

It Takes a System to Curate a Bit: Preservation Planning at the Princeton University Archives

Jarrett M. Drake

Digital Archivist, Princeton University

ALA-ALCTS-PARS Preservation Metadata Interest Group

Las Vegas, Nevada

June 28, 2014



Princeton University
Library

Princeton University Archives

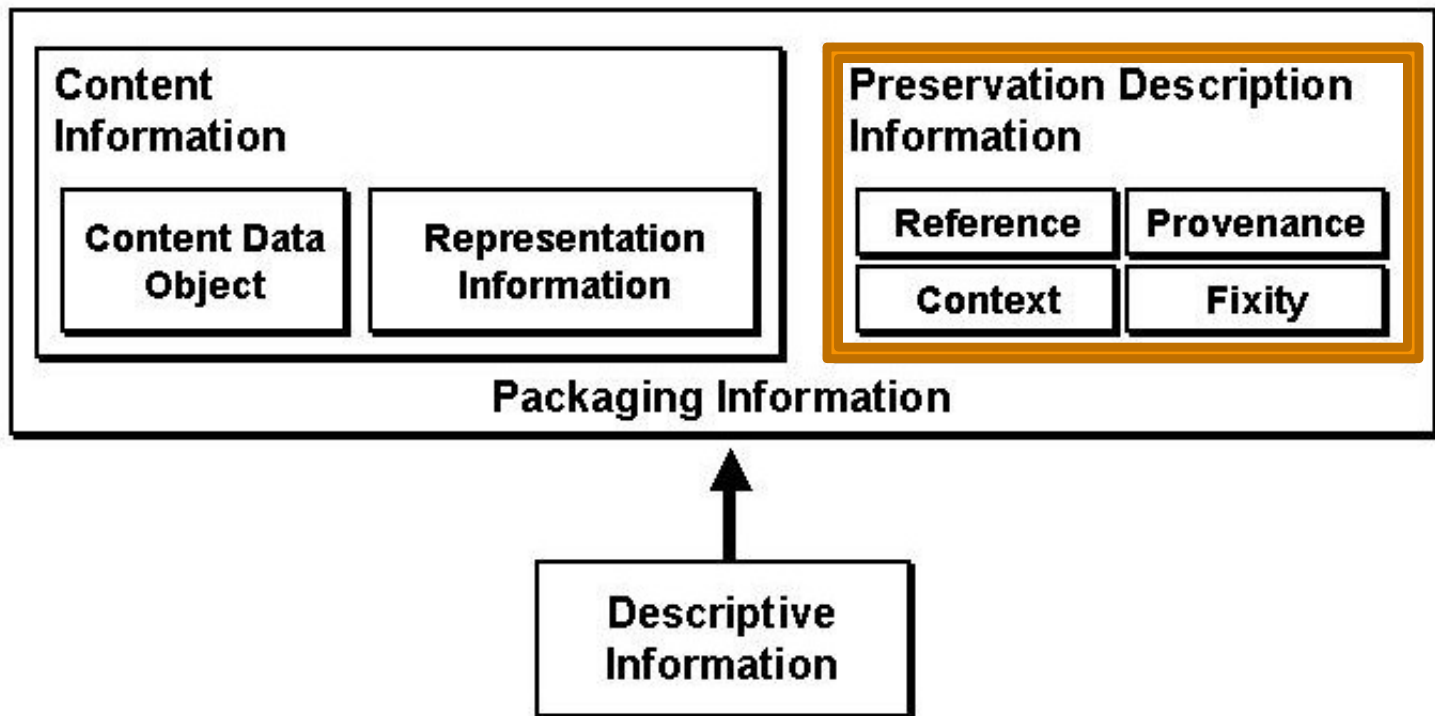
- Digital Curation Program est. in April of 2014
- ~100GB of 30,000 files in 20 collections

User Defined Text 4

2013-12-09: Virus scan performed on Disk001; no malware detected.

2013-12-11: Image of Disk001 created successfully. Directory printing of Disk001 created.

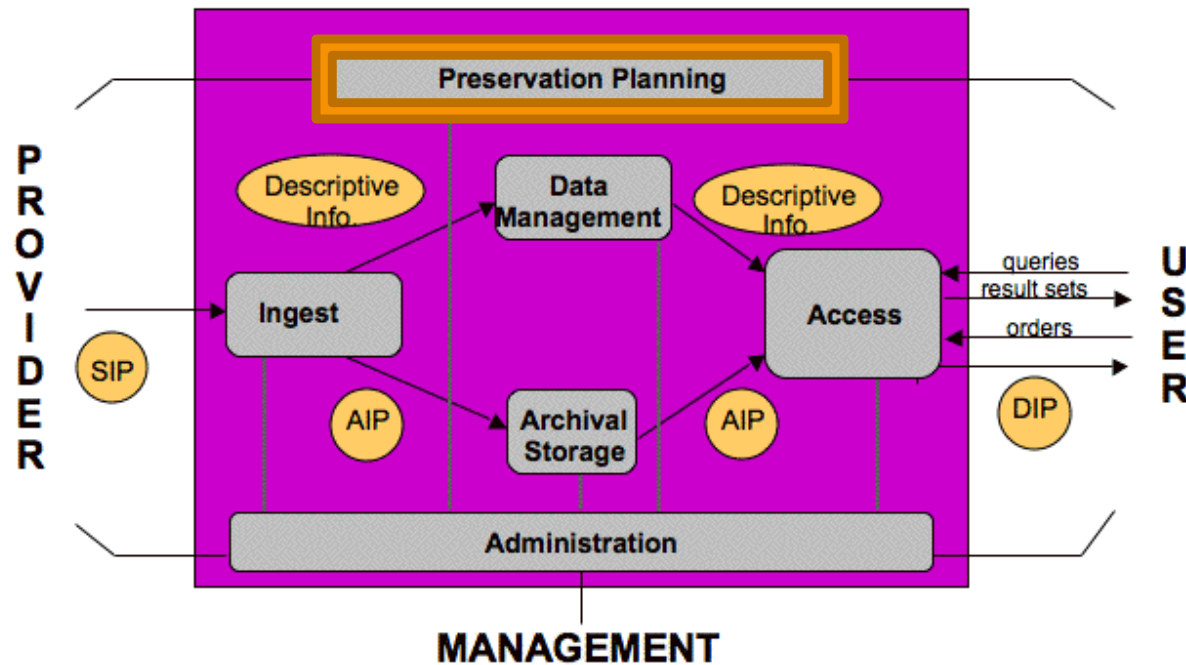
2013-12-12: Bag created on Lib-filesshare. Bag validated on Lib-filesshare.



"Preservation Metadata," 2nd Edition, Digital
Preservation Coalition

<http://dx.doi.org/10.7207/twr13-03>

Archival Functional Entities



SIP = Submission Information Package
AIP = Archival Information Package
DIP = Dissemination Information Package

<http://nssdc.gsfc.nasa.gov/nost/conf/archive21st/framework.html>

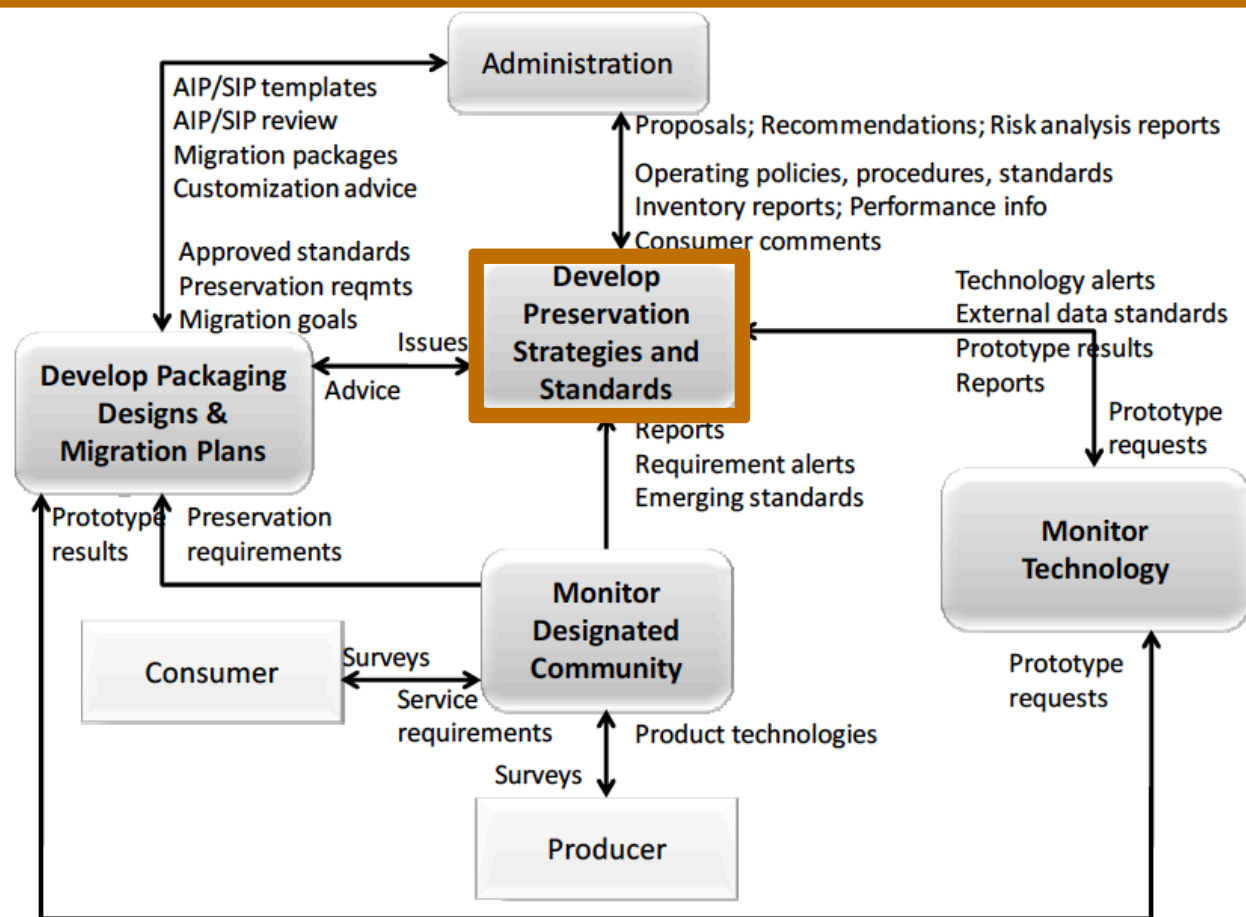


Figure 4-6: Functions of the Preservation Planning Functional Entity

<http://public.ccsds.org/publications/archive/650x0m2.pdf>

Example: North Carolina PMDO

	Element name	Obligation	Suggested Value Control
1	<u>Bit Depth</u>	Strongly recommended	Controlled vocabulary
2	<u>Checksum</u>	Strongly recommended	None (free text)
3	<u>Collection Source</u>	Required	Controlled vocabulary
4	<u>Color Space</u>	Optional	Controlled vocabulary
5	<u>Compression Degree</u>	Recommended, if applicable	None (free text)
6	<u>Compression Type</u>	Strongly recommended, if applicable	Controlled vocabulary
7	<u>Creation Hardware</u>	Strongly recommended (digitized); Optional (born-digital)	Controlled vocabulary
8	<u>Creation Software</u>	Recommended (digitized); Strongly recommended (born-digital)	Controlled vocabulary

<http://digitalpreservation.ncdcr.gov/PMDO.html>

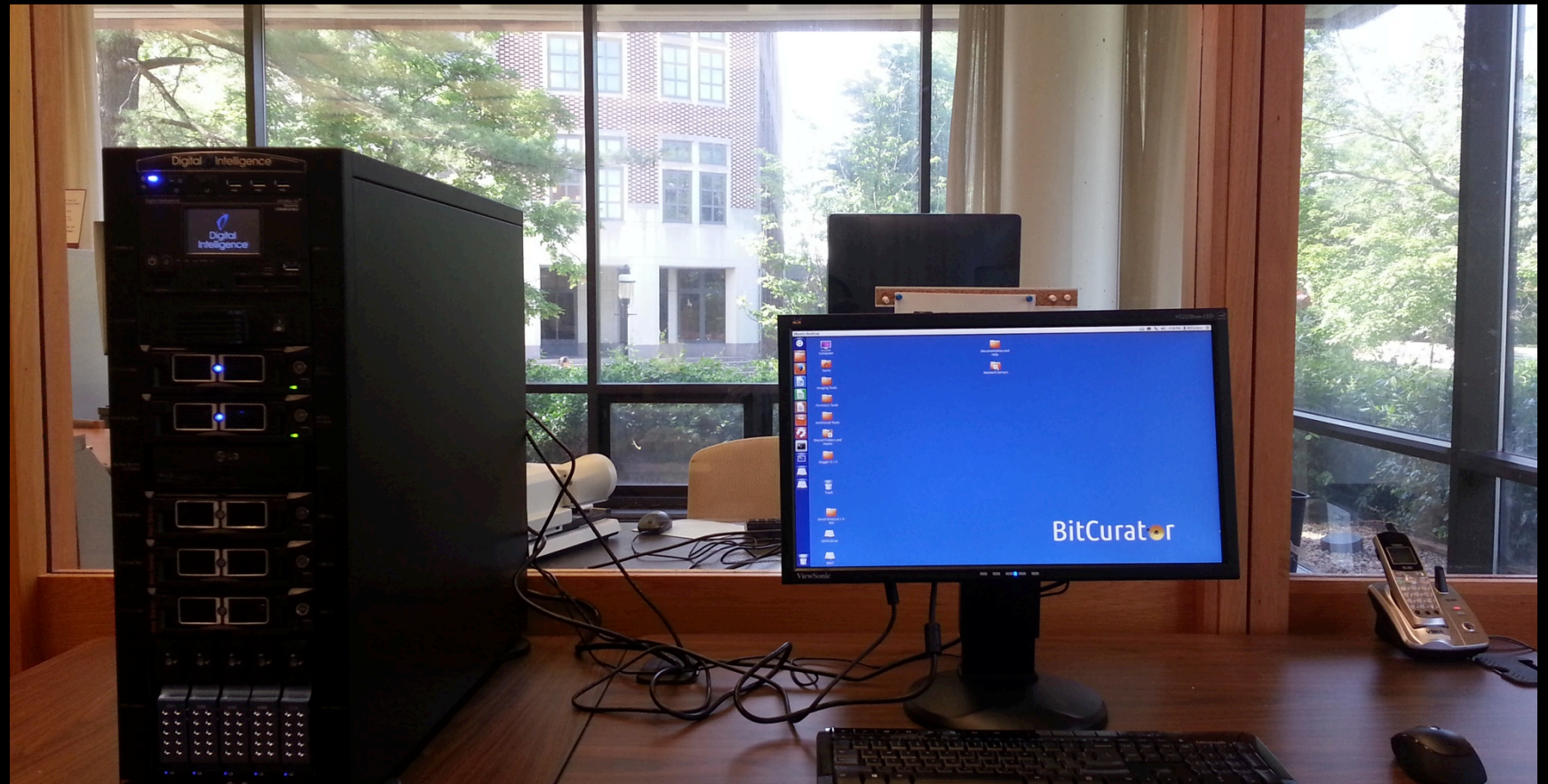
Current Practice

Guidance for Depositing Digital Records to the Princeton University Archives

Activities	Tier 1 (Format)	Tier 2 (Bit)
Conduct virus scan of media and maintain bitstream of uncorrupted files	•	•
Provide secure storage and backup in two separate locations	•	•
Produce, store, and validate checksums to assure file integrity and fixity	•	•
Generate preservation metadata to facilitate long-term management and access	•	•
Normalize file and perform active monitoring of trusted preservable formats	•	
Execute migration to successive formats at the point of format obsolescence	•	

http://www.princeton.edu/mudd/news/transf_donations/guidance.pdf

Future Practice



6/27/14

Princeton University Library

Native BitCurator Applications

fiwalk product of DFXML

```
<hashdigest type='md5'>cc93076a1c96ffd7083901473aec59c4</hashdigest>  
<hashdigest type='sha1'>3f8e16330b6aa614a87c2bc0956cfb96080d681e</hashdigest>
```

```
<object>  
  <objectIdentifier>  
    <objectIdentifierType>4ad51b3a-cc15-11e3-8d33-080027f8dfea</objectIdentifierType>  
    <objectIdentifierValue>/home/bcadmin/Desktop/sample.E01</objectIdentifierValue>  
  </objectIdentifier>  
</object>
```

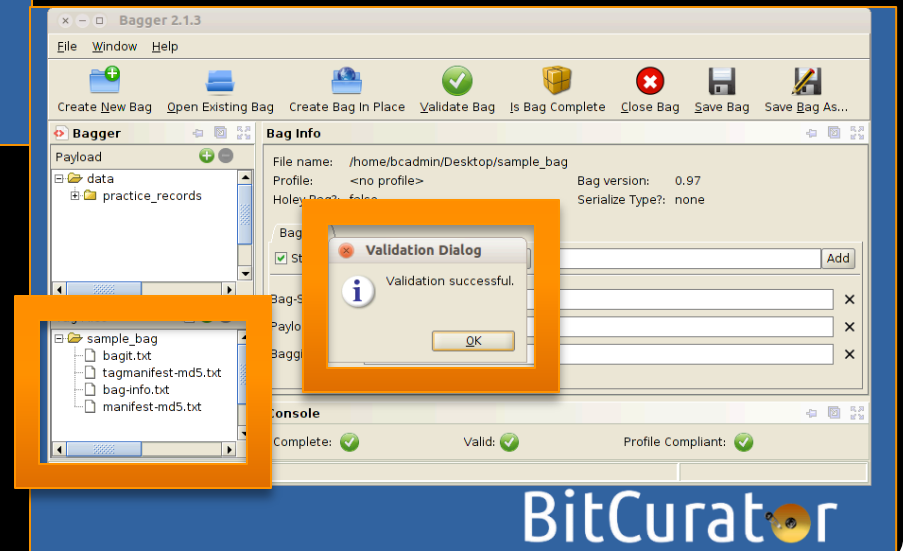
Hex editor view of a .vcf file

00000000	42	45	47	49	4E	3A	56	43	41	52	44	0D	0A	56	45	52	53	49	4F	4E	3A	33
00000016	2E	30	0D	0A	46	4E	3A	32	30	31	32	70	68	69	70	61	72	74	69	63	69	70
0000002C	61	6E	74	73	40	75	6D	69	63	68	2E	65	64	75	0D	0A	4E	3A	3B	32	30	31
00000042	32	70	68	69	70	61	72	74	69	63	69	70	61	6E	74	73	40	75	6D	69	63	68
00000058	2E	65	64	75	3B	3B	3B	0D	0A	45	4D	41	49	4C	3B	54	59	50	45	3D	49	4E
0000006E	54	45	52	4E	45	54	3A	32	30	31	32	70	68	69	70	61	72	74	69	63	69	70
00000084	61	6E	74	73	40	75	6D	69	63	68	2E	65	64	75	0D	0A	45	4E	44	3A	56	43

BEGIN:VCARD..VERSION:3
.0..FN:2012phiparticip
ants@umich.edu..N:;201
2phiparticipants@umich
.edu;;;..EMAIL;TYPE=IN
TERNET:2012phiparticip
ants@umich.edu..END:VC

Compatible Applications

Bagger validation of hash values



Opportunities

- Streamlined tools for processing non-media based digital accessions
- It takes a system to curate a bit, and it takes a community to curate a system!