In for a Penny, In for a Pound: The Management and Preservation of E-Theses in EPrints

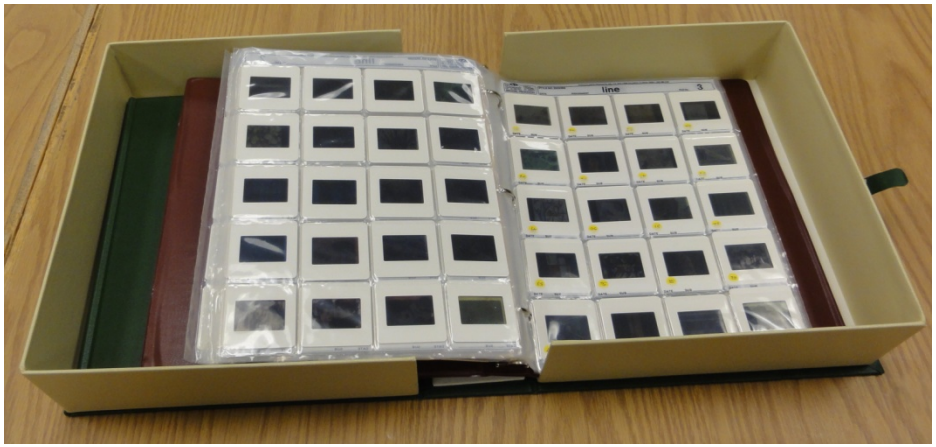
Slavko Manojlovich
Associate University Librarian (IT)
Memorial University of Newfoundland

Talking Points

- **Digitizing Analogue Collections**
- **EPrints vs CONTENTdm**
- **OCR of Analogue Collections**
- **PDF/A File Format**
- **Library Archives Canada Theses Canada Portal**
- **School of Graduate Studies E-Thesis Submission**
- **Discoverability**
- **Long-term preservation**

Digitizing Analogue Collections

- **Microfilm (colour maps, images, etc.)**
- **Supplementary Files**



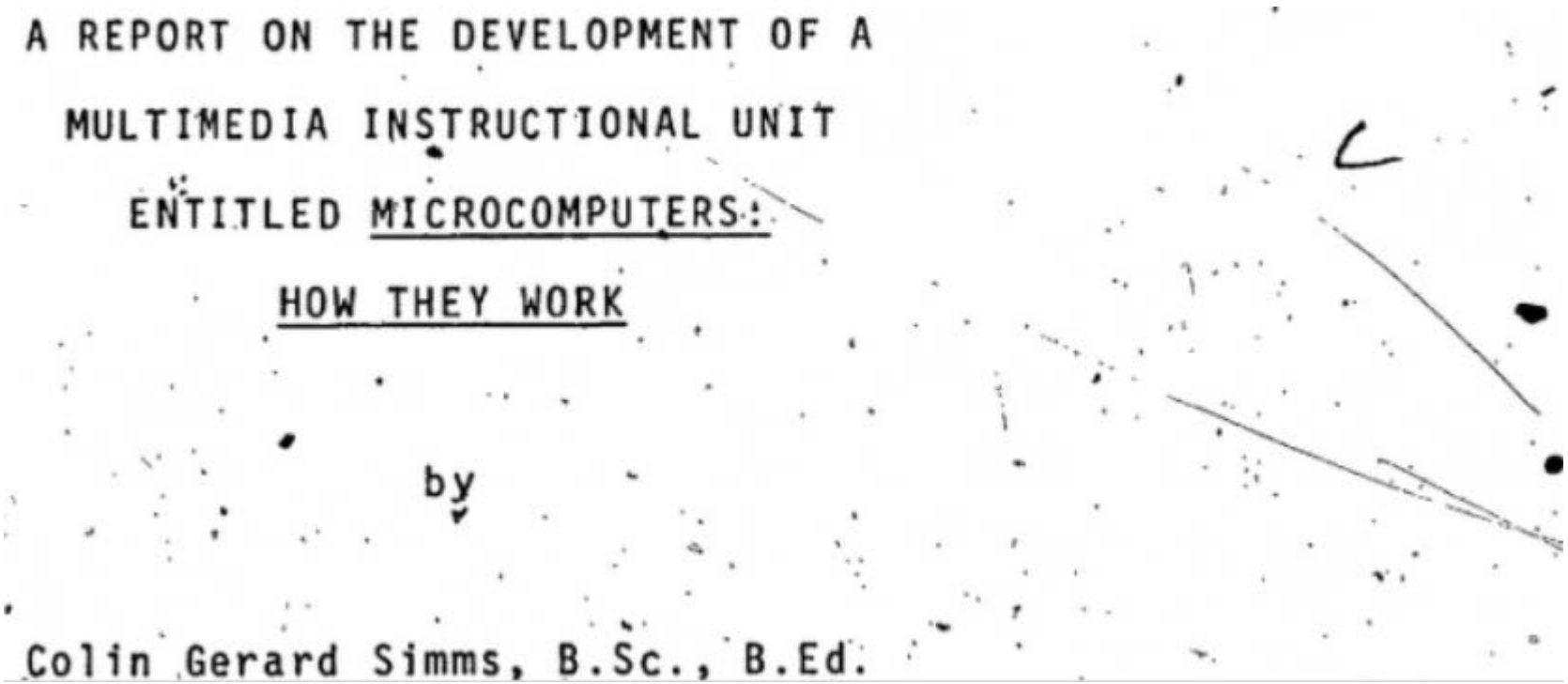
Digitizing Analogue Collections

- **Supplementary Files**

EPrints vs CONTENTdm

Feature	EPrints	CONTENTdm
Database	👍	👎
Indexing	👍	👎
Multiple Output Schemas	👍	👎
Long-term Preservation	👍	👎
Web-based staff client	👍	👎
Usage Statistics	👍	👎

Digitized microfilm and OCR



A REPORT ON THE DEVELOPMENT OF A

MULTIMEDIA INSTRUCTIONAL UNIT

ENTITLED MICROCOMPUTERS:

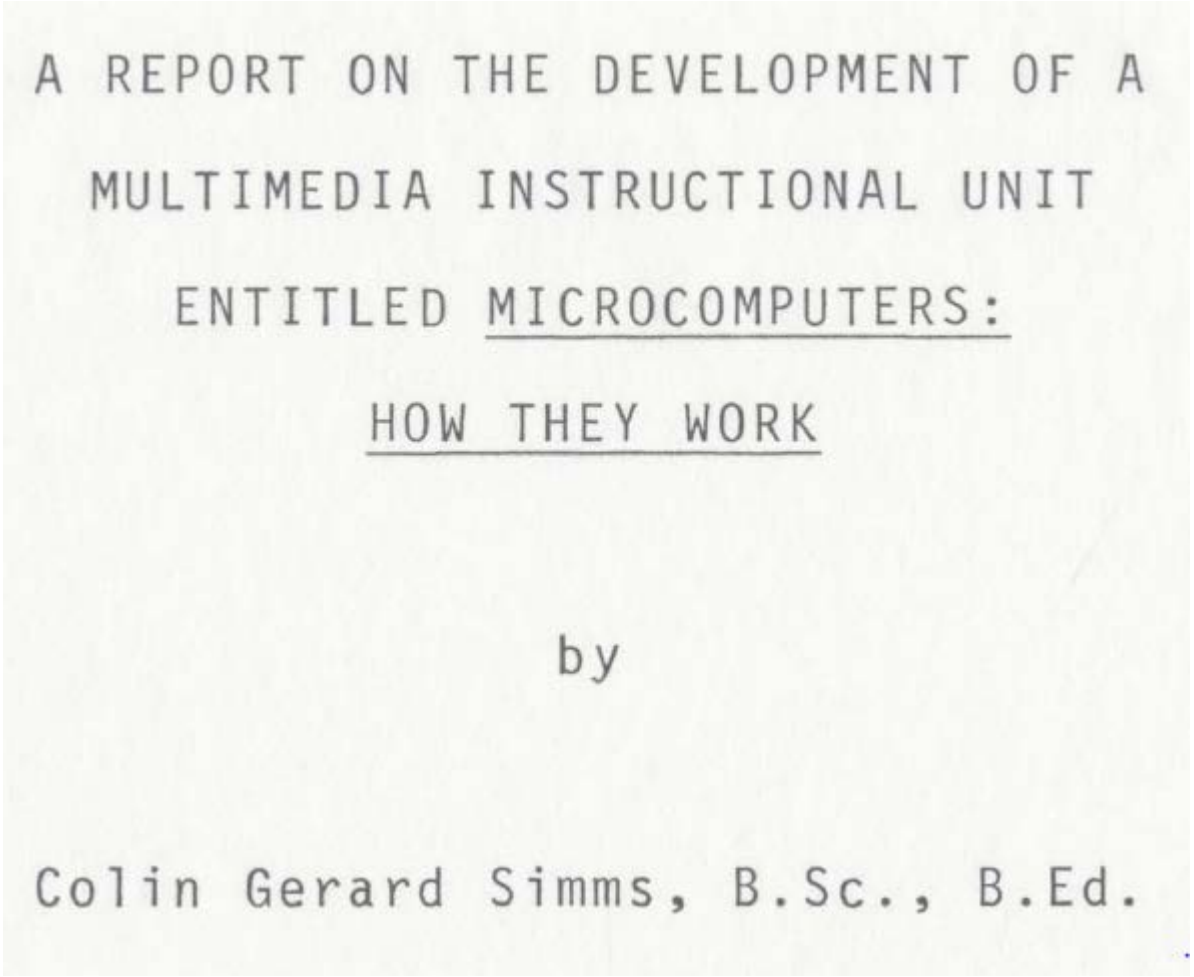
HOW THEY WORK

by

Colin Gerard Simms, B.Sc., B.Ed.

Microfilmed by Library Archives Canada in 1986

Digitized microfilm and OCR



A REPORT ON THE DEVELOPMENT OF A

MULTIMEDIA INSTRUCTIONAL UNIT

ENTITLED MICROCOMPUTERS:

HOW THEY WORK

by

Colin Gerard Simms, B.Sc., B.Ed.

Rescanned by the DAI in 2014

OCR of Analogue Collections

- **Pixel width (ppi) and OCR quality – 1200 ppi**

Mnu<rimo airns to he ~ distil><'tivdy compr~h~nsiv~ C~ nadian jotwn~l
of
Classical Studks. publishing Hticks :•nd rtvi~ws in l::>oth Fr-., nch ~n-d
English. On~ isstl<" annually is nur·rn."llly d~ v ott"d tu a r<:ha~ological
top·
k,;. incl uo.ling fidd o·q xwts. finds a nalpis. " "d th~ histm·y of ~ l'l in
<tn·
riq uity. The nrh~r - two issu~s wdrmnc wnrk in all .11·eas of intnest tu
scholm·s: this
includesbothtr·adirionalandinnnvativrrescat·chinphilology.
history.philosophy.f"'dagngy.and r-,ceptionstudies,aswellas
originalwork inano.l tr·anslationsintoC!-eekandlatin

OCR of Analogue Collections

- **Pixel width (ppi) and OCR quality – 1600 ppi**

Moust:iun aims to be a distinctivdy compn:hensive Canadian journal of Classical Studies. publishing anicles and t·eviews in both French and English. One issue annually is normally d~::voted to at·chaeologica l topics.
including field repm·ts. finds analysis. and the histm·y of at·t in antiquity.
The othe1· two issues welcome w(Wk in all areas of interest to scholars: this includes both traditional and innovative rtseat'ch in philology.
history. philosophy. pedagogy. and reception studies. as wdl as original wot·k in and translations into Greek and Latin

OCR of Analogue Collections

- **Pixel width (ppi) and OCR quality – 2000 ppi**

Mouseion aims to be a distinctively comprehensive Canadian journal of Classical Studies, publishing articles and reviews in both French and English. One issue annually is normally devoted to archaeological topics,
including field reports, finds analysis, and the history of art in antiquity. The other two issues welcome work in all areas of interest to scholars; this includes both traditional and innovative research in philology,
history, philosophy, pedagogy, and reception studies, as well as original work in and translations into Greek and Latin.

Pixel width is determined by page size and font.

PDF/A File Format

- **PDF/A-1**
- **PDF/A-2**
- **PDF/A-3**
- **See recent National Digital Stewardship Alliance Report:**

The Benefits and Risks of the PDF/A-3 File Format for Archival Institutions

Preserve Look and Feel doesn't always work: postscript→PDF/A

MEMORIAL UNIVERSITY

Contents

Abstract ii

Acknowledgements iii

List of Tables vii

List of Figures x

1 Introduction 1

1.1 Models for univariate multinomial data 2

Preserve Look and Feel doesn't always work: postscript→PDF/A

MEMORIAL UNIVERSITY

Contents

Acknowledgements iii

Abstract iv

1 Introduction 5

2 Background 10

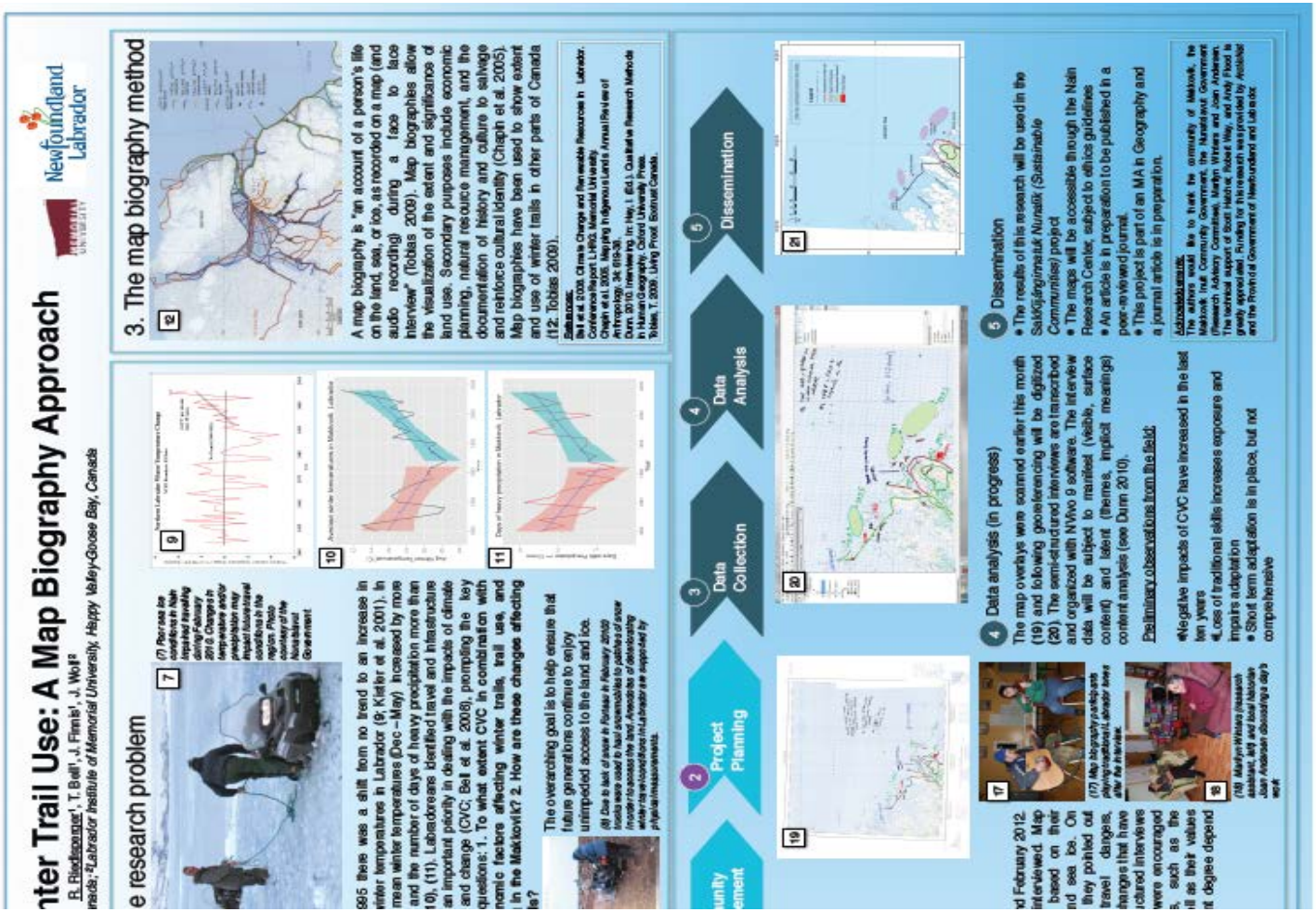
2.1 Linear and Nonlinear Feedback Shift Registers 10

2.2 Grain 12

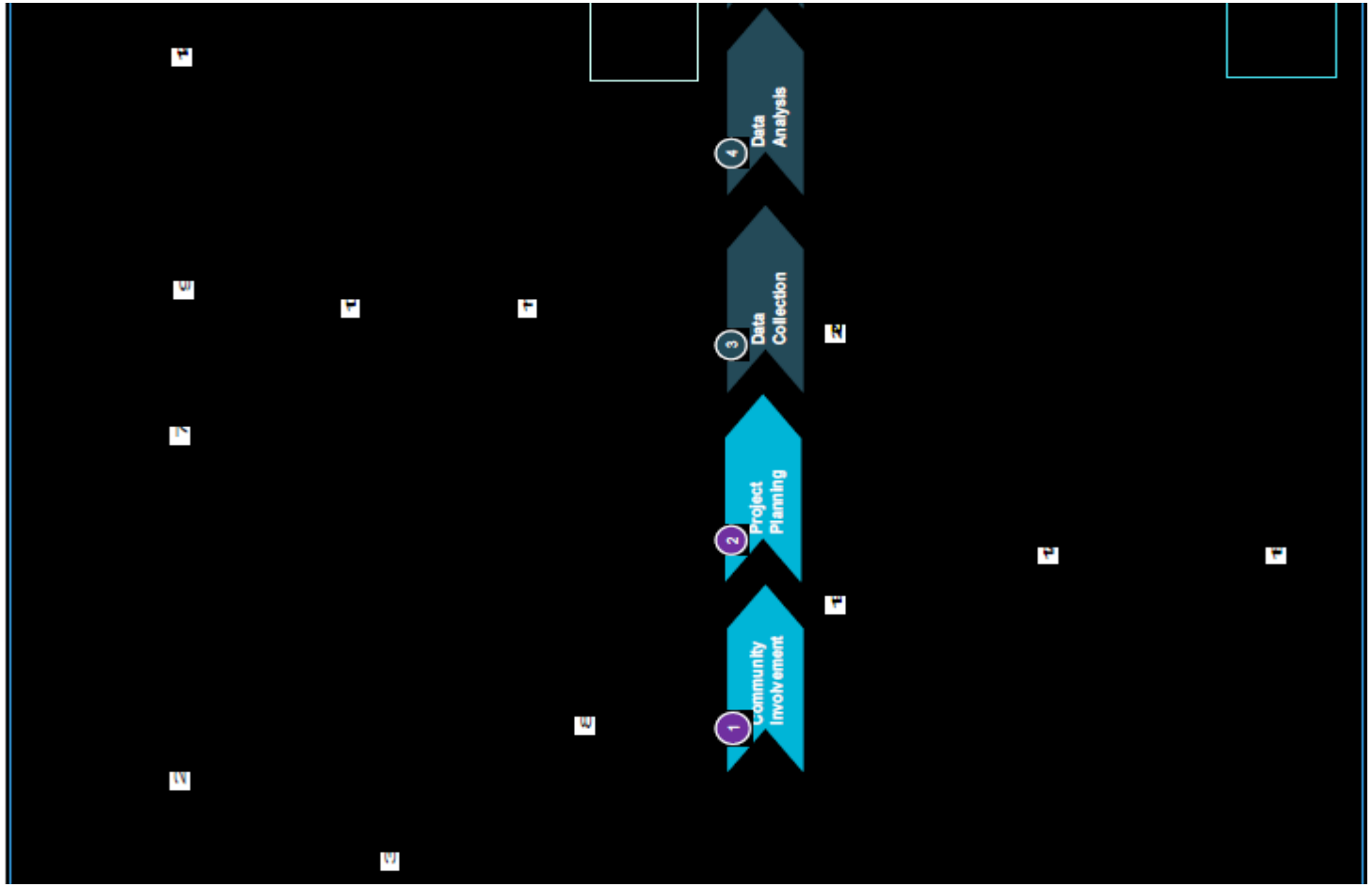
2.3 E0 17

2.4 LILI-128 18

Preserve Look and Feel doesn't always work for images → PDF/A



Preserve Look and Feel doesn't always work for images → PDF/A



Always check each page of a migrated PDF→PDF/A for problems.

Library and Archives Canada Theses Canada Portal

- **E-thesis only submission as of April, 2014.**
- **OAI-PMH harvesting requirements:**
  - ETD_MS metadata schema
  - Single PDF or PDF/A file

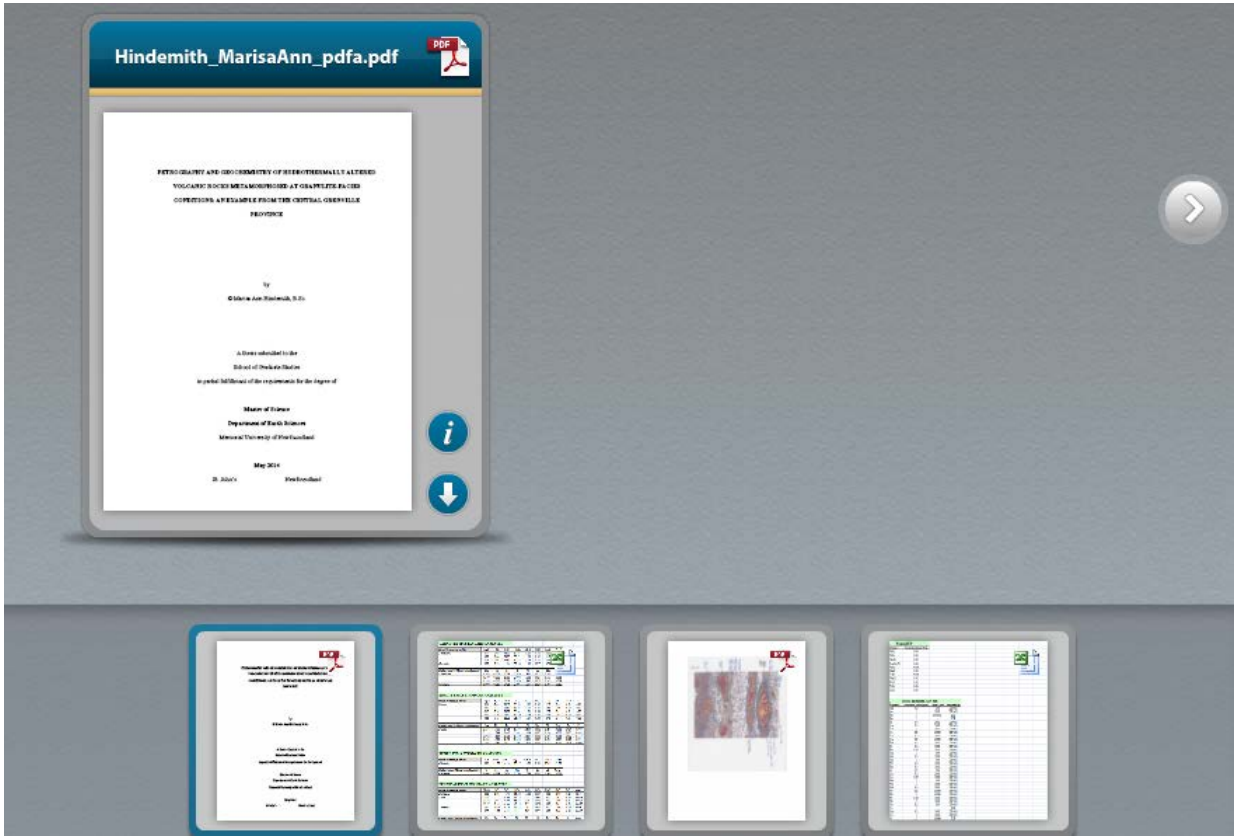
Library and Archives Canada Theses Canada Portal

- **ETD_MS**
  - Cannot use UKETD-DC schema
  - Loaded ETD_MS schema developed by Concordia University
- **Single PDF or PDF/A file**
  - Use Adobe PDF Portfolio container format to deliver thesis and supplementary files to Library Archives Canada. Supported by many non-Adobe readers and Adobe IOS reader.

Library and Archives Canada Theses Canada Portal

MEMORIAL UNIVERSITY

❖PDF Portfolio



Library and Archives Canada Theses Canada Portal

- **PDF Portfolio**

School of Graduate Studies E-Thesis Submission

dsadsad	File folder					6/2/2014 10:14 AM
folder-1	File folder					6/2/2014 10:14 AM
folder-2	File folder					6/2/2014 10:14 AM
folder-the-first	File folder					6/2/2014 10:14 AM
root	File folder					6/2/2014 10:14 AM
appendices	Microsoft Excel Comma S...	1 KB	No	1 KB	0%	6/2/2014 10:14 AM
converted	Adobe Acrobat Document	1,834 KB	No	1,834 KB	0%	6/2/2014 10:14 AM
keywords	Microsoft Excel Comma S...	1 KB	No	1 KB	0%	6/2/2014 10:14 AM
metadata	Microsoft Excel Comma S...	1 KB	No	1 KB	0%	6/2/2014 10:14 AM
thesis	Microsoft Excel Comma S...	1 KB	No	1 KB	0%	6/2/2014 10:14 AM
thesis	Adobe Acrobat Document	1,398 KB	No	1,398 KB	0%	6/2/2014 10:14 AM

E-Thesis submission → Eprints XML and BagIt AIP

Discoverability

- **Library catalogue: transform EPrints xml to MARC21.**
- **Theses Canada Portal: EPrints OAI-PMH ETD_MS schema.**
- **Summon Discovery Service: EPrints OAI-PMH qualified Dublin Core schema.**

Long-term Preservation

- **Media type preservation plan for theses supplementary files (e.g. MPEG-4 H.264 access file format for video); ongoing problems with new versions of browsers.**
- **Research data curation plan for theses with associated data sets (spreadsheets, access databases, software, etc.)**

Long-term Preservation

- **M-DISC (1,000 year storage solution)**
- **SAFE PLN private LOCKSS network: (BagIt archive information package)**