

# ARCHIVING & PRESERVATION OF OPEN ACCESS SCHOLAR- LED PRESS MONOGRAPHS

**Dr Miranda Barnes**  
**COPIM Project**  
**Loughborough University**

**SCURL Conference**  
**15<sup>th</sup> June 2022**



# ‘WORK PACKAGE’ 7: ARCHIVING & PRESERVATION

- There are [7 work packages in COPIM](#).
- Work package 7 is dedicated to the key challenges associated with archiving research monographs in all their variation and complexity, and work towards developing new solutions.
  - What are the boundaries of a “book”?
  - How are complex digital monographs preserved?
  - What are the risks to archived content?
- The work package includes preservation experts, librarians, scholars and publishers from Loughborough University, Open Book Publishers, the Digital Preservation Coalition, the British Library, and Jisc.



# “LEAKY METADATA PIPELINE”



Image: Photo by [Luis Quintero](#) from [Pexels](#)

- Examining the “leaky metadata pipeline” between parties within the archiving & preservation process.
- We are looking at three-party situations:
  - Publisher
  - Aggregator
  - Digital Preservation Service
- Metadata files for OBP, OAPEN, and Portico for two monograph titles we have been working with
  - *Rameau’s Nephew*
  - *Image, Knife, and Gluepot*
- Analysis is ongoing; there appears to be:
  - expected differences (where field evolves to reflect updated information)
  - missing content; and
  - differing content (where same field has similar but ultimately different information, such as “licence name” (Open License vs. CC License), subject code, heading text, etc)
- A summary report will be published when analysis is complete.

## FILE FORMATS: SMALL PUBLISHERS & OA MONOGRAPH PRESERVATION

PDF	XML	EPUB/MOBI
<ul style="list-style-type: none"><li>• Most publishers use PDF/widely adopted</li><li>• Closely mimics familiar book format, but enhanced content may be embedded</li><li>• <b>Some preservation implications for embedded content</b></li><li>• Different formats (PDF vs PDF/A, etc.)</li></ul>	<ul style="list-style-type: none"><li>• Preferred by key preservation players</li><li>• Integrates data, metadata and structure</li><li>• More ideal for long-term preservation than PDF</li><li>• Relationships between content can be defined</li><li>• <b>Many small publishers don't have tech or resource to convert OA monographs to XML</b></li></ul>	<ul style="list-style-type: none"><li>• <b>Proprietary formats</b>, which means there are complications in terms of open access &amp; infrastructure</li><li>• These formats may be useful as part of the preservation process, but not ideal for OA</li></ul>



# ACCESS VS. ARCHIVE COPY

- Because complex digital monographs may contain multiple file types and large file sizes, discussion continues about an access vs. archive copy
- For many access users, a smaller, simpler file “package” could be sufficient (PDF with links, smaller images)
- However, for full preservation, XML, Zip, or other option with all file components actively preserved would be preferable



# EMBEDDED & LINKED CONTENT

**Complex OA monographs** are born-digital publications that cannot be replicated in print form. Moving beyond digital enhancements, complex OA monographs rely on multimodal content and the web's interactive nature to create a distinctive reading experience that often deviates from the linear structure of a printed book. ([Elliott, 2015](#))

Complex digital monographs may have:

- Embedded content, including images, videos, audio files, geospatial data, 3-D models (and others)
- Linked content, QR codes

**What are the boundaries for preservation? Whose responsibility is it to preserve this content, how, and where? If not in the same location as the monograph itself, how might all components be linked together? (How might one distribution platform assist?)**

# OPEN ACCESS PUBLISHING = OA PRESERVATION?

- The majority of preservation platforms keep the archived versions in an inaccessible archive (a “dark archive”) until a trigger event is reached and the work is then made available openly.
- This is generally the default for digital preservation archives, as the majority of content being preserved is proprietary and the publishers wish for all preserved content to remain secure until some future point where they may no longer exist.
- **Should open access publishing mean open access preservation and archiving?**
- **How might a network of repositories contribute to the possibility of this becoming reality?**



[Photo by Tim Douglas](#)

# REPOSITORY WORKFLOWS

- Two OBP monographs, along with their supplementary material, have been deposited in Loughborough University's test instance of Figshare. This has been done both manually and automated via API.
  - [Rameau's Nephew](#)
  - [Image, Knife & Gluepot](#)
- The manual deposit of both books and their accompanying essential and supplementary materials, as well as manual input of metadata, took several days
- This raised the necessity of developing a way of automating the process, particularly for small publishers who often have the least resource, in terms of staffing, expertise, and technology







\_\_\_\_\_

11/11/2019




# IMAGE, KNIFE AND GLUEPOT (K. RUDY)

 [Browse](#)  [Upload](#) [My data](#)  

 **Image, Knife, and Gluepot: Early Assemblage in Manuscript and Print**

[+ Follow](#) Published on 07.10.2021 - 11:21 by [Miranda Barnes](#)

[USAGE METRICS](#) 

5 views

Originally published by Open Book Publishers.

This collection in includes the PDF and XML formats of the monograph.

Kathryn M. Rudy. Image, Knife, and Gluepot: Early Assemblage in Manuscript and Print. Cambridge, UK: Open Book Publishers, 2019, <https://doi.org/10.11647/OBP.0145>


Chapter Abstracts:

1. Cut, Pasted, and Cut Again: The Fate of 140 German and Netherlandish Single-Leaf Prints at the Hands of a Limburg Franciscan and a Modern Connoisseur


Chapter 1 reconstructs a manuscript by beghards in Maastricht. (Beghards were men who followed St Francis, considered themselves to be Franciscans, and who lived in community in towns and cities. They did not beg but made their living through their trade and labour.) By reconstructing this book, I can show how the beghards learned to integrate the new technology of printed images into the making of their manuscript. That book straddles the two technological moments and sees the creators adapting midstream.

2. A Novel Function for the Calendar in Add. Ms. 24332

A short Chapter 2 discusses the significance of the unusual calendar the beghards constructed, which goes some distance to help us understand how they thought in terms of fungible categories: these innovators in the realm of print technology were also inventing new ways of organising information. They experimented with how the book

 **Loughborough University**


MEMBERS (1)

 [Miranda Barnes](#)




[cite all items](#)


SHARE

[facebook](#) [twitter](#) [linkedin](#) [email](#)




[need help?](#) [+ Follow this search](#)

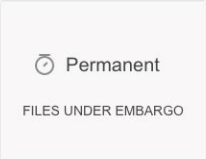
4 results found sort by: Date added   



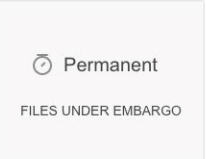
**Image, Knife, and Gluepot: Early Assemblage in ...**  
Monograph posted on 07.10.2021 in Loughborough University  
[Kathryn M. Rudy](#)



**Image, Knife, and Gluepot: Early Assemblage in ...**  
Monograph posted on 07.10.2021 in Loughborough University  
[Kathryn M. Rudy](#)



**Image, Knife, and Gluepot: Early Assemblage in ...**  
Monograph posted on 07.10.2021 in Loughborough University  
[Kathryn M. Rudy](#)



**Image, Knife, and Gluepot: Early Assemblage in ...**  
Monograph posted on 07.10.2021 in Loughborough University  
[Kathryn M. Rudy](#)



# THOTH PRESERVATION NETWORK

- THOTH is working towards possible developments that would allow for dissemination of files in addition to metadata.
- The hope is that this could allow for pushbutton deposit of OA monographs and accompanying materials into a repository of choice by publishers, as “just another” form of dissemination.
- Longer-term goal is developing a willing network of university repositories that would be happy to receive and archive the OA monographs of small and scholar-led publishers.
- While we acknowledge repository archiving is not a “perfect” solution or truly preservation, for many small OA monograph publishers with little or no preservation in place, this would allow for a reduced risk to this essential contribution to the scholarly record.
- We have confirmed interest from three University libraries, along with informal interest from other parties.



# THOTH PRESERVATION NETWORK

- We have developed successful proof-of-concept for the three major repository platforms: Figshare, Eprints, and Dspace
- This has been done via workflow experimentations with API, Python, and the SWORD Protocol by a member of our developer team working with Thoth
- At it's simplest, Thoth will have links to both the textual material and the associated material, and the system will use those links to transfer the available content to a partner repository.
- Our aim is to simplify this process to a single "Send to archive" button. We are hopeful that this will be achievable by the end of the project in April 2023.

ANY QUESTIONS?





# THANK YOU!

If you or your University repository might be interested in participating in the Thoth Archiving Network as it develops, please get in touch!

**Dr Miranda Barnes**

Twitter:

**@MirandaB\_OA**

Email address:

**m.l.barnes@lboro.ac.uk**

