

Finding the Golden Mean: An Efficient Model for Improving Discovery and Access for Legacy Theses and Dissertations in a Medical Library's Institutional Repository





Presenter


Brandon T. Pieczko, MSLIS

Digital and Special Collections Librarian

Ruth Lilly Medical Library

Indiana University School of Medicine

Indianapolis, IN

 [0000-0003-0054-7260](https://orcid.org/0000-0003-0054-7260)



1

Project Background



Project Background

- Ongoing challenge: providing efficient and cost-effective access to student and faculty scholarship, including legacy print theses and dissertations
- Usage statistics for print theses and dissertations (Jul 2004 - Oct 2023)
 - Total uses: 495
 - In-house uses: 156
 - Circulation checkouts: 334
 - ILL requests: 5 (since 2018)



2

Project Planning



Project Planning

- Project planning started in June 2022
 - Approximately 1,400 bound print theses/dissertations from IU School of Medicine (1953-2008)
 - Digitize and upload to ScholarWorks (DSpace institutional repository)
- Initial cost estimate for full-text digitization
 - \$0.75 per page scanned
 - 1,400 volumes X 150 pages average per volume = 210,000 pages
 - Total time investment: 11,200 hours (8 hours per volume)
 - Total cost investment: \$157,500 (\$112.50 per volume)
- Zoiks! (neither efficient nor cost-effective)



Our “Golden Mean” Approach

- Focus on exposing the descriptive metadata for all 1,400 items
 - Extract bibliographic information and create individual metadata records in ScholarWorks
- Create “preview” scans of the frontmatter for each item (~20 minutes per item)
 - Title page, list of committee members, committee member signatures, abstract, table of contents
- Implement “scan-on-demand” service model for full-text requests
 - *“This document only includes an excerpt of the corresponding thesis or dissertation. To request a digital scan of the full text, please contact the Ruth Lilly Medical Library’s Interlibrary Loan Department (rlmlill@iu.edu).”*



Addressing Copyright Concerns (sort of)

- Ask for forgiveness rather than permission
- Contacting individual copyright holders would be too onerous
- Disclaimer for full-text scans
 - *"If you are the author and/or copyright holder of this item and would like to submit a request to have it removed from the university's institutional repository, please contact the Ruth Lilly Medical Library (medlref@iupui.edu)."*



3

Project Implementation



Metadata Extraction and Cleanup

Exported bibliographic records for theses and dissertations from library catalog as a single MARC (.mrc) record

```
LEADER 00000cam a2200000 a 45 0
001 249693
008 920326s1970 xx a b 000 0 eng d
035 a| (Sirsi)249693
035 a| (InU)ACK4003IM
035 a| (OCoLC)ocm25532162
040 a| IUM c| IUM
049 a| IUMB
100 1 0 a| Brahmi, Zacharie, d| 1942- =| ^A3590973
245 1 0 a| Single cell studies of cross-reactive antibodies to closely related haptens
/ c| by Zacharie Brahmi.
260 0 a| [S.l. : b| s.n.], c| 1970.
300 a| 103 leaves : b| ill. ; c| 28 cm.
502 a| Thesis (Ph. D.)--Indiana University, 1970.
500 a| Vita.
504 a| Includes bibliographical references.
650 2 a| Haptens. =| ^A3696979
650 2 a| Cross Reactions. =| ^A3674269
650 2 a| Antibodies. =| ^A3668166
710 2 0 a| Indiana University School of Medicine. b| Department of Microbiology =|
^A1466703
998 a| 3/26/1992 b| 9/13/2000 c| 3/26/1992
996 a| 49
999 a| XX(249693.2) ij 30000022806784 jj XX(249693.2) II THESES m|
I-MEDICINE ni THESES w| ASIS
```

Single cell studies of cross-reactive antibodies to closely related haptens

Author [Brahmi, Zacharie, 1942-](#)
Title [Single cell studies of cross-reactive antibodies to closely related haptens / by Zacharie Brahmi.](#)
Format Book
Published [S.l. : s.n.], 1970.
Description 103 leaves : ill. ; 28 cm.
Thesis (Ph. D.)--Indiana University, 1970.
Other contributors [Indiana University School of Medicine. Department of Microbiology](#)
Notes Vita.
Includes bibliographical references.
Medical subjects [Haptens.](#)
[Cross Reactions.](#)
[Antibodies.](#)

Holdings

Library [Indpls - Ruth Lilly Medical Library](#)
Location [Theses](#)
Floor [1st Floor](#)



Metadata Extraction and Cleanup

Used “Export Tab Delimited Text” feature in MarcEdit to export selected fields into comma separated value (CSV) file

The screenshot shows the 'Export Tab Delimited Records' dialog box in MarcEdit, specifically the 'Settings....' window for 'Step 1: Set File Paths'. The window title is 'Export Tab Delimited Records'. Under the heading 'Export Tab Delimited Records', the first step is 'Set File Paths'. There are two text input fields: the first contains the file path 'C:\Users\bpiec\Desktop\20230126.RLML.theses.mrc' and has a folder icon to its right; the second contains 'C:\Users\bpiec\Desktop\20230126.RLML.theses.csv' and has a save icon to its right. Below these are two sets of dropdown menus: 'Select Field Delimiter' is set to 'Comma (,)', and 'In field delimiter' is set to '|'. There is also a 'Contextual delimiter' dropdown which is currently empty. At the bottom, there are two buttons: 'Next >>' and 'Close'.

The screenshot shows the 'Export Tab Delimited Records' dialog box in MarcEdit, specifically the 'Settings....' window for 'Step 2: Define Fields/Subfields to Export'. The window title is 'Export Tab Delimited Records'. Under the heading 'Export Tab Delimited Records', the second step is 'Define Fields/Subfields to Export'. A list box contains the following field identifiers: '100', '245\$a', '245\$b', '260\$c', '502', '650\$a', and '710\$a'. The '100' field is currently selected. Below the list box is a checked checkbox labeled 'Normalize field data'. Underneath this checkbox is a yellow highlighted area containing two dropdown menus labeled 'Field:' and 'Subfield:', and two buttons: 'Add Field' and 'Delete Field'. At the bottom, there are three buttons: 'Back', 'Export', and 'Close'.



Metadata Extraction and Cleanup

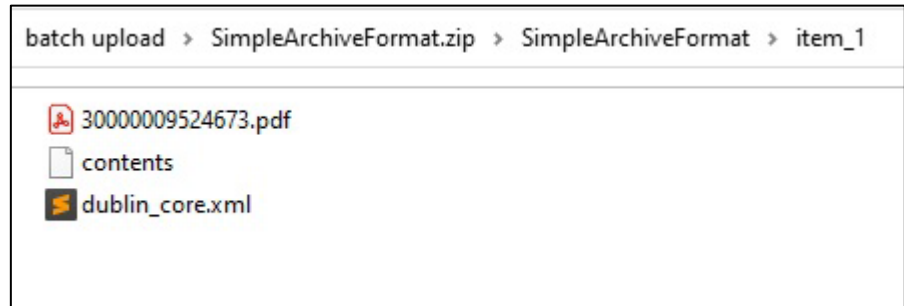
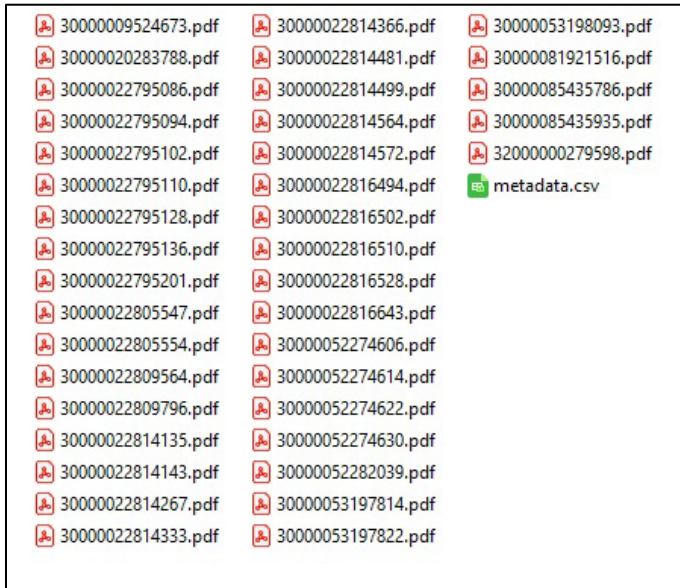
Used LibreOffice Calc (not Microsoft Excel) to clean up the bibliographic metadata in the CSV spreadsheet in preparation for metadata crosswalk to Dublin Core

METADATA CROSSWALK		
MARC Field	DSpace Dublin Core Field	Example
100 a	dc.contributor.author	Brahmi, Zacharie
245 a	dc.title	Single cell studies of cross-reactive antibodies to closely related haptens
245 b		
260 c	dc.date.issued	1970
	dc.degree.date	1970
502 a	dc.degree.level	Ph.D
	dc.type	Dissertation
650 a	dc.subject	Haptens Cross Reactions Antibodies
710 a	dc.degree.grantor	Indiana University
710 b	dc.degree.discipline	Department of Microbiology
999 i	filename	30000022806784
*Added field	dc.description	This document only includes an excerpt of the corresponding thesis or dissertation. To request a digital scan of the full text, please contact the Ruth Lilly Medical Library's Interlibrary Loan Department (rlmlill@iu.edu).
*Added field	dc.language.iso	en_US



Metadata Extraction and Cleanup

Batch uploaded PDF scans and corresponding CSV metadata to ScholarWorks repository using [Simple Archive Format Packager](#) command line script for DSpace



4

Project Outcomes



Project Outcomes

679 theses/dissertations (~50%) partially scanned and uploaded to ScholarWorks since May 2023

Usage statistics for theses and dissertations in ScholarWorks (May 4 - Jul 18, 2023)

- **Total uses: 37,677**
 - Item page views: 30,277
 - Item bitstream downloads: 7,400

Received first request for full-text scan on September 26, 2023:
<https://hdl.handle.net/1805/33601>

The screenshot shows the IUPUI ScholarWorks website. The header includes the IUPUI logo and 'Indiana University'. Navigation links include 'Communities & Collections', 'Browse ScholarWorks', and 'Statistics'. The main content area is titled 'School of Medicine Theses and Dissertations' and features the Indiana University School of Medicine logo. Below the title, there is a 'Browse' section with tabs for 'Recent Submissions', 'By Issue Date', 'By Author', 'By Title', 'By Subject', 'By Date Submitted', and 'By Subject Category'. The page indicates 'Now showing 1 - 10 of 641' items. Two items are visible in the list:

- Item**
Studies on triadin and insights into its role in Ca²⁺ Release
(1999) Kobayashi, Yvonne Monique
- Item**
The role of carotid body chemoreceptors on the vascular capacitance system
(1985) MacAnespie, Carol L.

5

Project Next Steps



Next Steps

- Continue creating preview scans and uploading to ScholarWorks until complete
- Add ScholarWorks links to bibliographic records in local library catalog and WorldCat
- ScholarWorks metadata enrichment (adding MeSH subject terms)
- Continue filling requests for full-text scans and uploading them to ScholarWorks



Resources

- Dietz, Peter & DSpace Contributors (2019 March 12). SAFBuilder – Item Packager from CSV. <https://github.com/DSpace-Labs/SAFBuilder>
- IUPUI ScholarWorks (2023). Indiana University School of Medicine Theses and Dissertations. <https://hdl.handle.net/1805/32777>
- LibreOffice (2023). Calc. <https://www.libreoffice.org/discover/calc/>
- Reese, Terry (2023 August 15). MarcEdit: Your Complete Free MARC Editing Utility. <https://marcedit.reeset.net/>
- Van Kleeck, David (2022 July). RDS@UF: Using MarcEdit to Convert .mrc File to Tab Delimited File for Excel. George A. Smathers Libraries, University of Florida. <https://guides.uflib.ufl.edu/c.php?g=147853&p=6878235>

THANKS!

Brandon T. Pieczko

Digital and Special Collections Librarian

Ruth Lilly Medical Library, Indiana University School of Medicine

<https://orcid.org/0000-0003-0054-7260>

Questions?





CREDITS

Presentation template by [SlidesCarnival](#)



This work is licensed under a [Creative Commons Attribution 4.0 International License](#).