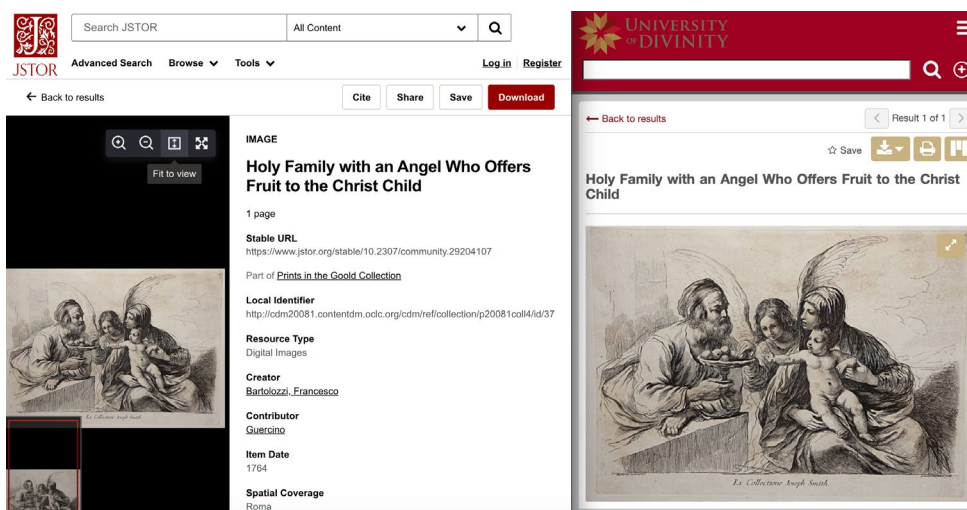Figure 2. CONTENTdm and WorldCat showing the same item through OAI-PMH

Figure 3. JSTOR and CONTENTdm showing the same item through OAI-PMH

Application Programming Interfaces (APIs)

Various APIs[2] are useful for both the general collections at Mannix Library but because of the ability for an API to display content from one system in another they are especially useful for special collections, where material is usually hosted externally, or utilising an existing system to display information in a specific manner.

[2] Application Programming Interface https://www.wikiwand.com/en/API

WorldCat

In order to construct search widgets for WorldCat discovery, part of the WorldCat REST[3] API architecture (https://developer.api.oclc.org/wcv2) can be utilised. While Mannix Library uses this architecture for the whole collection, we can further refine this service for use with special collections. The special collection for ISCAST Network (http://iscast.org/library), for example, has a specific utilisation of this architecture which points only to their physical collection. In this case instead of requesting JSON[4] from the server, HTML is called from the server and displayed in the native interface.

SUSHI

The SUSHI COUNTER[5] API (https://app.swaggerhub.com/apis/COUNTER/counter-sushi_5_0_api/1.0.0) uses REST to allow statistical data about usage of digital collections to be displayed as a structured dataset from the raw JSON response. This allows for an automated workflow through the Wolfram|One platform, rather than manually dealing with CSV or Excel Spreadsheets. Programmatic calls to the server can be made and then manipulated into a dataset and then subsequently a visualisation. While this service is generally used for born digital items, such as electronic journals and eBooks, if a collection is being harvested to JSTOR for example, the usage will show as COUNTER compliant statistical data, which is able to be retrieved through an API call.

```
sushi11 =
  URLExecute[
   "https://www.jstor.org/sushi/reports/pr?requestor_id=
          █████████r&customer_id=█████████u&begin_date=
      2019-01-01&end_date=2019-12-31&Total_Item_Requests
      ", "RawJSON"];
```

Figure 4. The SUSHI API Call asking for a response in Raw JSON

---

[3] Representational State Transfer https://www.wikiwand.com/en/Representational_state_transfer
[4] JavaScript Object Notation https://www.wikiwand.com/en/JSON
[5] COUNTER is a standard that allows for consistent usage report data across different platforms
https://www.projectcounter.org/

Platform	Item_ID		Database	Publisher	Performance		
	Type	Value			Period		Instan
							Metric_Type
EBSCOhost	Proprietary	EBSCOhost:edsacl	ACLS Humanities E–Book	American Council of Learned Societies	Begin_Date 2020–08–01		Total_Item_Investigations
					End_Date 2020–08–31		
					8 total ‣		
EBSCOhost	Proprietary	EBSCOhost:edspdh	APA PsycArticles	American Psychological Association	Begin_Date 2020–07–01		Total_Item_Investigations
					End_Date 2020–07–31		
					4 total ‣		
EBSCOhost	Proprietary	EBSCOhost:asx	Academic Search Index	EBSCO Publishing	Begin_Date 2020–08–01		Total_Item_Investigations
					End_Date 2020–08–31		
					8 total ‣		
EBSCOhost	Proprietary	EBSCOhost:h7h	Atla Historical Monographs Collection: Series 1	American Theological Library Association	Begin_Date 2020–08–01		Total_Item_Investigations
					End_Date 2020–08–31		
					8 total ‣		
EBSCOhost	Proprietary	EBSCOhost:h8h	Atla Historical Monographs Collection: Series 2	American Theological Library Association	Begin_Date 2020–08–01		Total_Item_Investigations
					End_Date 2020–08–31		
					8 total ‣		
EBSCOhost	Proprietary	EBSCOhost:lsdar	Atla Religion Database with AtlaSerials PLUS	Undefined	Begin_Date 2020–08–01		Total_Item_Investigations
					End_Date 2020–08–31		
					8 total ‣		
EBSCOhost	Proprietary	EBSCOhost:lsdah	AtlaSerials PLUS, Religion Collection	Undefined	Begin_Date 2020–08–01		Total_Item_Investigations
					End_Date 2020–08–31		
					7 total ‣		
EBSCOhost	Proprietary	EBSCOhost:edsaup	Australian Public Affairs – Full Text	RMIT Training Pty Ltd trading as RMIT Publishing	Begin_Date 2020–08–01		Total_Item_Investigations
					End_Date 2020–08–31		
					3 total ‣		
EBSCOhost	Proprietary	EBSCOhost:edsjbk	Books at JSTOR	JSTOR	Begin_Date 2020–08–01		Total_Item_Investigations
					End_Date 2020–08–31		
					8 total ‣		
EBSCOhost	Proprietary	EBSCOhost:edsble	British Library EThOS	British Library Board	Begin_Date 2020–08–01		Total_Item_Investigations
					End_Date 2020–08–31		
					8 total ‣		

rows 1–10 of **52**

Figure 5. JSON Converted to a Dataset by Wolfram One

IIIF Presentation API and JsonStorage API

The IIIF[6] Presentation API (https://iiif.io/api/presentation/2.0/) is heavily used with CONTENTdm. The API itself makes use of Linked Data as manifests in JSON-LD[7] format. Items uploaded and described on CONTENTdm are displayed using an IIIF server. Typically, this is used to bring a high-resolution image viewer into the native catalogue interface (WorldCat Discovery[8]).

---

[6] International Image Interoperability Framework (https://iiif.io/)
[7] JavaScript Object Notation for Linked Data (https://www.wikiwand.com/en/JSON-LD)
[8] The example shown here is accessible from the following link ad clicking the diagonal viewer arrows on the thumbnail image https://divinity.on.worldcat.org/search?databaseList=&queryString=no%3A1142635892#/oclc/1142635892
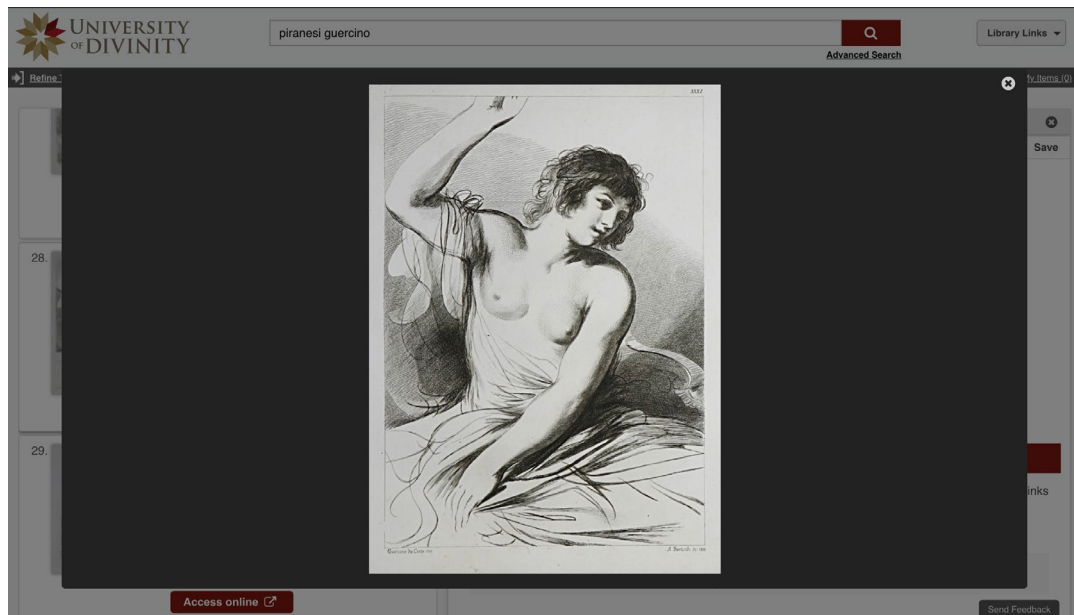
Figure 6. WorldCat Discovery displaying an image using the IIIF Server

The API is also used to display the alternate viewer Mirador (https://projectmirador.org), which is an added integration for CONTENTdm. Mirador can be used both within CONTENTdm and also externally, using the API to use the data output by CONTENTdm. This allows extensive use of the viewer in various applications including learning management systems such as Moodle (https://moodle.org).
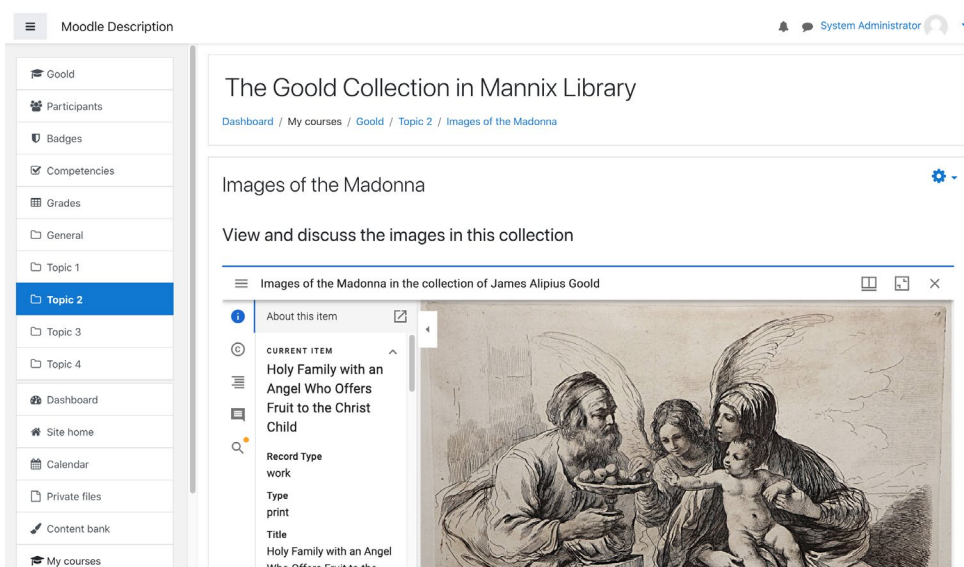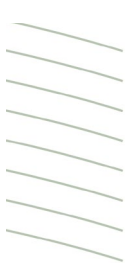

Figure 7. An example of Mirador embedded into Moodle

The viewer can be styled and has a gallery function, making it especially good for a virtual exhibition style web site. Crucially, this element engages with the professional competency around "developing innovative teaching strategies and scholarly tools utilizing information technologies and special collections materials".
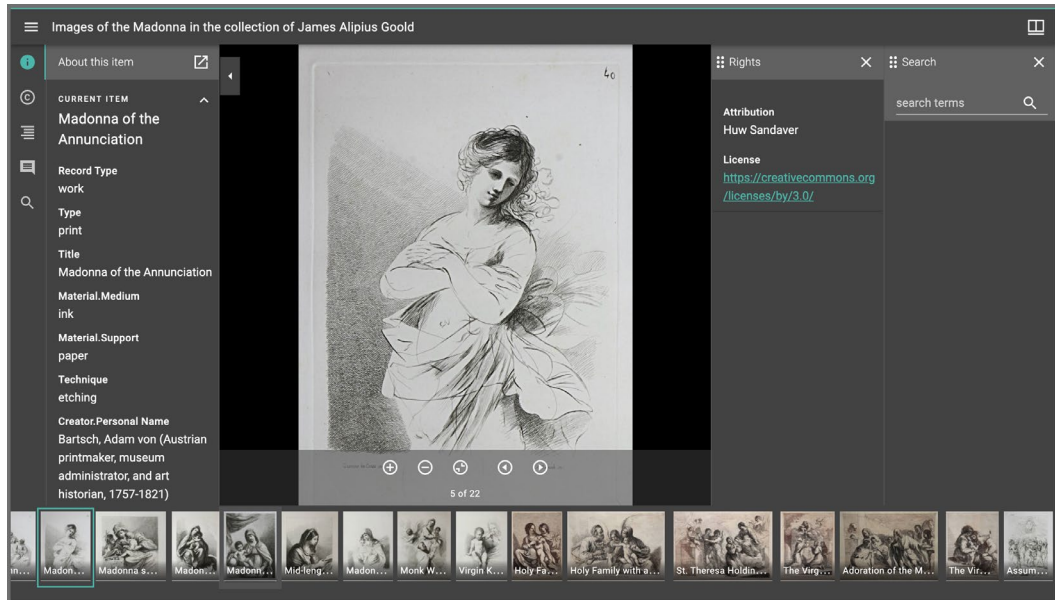
Figure 8. Mirador displaying a manifest

Further to this with an annotation server installed, librarians and academics can mark up images and texts with explanatory information and transcription. The presentation API allows a further search API to be embedded into the manifest. Annotations allow hyperlinks, images, audio and other HTML elements to be embedded into the information panel, which can greatly improve the user understanding of the elements of the collection.
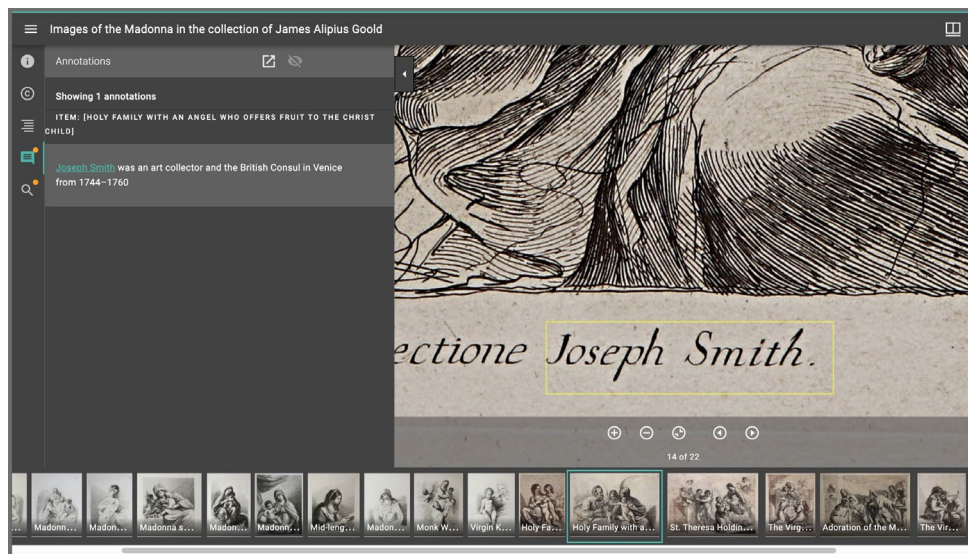
Figure 9. Mirador displaying an annotation
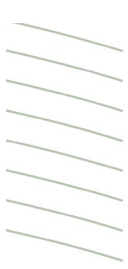
Manifests can be created and edited in an online application created by Oxford University https://digital.bodleian.ox.ac.uk/manifest-editor/ however, for full functionality the JSON-LD file must be directly edited and stored as an URI[9]. JsonStorage.net can be used to create an URI for the manifest.
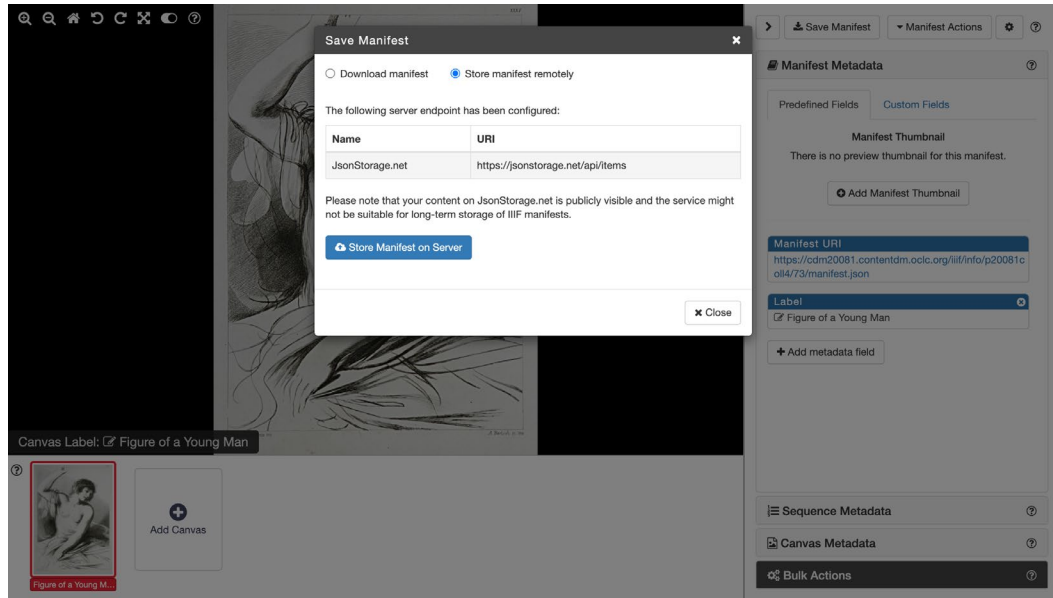
Figure 10. Visual Manifest Editor used to store a manifest as a URI on JsonStorage.net

Free API development software such as Postman (https://www.postman.com/) assists with this, as it allows retrieval and manipulation of the JSON-LD file directly from the storage server.

---

[9] Uniform Resource Identifier https://www.wikiwand.com/en/Uniform_Resource_Identifier
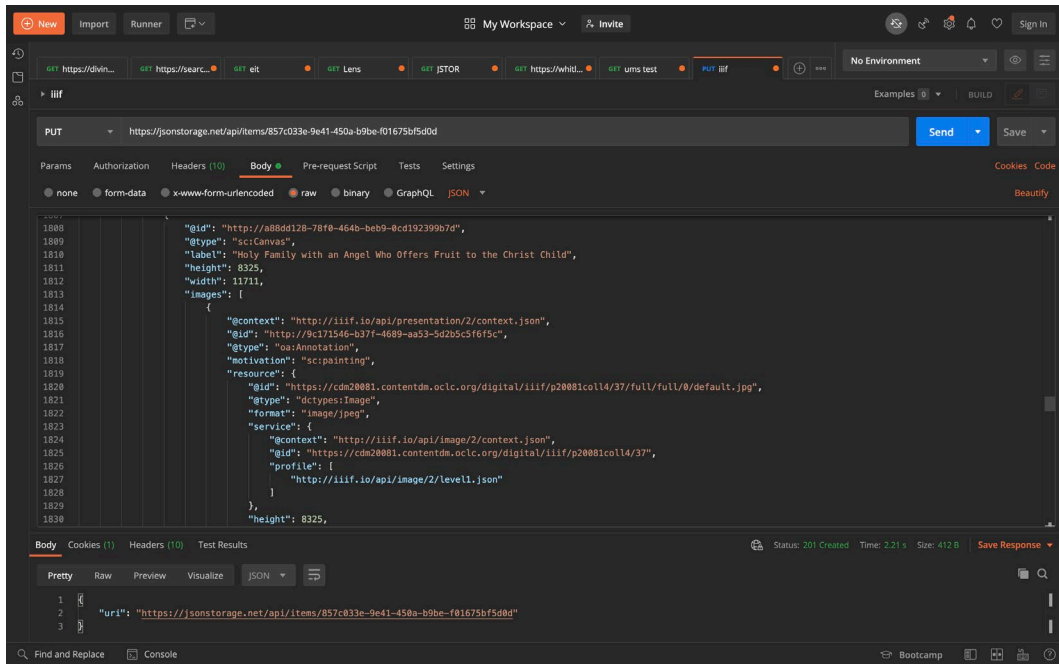
Figure 11. Postman using the JsonStorage API to manipulate data

Once the custom manifest is created, various manipulations of the material can be made, such as alterations in brightness, inverting colours. This can have benefits to the researcher since these interventions often highlight underlying imagery not always seen in the normal scan, such as watermarks.
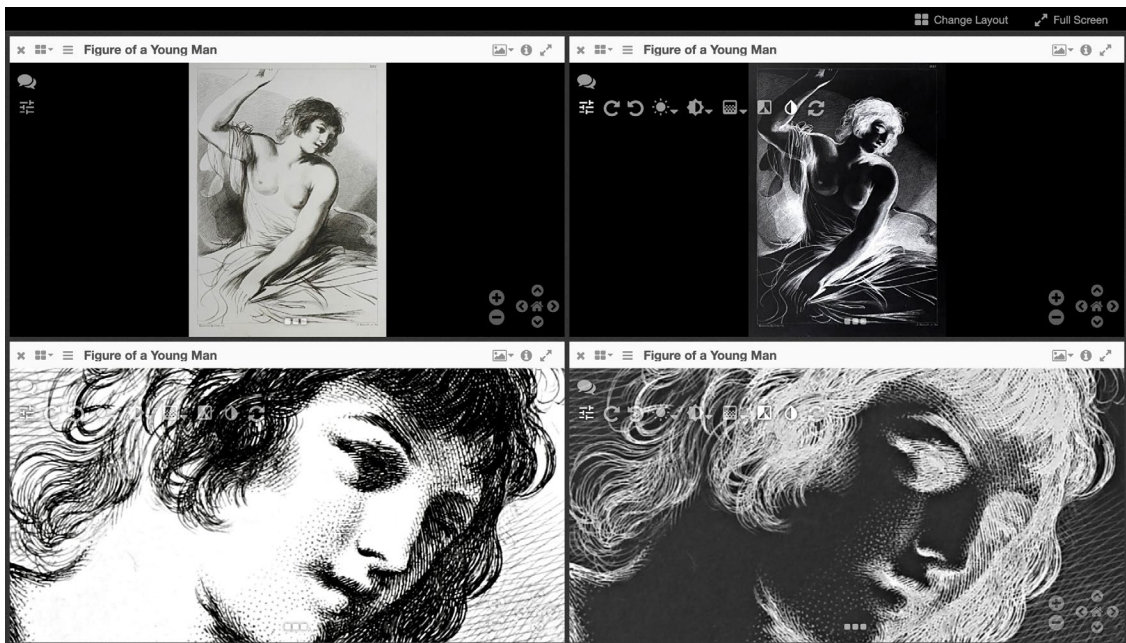
Figure 12. Mirador proxied by CONTENTdm with the image manipulation toggle
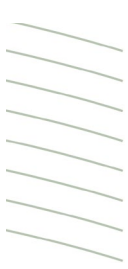
Wolfram|One/Cloud additional uses for Special Collections

While any kind of API can be utilised from within Wolfram|One, the entire platform has functions of specific benefit for special collections. Conservation is one area where Wolfram can help visualise how well the environment is suited to where the collection is being stored. With a simple electronic device, environmental data can be recorded and compared against the actual trend. The goal is to keep the conditions from variating over five degrees warmer or cooler. Wolfram's visualisations help to show any warning trigger points for checking the status of more fragile bindings that might be vulnerable to such change.
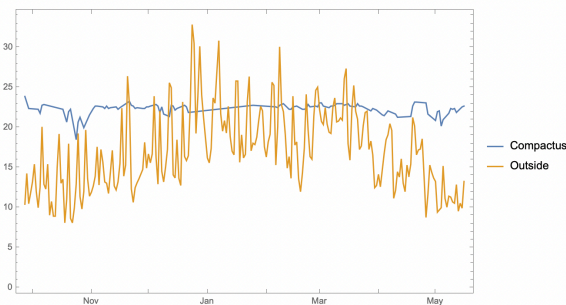

Figure 13. Conservation Environmental Data using Wolfram|One

It is also possible to analyse images in order to achieve a greater understanding of the collection. For example, with chromolithographs the chromaticity (a chart displaying the properties of the colour used in an image) can be examined. In the case presented here we can compare two artists, one known to be part of the Barbizon school, a realist movement during the mid-to-late 19th century. The examination of the images shows that realist artist does indeed have a narrower and more naturalistic use of colours in the lithographs (represented by the top two graphs in the set). It is also possible to examine an image in terms of spread of ink or colours in an image. The sanguine toned prints digitised as part of the Goold collection in the Mannix Library display three colour tones – Wolfram can break the image down to show the distribution of inks. It is also possible to leverage machine learning to try and understand representations of emotion in an image. The resulting processing shows the strongest face in the central
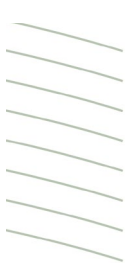
part of the image, the faces are designed to promote an empathetic response in the viewer and that if edges are detected in the image, we see a conversion in the central part of the image promoting a sense of balance. These are all specific traits of a Renaissance style image (something that the engraver was emulating). In fact, of all the images in the folio, the image presented in the examination has the most defined sense of balance and empathy which most definitely explains why many people are drawn to it as one of the strongest images in the collection.
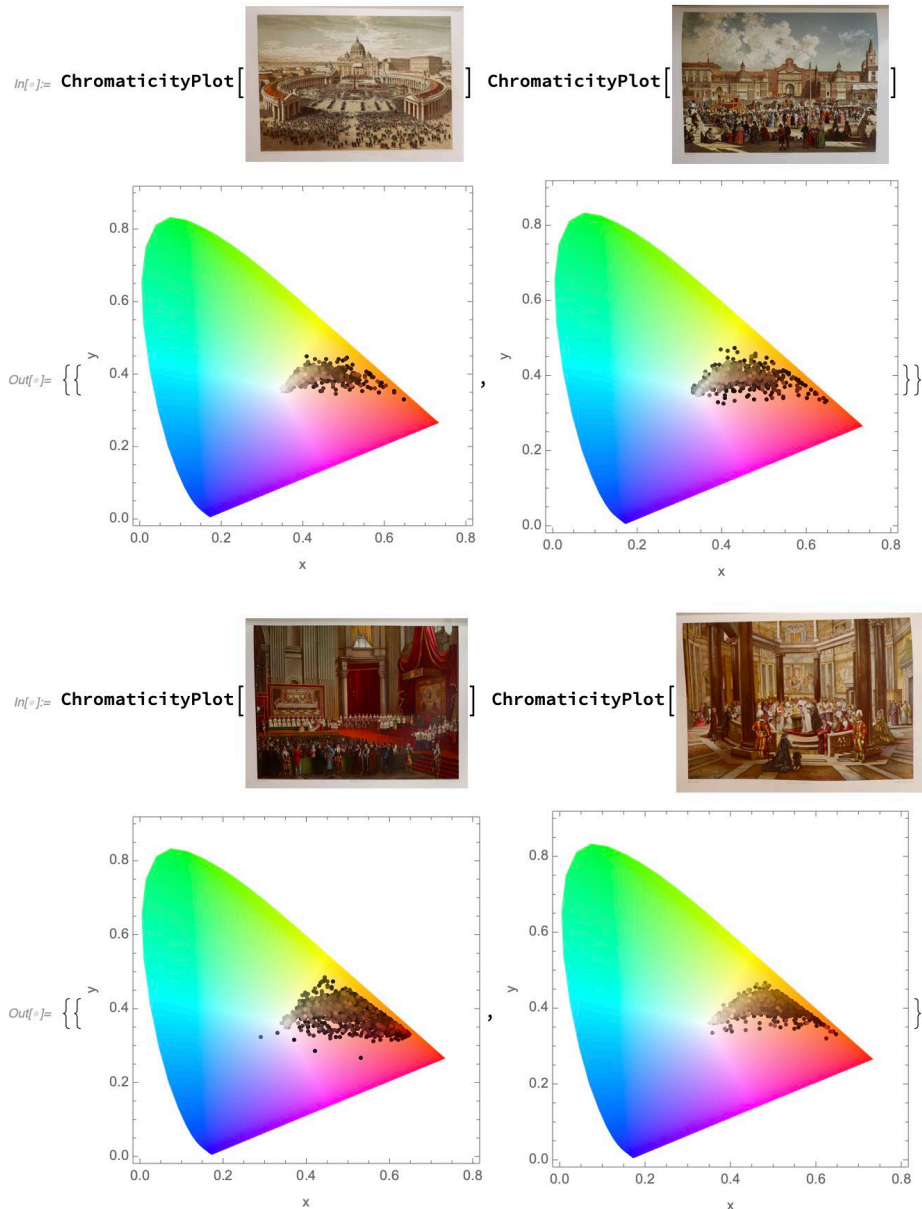

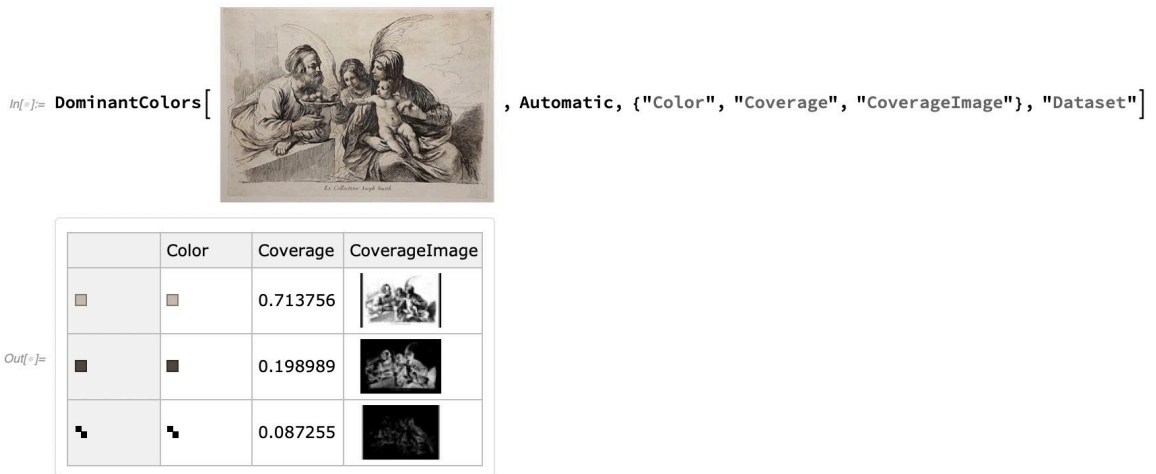Figure 14. Chromaticity in four chromolithographs measured by Wolfram One

In[ ]:= DominantColors[, Automatic, {"Color", "Coverage", "CoverageImage"}, "Dataset"]

	Color	Coverage	CoverageImage
▢	▢	0.713756	
▪	▪	0.198989	
◣	◣	0.087255	

Out[ ]=

Figure 15. Dominant Colours measured by Wolfram One and converted to a Dataset

In[ ]:= FindFaces[OCN1142635756, {"Image", "Strength", "Emotion"}] // Dataset

Image	Strength	Emotion
	0.976685	sadness
	0.932064	sadness
	0.843959	sadness
	0.825374	sadness

Out[ ]=

In[ ]:= HighlightImage[OCN1142635756, MapThread[Labeled, {FindFaces[OCN1142635756], FacialFeatures[OCN1142635756, "Emotion"]}]]

Out[ ]=
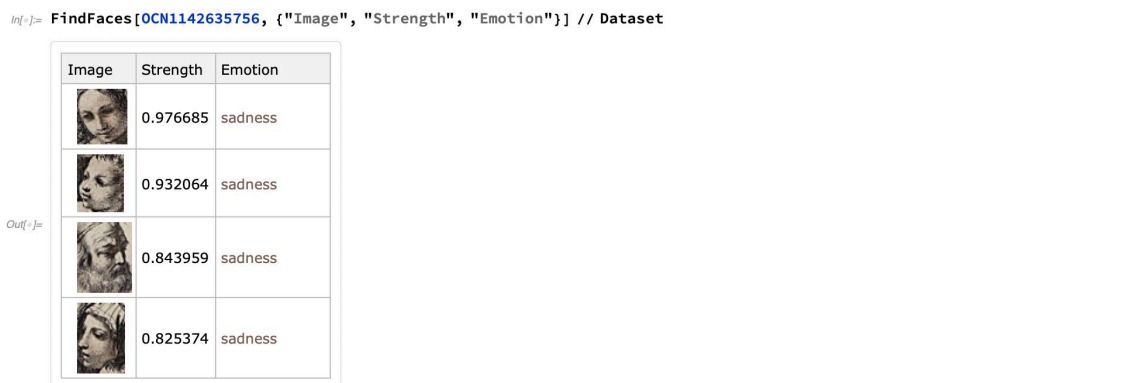
Figure 16. Emotion as recognised by a machine learning model by Wolfram One

```
In[•]:= HighlightImage[OCN1142635756, region]

Out[•]=
```
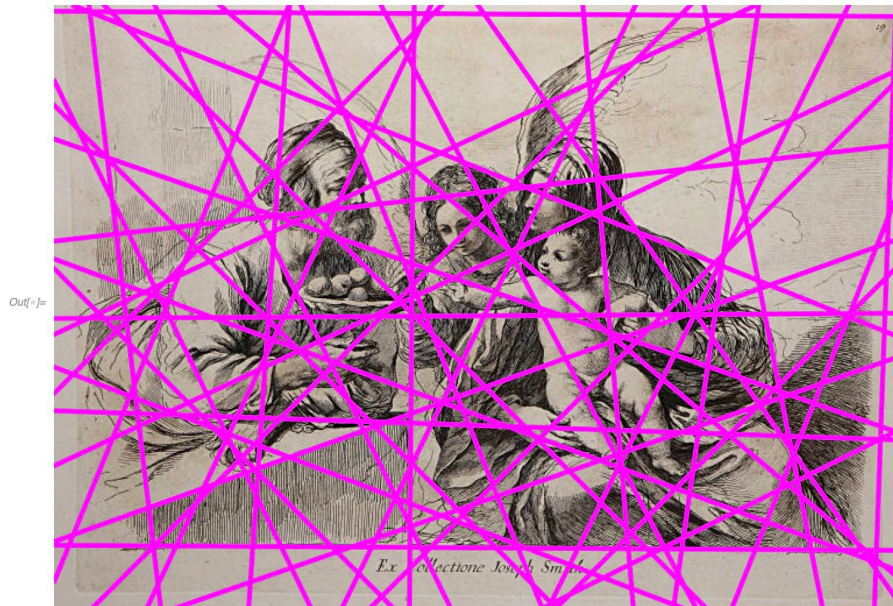
Figure 17. Edge Detection shows how the artist worked to achieve a focal point

Metadata

Metadata is fundamental to the discovery of special collections. While open access parts of library collections can be browsed, typically special collections are held in closed storage spaces within libraries, for conservation and protection. The biggest intersection of information technologies and special collections is how metadata is treated. In order to correctly describe an item, the cataloguer should completely understand what they are describing.

Case Study 1: Incipit Racionale diuinorum officiorum.

Incunabula represents a particular challenge for rare books cataloguing, mainly due to the fact that more modern conventions in regard to printed publication detail are mostly absent from these items. In order to catalogue the material, a fair degree of research must be undertaken to ensure accuracy.

This backlogged item from Mannix Library had next to no information about the item other than a supposed Hain number[10]. A blurb from an auction catalogue or book seller mentioned this item was supposedly Hain 6463. This detail allows for verification of the title, and some detail about the place of publication and the typography. Hain doesn't mention dates or printers in this particular set of

---

[10] Hain numbers come from Ludwig Hain's *Repertorium Bibliographicum* (https://catalog.hathitrust.org/Record/003915335)

information. The auctioneer's blurb suggested that this item was printed by Berthold Ruppel, an apprentice of Johannes Gutenberg, notable for being the first printer in Basel, Switzerland.

Figure 18. Hain 6463

Ruppel, operated in Basel from about 1468 onwards so, in the absence of any other evidence about the printing, the watermark was consulted. The watermark in the work is a bull's head with a tau cross above it. It can be explicitly dated to 1468 from the Piccard watermark database[11]. The mark is associated with a Basel papermaker called Anton Gallician known to have supplied Ruppel[12]. The watermark confirms two facts then, that the printer was likely Ruppel in 1468. Ruppel often worked with another printer called Michael Wenssler, however these works tended to have different types according to who was printing and the work in question doesn't have any evidence of the work of multiple printers. It is most likely that this isn't what the Incunabula Short Title Catalogue links as Hain 6463 (https://data.cerl.org/istc/id00415000), but an earlier edition printed solely by Ruppel.

---

[11] https://www.piccard-online.de/detailansicht.php?PHPSESSID=&klassi=002.003.008.007.001&ordnr=72524&sprache=

[12] Allan Stevenson, *The Problem of the Missale Speciale*, *Bibliographical Society Publication, 1963-1964*, Special Am, Pittsburgh SE: Thomas C. Pears, 1967.
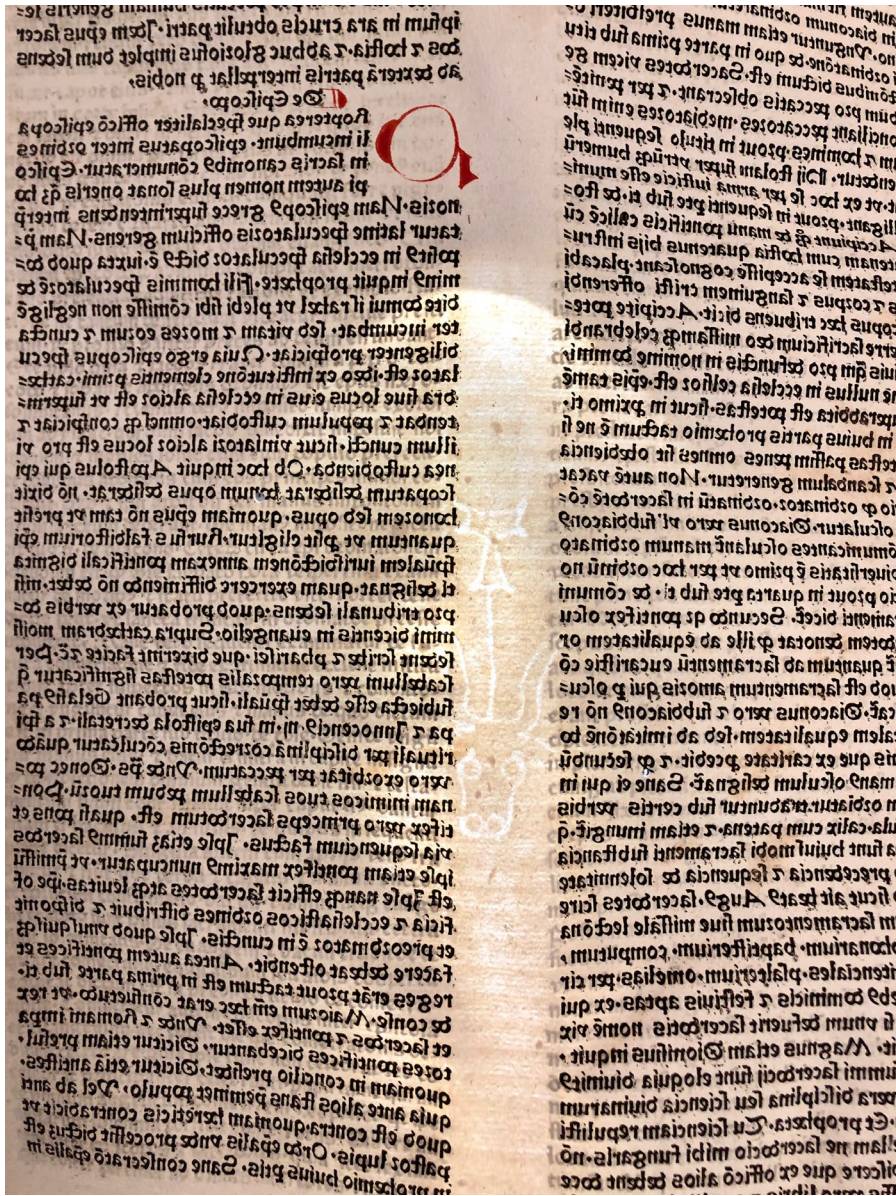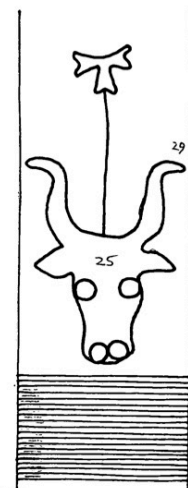
Figure 19. Bull's head Tau watermark

Figure 20. Watermark dated from Piccard Database

Having gathered enough evidence to confidently assign metadata, given that this is an extremely rare item that has previously been undescribed in any literature, a full DCRM(B) (http://rbms.info/dcrm/dcrmb/) record needed to be created. In the MARC we assigned both rda and dcrmb as the standards used, since they are complementary. The descriptive standard for rare books allows for correct transcription of the incipit, since a u is used in place of a v. Key to the DCRM(B) are extensive notes on evidences about the printing of the work, provenance, citations and binding, as well as properly attributed relator terms. Additionally, RBMS controlled vocabularies (http://rbms.info/vocabularies/index.shtml) are applied for richer data for researchers and other special collections librarians. Embedding Linked Data URIs into the record helps to open up the information out of the catalogue. MarcEdit can be used for this purpose with the build links function.
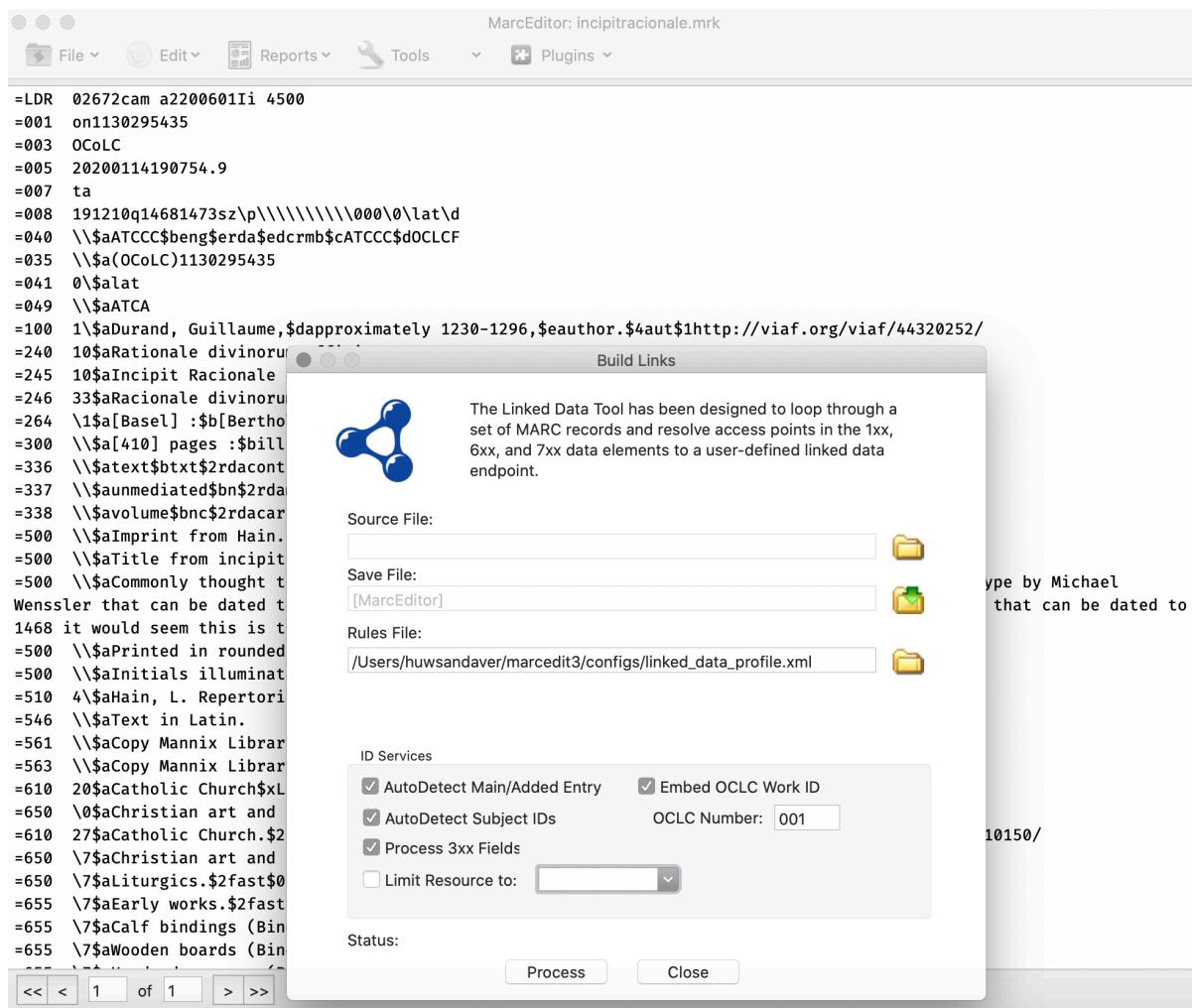
Figure 21. Linked Data function in MarcEdit

The resulting full level MARC record then provides access to the richest set of metadata that can be provided for the item, complete with Linked Data URIs. Linking this item with a digitisation is also possible with annotations. Since this is a newly discovered edition, researchers may be interested in studying the item in detail. One of the challenges with older works and digitisation is that Optical Character Recognition will fail if the typography is unusual. This can be overcome with an annotation server, which allows transcription on a line by line basis with search capability.

```
=LDR  02672cam a2200601Ii 4500
=001  on1130295435
=003  OCoLC
=005  20200114190754.9
=007  ta
=008  191210q14681473sz\p\\\\\\\\\000\0\lat\d
=040  \\$aATCCC$beng$erda$edcrmb$cATCCC$dOCLCF
=035  \\$a(OCoLC)1130295435
=041  0\$alat
=049  \\$aATCA
=100  1\$aDurand, Guillaume,$dapproximately 1230-1296,$eauthor.$4aut$1http://viaf.org/
viaf/44320252/
=240  10$aRationale divinorum officiorum
=245  10$aIncipit Racionale diuinorum officiorum.
=246  33$aRacionale divinorum officiorum
=264  \1$a[Basel] :$b[Berthold Ruppel],$c[between 1468 and 1473]
=300  \\$a[410] pages :$billuminations,$c42 cm (fol.)
=336  \\$atext$btxt$2rdacontent
=337  \\$aunmediated$bn$2rdamedia
=338  \\$avolume$bnc$2rdacarrier
=500  \\$aImprint from Hain.
=500  \\$aTitle from incipit on leaf [a1].
=500  \\$aCommonly thought to have been printed by Berthold Ruppel in 1468, some
editions include type by Michael Wenssler that can be dated to 1477. The present
edition only has one example of type and a watermark that can be dated to 1468 it would
seem this is the earlier edition solely printed by Ruppel.
=500  \\$aPrinted in rounded gothic type in two columns.
=500  \\$aInitials illuminated and rubricated in red.$5ATCCC
=510  4\$aHain, L. Repertorium Bibliographicum,$c6463
```

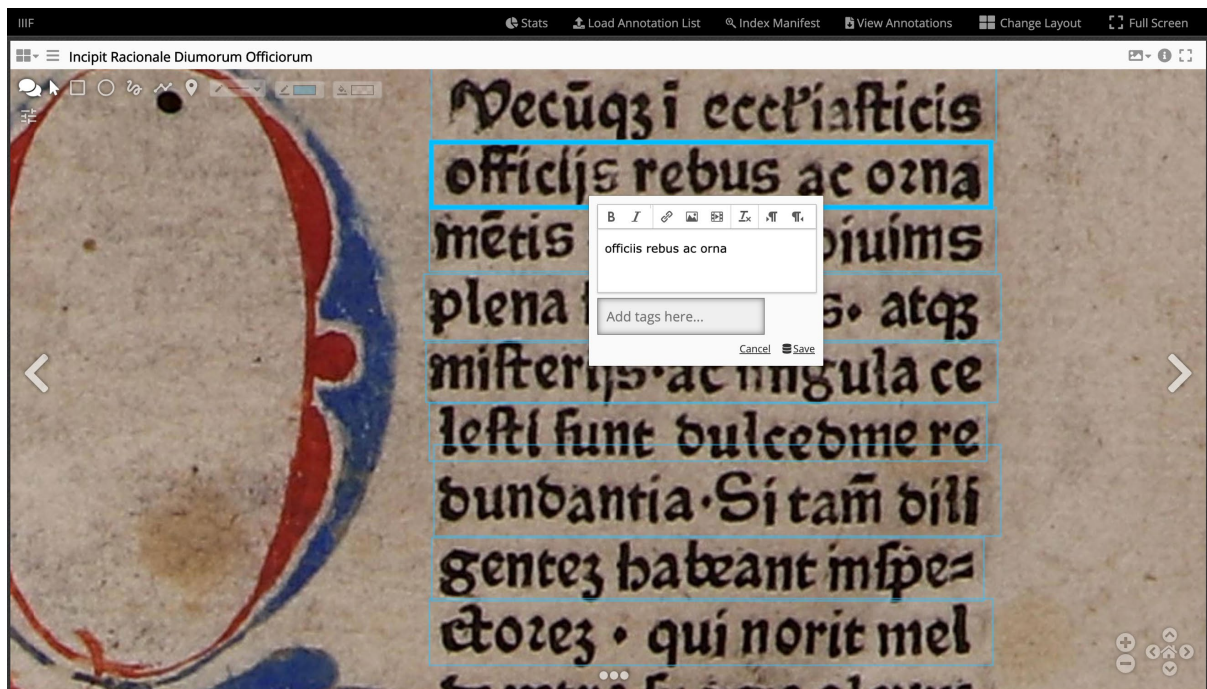Figure 22. Full MARC record for Incipit Racionale Diuimorum Officiorum

Figure 23. Transcription via Annotation Server

Case Study 2: Holy Family with an Angel Who Offers Fruit to the Christ Child

Separately digitising elements of bibliographic items is an increasingly popular way of representing a collection. The case study presented here is Francesco Bartolozzi's *Holy Family with an Angel Who Offers Fruit to the Christ Child* an engraving made in sanguine tone after the Renaissance artist Guercino and printed by Giovanni Battista Piranesi. In Mannix Collections the metadata is handled by CONTENTdm using Dublin Core (https://dublincore.org/). CONTENTdm makes it mandatory to use controlled vocabularies with certain fields. Since primarily the digitisations are representations of art objects or textual objects presented as material culture, we use Getty controlled vocabularies as they are the most appropriate for this kind of material. Subjects are set by the Getty Art and Architecture Thesaurus, names by the Getty Union List of Artist Names and places by the Getty Thesaurus for Geographic Names. Best practices for the use of other Dublin Core fields are used, such as the DCMI Type Vocabulary and DCMI recommended vocabularies for format.

Figure 24. CONTENTdm metadata entry form

Once the data is in CONTENTdm we can then use the IIIF presentation API to manipulate it, for example into a virtual exhibition using Mirador. Using the online Manifest Editor, it is then possible to manipulate the metadata to a more appropriate format for an exhibition. In the case of the prints in

the Mannix Collections the Visual Resources Association core (https://www.loc.gov/standards/vracore/) known as VRA Core is a more appropriate standard as they are able to display data that is much more comparable to a museum didactic label. Manifests can use any schema desirable for use with the images, so while the core database, such as CONTENTdm or another repository may insist on Dublin Core, it's possible to re-imagine your collection with otherwise minimal intervention, while keeping the data hosted centrally.

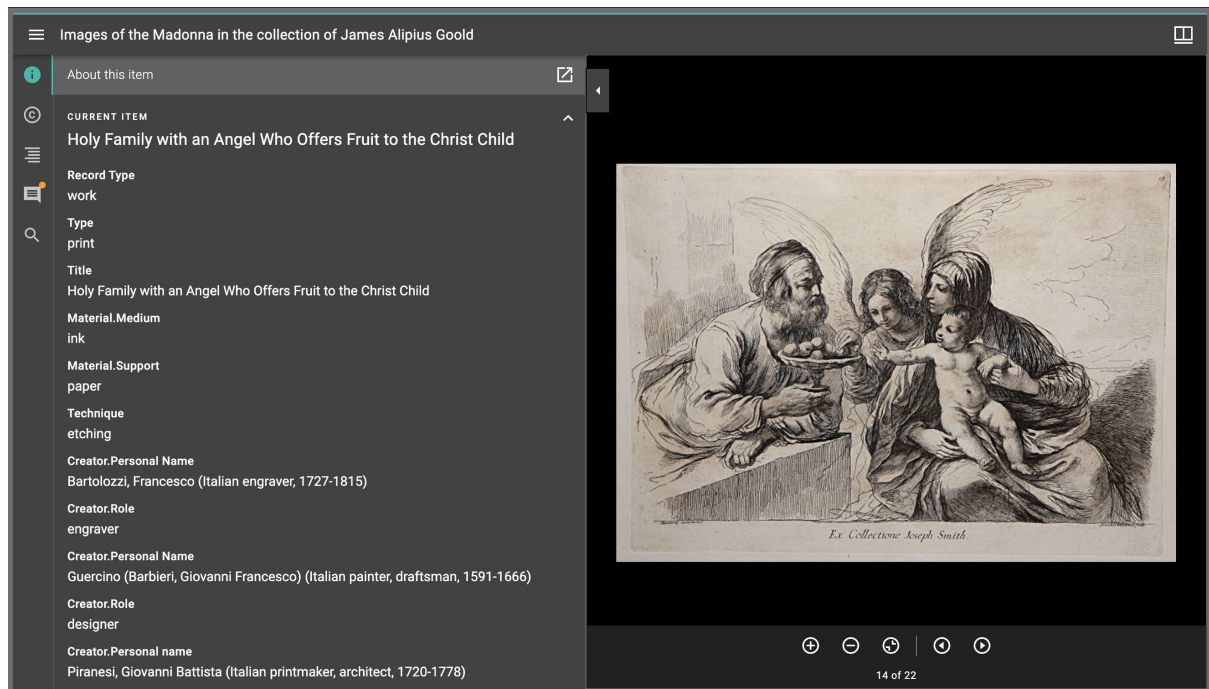Figure 205. VRA Core Metadata entered via Visual Manifest Editor

Figure 216. VRA Core metadata displaying as a virtual exhibition item with Mirador

Conclusion

There are many unique opportunities to break specialised material held in libraries out of their silos. While there may be a distinct learning curve around newer metadata schemas and information technology, and some costs involved, engaging with the collection in new ways using new tools can lead to greater engagement and interest in the material, and not least, a greater understanding of the collection and what is in it.

Bibliography

'RBMS Competencies for Special Collections Professionals: Approved by the ACRL Board of Directors, July 2008'. *College & Research Libraries News* 69, no. 10 (2008): 622–29. https://doi.org/10.5860/crln.69.10.8083

Hain, Ludwig. *Repertorium Bibliographicum, in Quo Libri Omnes Ab Arte Typographica Inventa Usque Ad Annum Md. Typis Expressi, Ordine Alphabetico Vel Simpliciter Enumerantur Vel Adcuratius Recensentur.* Repertorium Bibliographicum. Stuttgartiae: J.G. Cotta; [etc., etc.], 1826. https://catalog.hathitrust.org/Record/003915335

Stevenson, Allan. *The Problem of the Missale Speciale.* [Bibliographical Society Publication for the Years 1963 & 1964.]. Pittsburgh: T. C. Pears III, 1967.