

PRESERVATION PLANNING

Chris Grygiel,
Digital Archive Project Assistant,
Special Collections



Project Outline

To develop ingestion procedures for the preservation of born digital material in existing collections, as well as developing, testing and refining workflows for new accessions and deposits.

Survey to establish existing material

- Lack of agreed standards has led to a range of terminology to date.
 - Search terms in Emu (our collections management software), in a range of fields (title, scope and content, extent and medium) with a range of terms (floppy, disc, disk, CD, HD, hard drive, etc...)
- Discussion with archivists
 - Agree terminology standards.
 - Establishing status of new acquisitions/accessions not yet catalogued, update regularly on developments and any new procedures as and when they arise.

Findings

Born digital items present in 68 different collections:

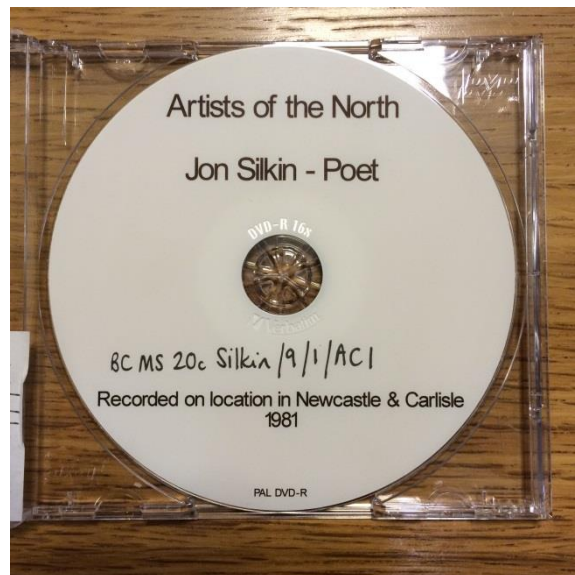
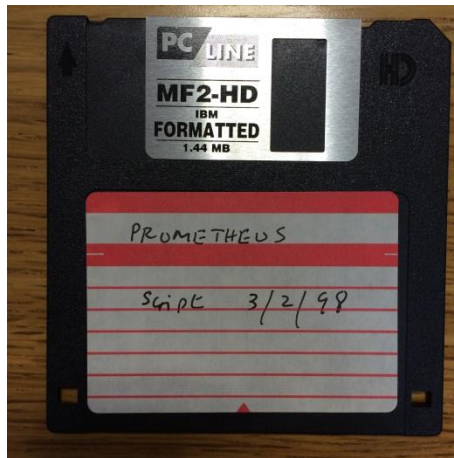
- Optical Media (DVDs, CDs, Minidiscs)

- Floppy disks (3.5" & 5.25")

- External Hard Drives (20GB-80GB)

- Folders on network drive

- Ongoing regular accruals of senate, court and council (university archive) minutes, exam papers, prospectuses and staff newsletters (as pdf/.doc(x)).

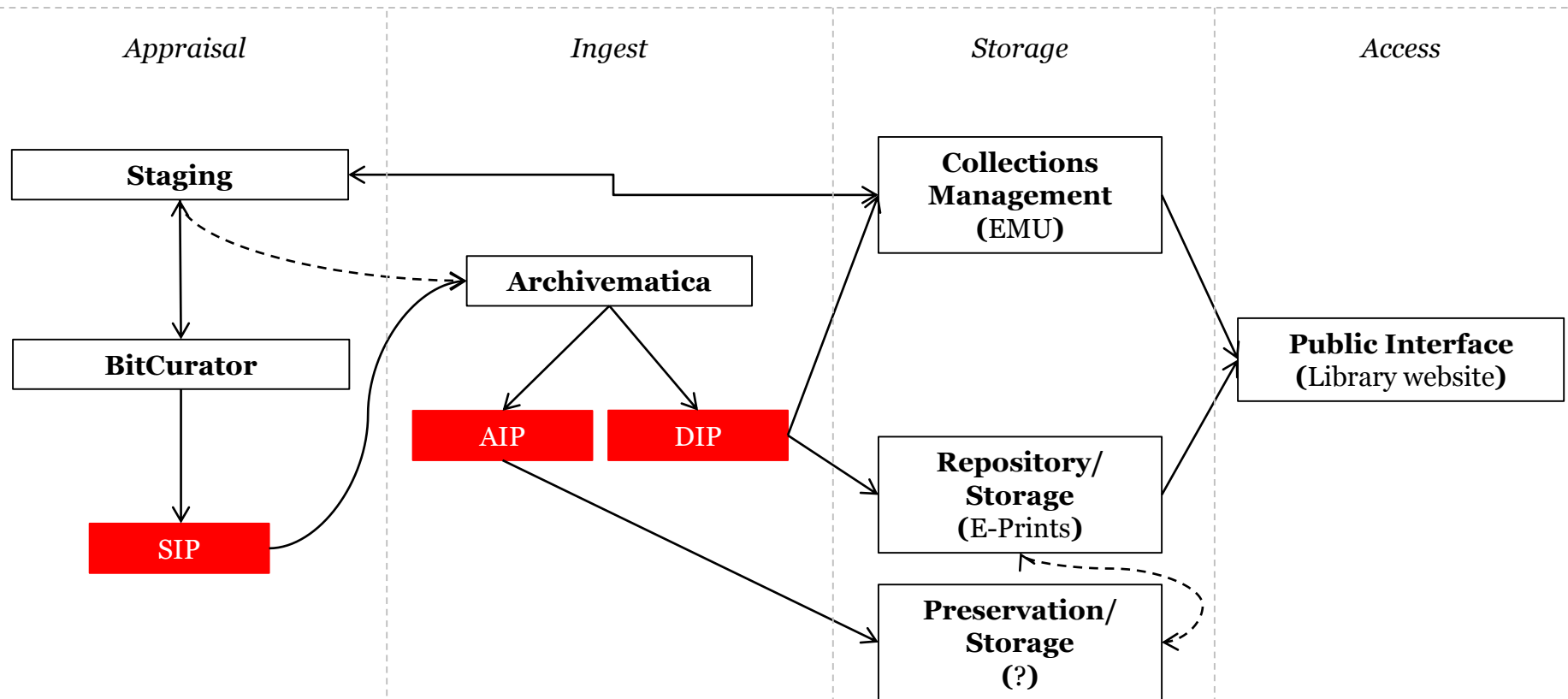
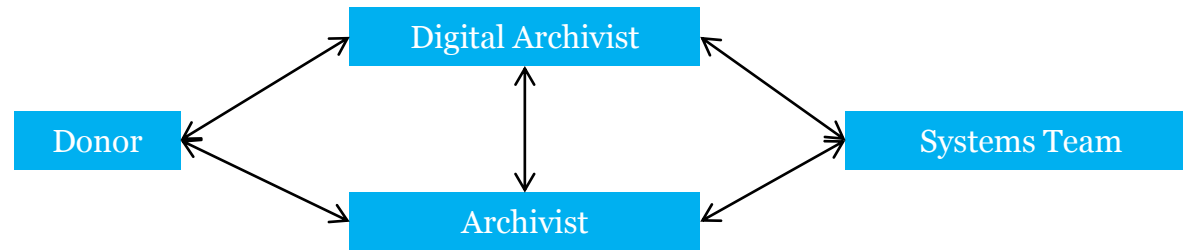


Immediate digital duplicate to
quarantine PC
naming convention to follow existing
material (lua_adm_010_343)

Name

- BC MS 20c Pybus_5_3_3
- BC_MS_20c_Crossley-Holland_5_3_4_6
- BC_MS_20c_Harrison_04_PRO_04
- BC_MS_20c_Heath-Stubbs
- BC_MS_20c_Silkin_6_1_29
- BC_MS_20c_Silkin_9_1_AC1
- BC_MS_20c_Silkin_9_2_AC1
- BC_MS_Cottingley_Fairies
- Holocaust Survivors External HD TESTING DATA(SAMPLE)
- LIDDLE_WW1_ADD_022
- LIDDLE_WW1_ADD_113
- LIDDLE_WW1_GS_0105_Baynham_CT
- LIDDLE_WW1_GS_0874_Juggins
- LIDDLE_WW1_GS_1824_Merrett_HC
- LRA_MS_1172_102a
- LRA_MS_1231_601
- LRA_MS_1452_31
- LRA_MS_1459_62
- LRA_MS_1488_3
- LRA_MS_1489_3
- LUA_DEP_BIO_1_1_Ida Roper Herbarium
- MS 1793_10
- MS 1938-384 Eurich Papers
- Quaker Digital Archives
- QUAKER POST BC
- SUR_46
- SUR_429 Dr Patrick Ainley Degrees of Difference

Leeds Workflow MODEL



We will be handling a range of content

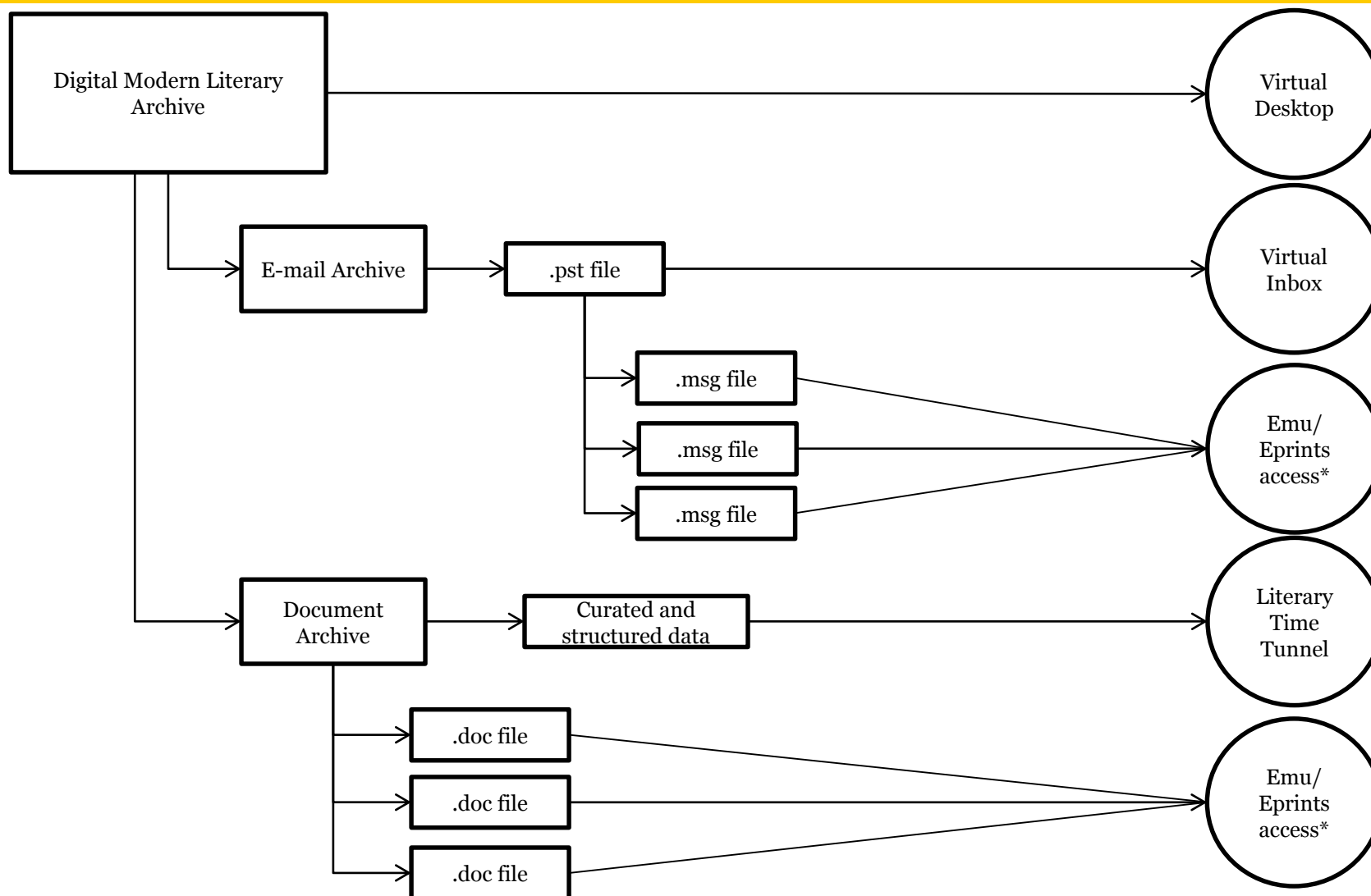
- Simple, archival stuff
- Self contained collections
- Very complicated deposits (modern literary archives)

So...

Development of a range of ingest models to handle this diversity.

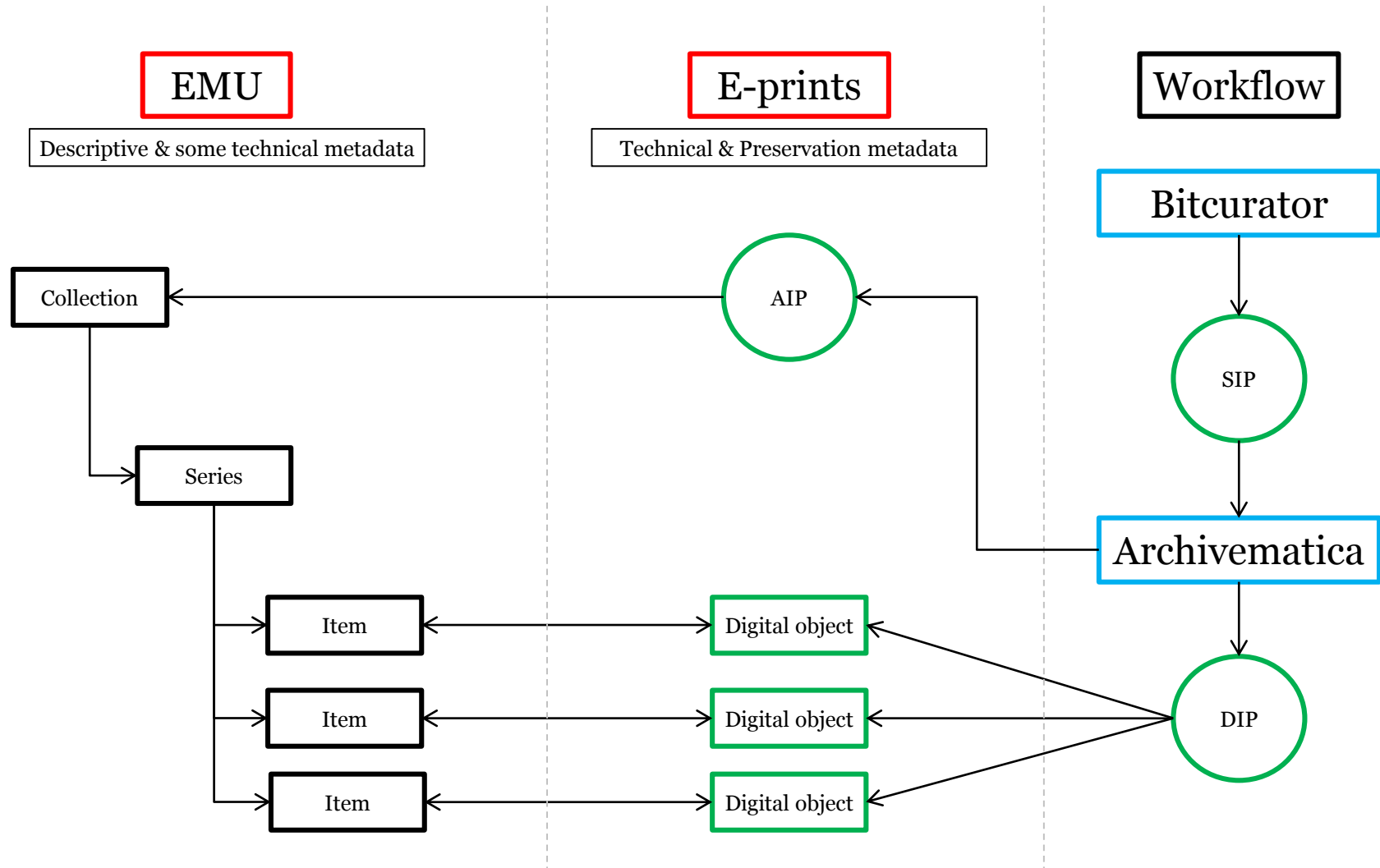
Ingest models

Model number	1	2	3	4
Provisional content type	Administrative/Organisational	Slightly more complicated	Research Collection (Quite rich data)	Creative Archives. A thoroughness – extensive, varied, research potential (digital humanities) Data mining ‘pocketed’ content.
Preservation methodology	None necessary (data already provided in a preservation format)	Migration into simple/preservation file formats	Emulation and migration	Emulation and migration
Structure/Hierarchy	Simple hierarchical file structure			Complex
Example file level content	Text of an email transcribed to a word processing document (.txt, .doc, .docx) or saved as a .pdf	.msg file	Folders of .msg files, in hierarchical folder structure	.pst file (providing comprehensive context and structure).
Complexity	Routine			Exciting/very complicated
Developmental possibilities?	Little or no value in developing a user resource			A lot of value in developing the resource
Cost to Acquire?	?	?	?	?
Content example	University Archive material	Quaker Collection	Holocaust Survivors Archive	Modern Literary/Artwork Archive/Collection



* Typical file - accessed through catalogue.

Bundled Content



Further Developments

Establishing policy for Donor interactions:

- Format policies
- Donor Questionnaires
- Donor/Deposit/Purchase agreements

Developing Procedures:

- Developing a fully functional ingestion workflow (!)
- Establishing metadata guidelines across content (alongside digitisation team)

Investigating Access:

- Researching tools and method
- Developing connections with schools and departments.

Ongoing Preservation:

- Investigate tools
- Implement policies