



Documentation of Born-Digital Archival Collections In Centralized Processing Environments

CurateGear 2016,
Chapel Hill, North Carolina

Susan Malsbury
The New York Public Library

Donald R. Mennerich
NYU Libraries



Why Centralize?

Efficient Collection Management

Consistent Accessioning, Arrangement & Description

Project planning and resource allocation

Curatorial staff are free to focus on curatorial work

Rodgers
and
Hammerstein
Archive of
Record
Sound

Jerome
Robbins
Dance
Division

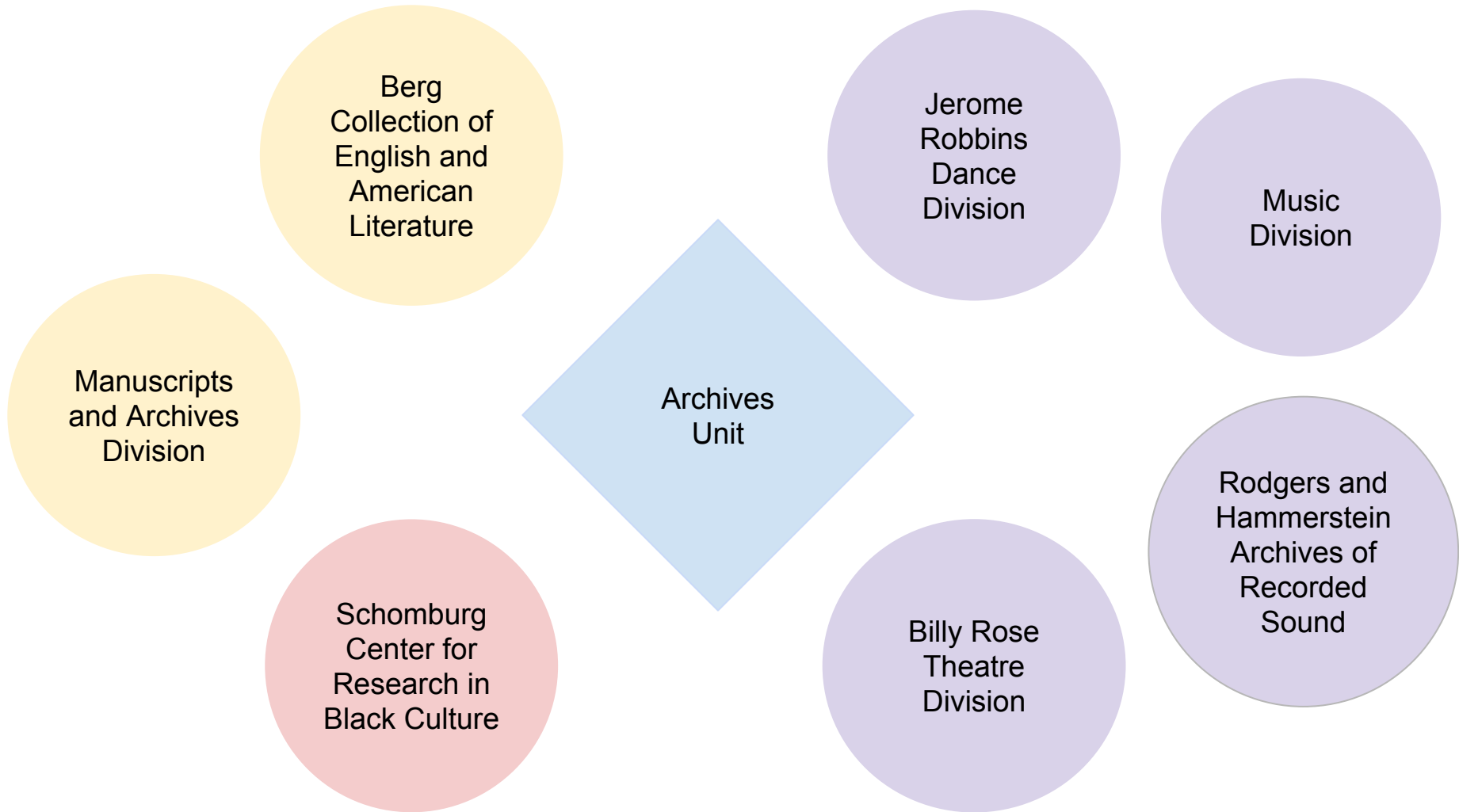
Berg
Collection of
English and
American
Literature

Music
Division

Manuscripts
and
Archives

Schomburg
Center for
Research in
Black Culture

Billy Rose
Theatre
Division



University
Archives

Archival
Collection
Management

Tamiment
Wagner

Fales
Library &
Special
Collections

Education and outreach

Curators

Reference staff

Processing archivists

Project Planning

Separate processing queues / balancing expectations

Resource allocation

Increased R & D

Preservation services

Gaps between acquisition and accessibility

Collection and Institutional Knowledge

Born-digital is a an emerging problem space in archives

Understanding collections is key to obtaining processing resources

Establishing the provenance of digital materials is critical to establishing and maintaining the integrity and authenticity of records

Donor Relations

Transparency in operations is key in building trust in the repository

Documentation Examples

Transfer/Submission Agreements

What is the collecting repository acquiring?

What are the formats?

How much?

How often?

1. Transfer Agreements - PAIMAS

Project context and Information objects

- What are the objects
- What is the PDI
- What description exists

Obligations

- Legal ownership
- Rights to preservation
- License agreement

Transfer conditions

- SIP structure
- Specify data transfer session
- Specify data objects
- Test transfer

Validation

- Systematic (checksums)
- In-depth (intelligibility)

1. Information to be transferred
2. Data objects and standards
3. Object references
4. Quantification
5. Security conditions
6. Legal and contractual conditions
7. Transfer operations
8. Validation
9. Timetable
10. Impact on repository
11. Critical points

Curatorial Resources / Help Docs

General Considerations:

- Attempt to approach born-digital materials in the same way as other archival materials. What we do doesn't change; how we go about it may.
- Make it regular practice to ask donors whether their collection includes digital material they maintain or digital materials stored in cloud services and websites they have created or manage, as these are often overlooked by donors unless the donation predominantly consists of born-digital materials.
- Clarify expectations for byte-level preservation and local access, explaining the limitations of certain technologies and the capabilities of our program.
- Consider asking for a small sample of files for testing purposes to identify technical or administrative issues prior to donation.

Digital Material Creation:

- Are you solely responsible for creating your files? If not, who else is and what is their role?
- Is the collection solely born-digital, or a mixture of born-digital, paper, and other formats?
- Is material unique or have copies been kept in another format (e.g. print-outs)?
- Has physical material from the collection been digitized or duplicated in the digital records; if so what and how much?

Curatorial Resources/ Help Docs

Types of Digital Material:

- What types of content do you create (e.g. research notes, original writing, correspondence, diaries, artwork)?
- What types of materials are present: photographs, office documents (word processing, spreadsheets), harvested web content, email messages, static data sets, dynamic data sets, digital art, digital media, presentations (Powerpoint), and other electronic records?
- What [file formats](#) are present? Word, Excel, Rich Text, Plain Text, PDF, TIFF, BMP, JPEG, MOV, MPEG, WAV, MP3, MP4, HTML, XML are some common file formats.
- Do you create records using specific applications or software programs (e.g. Photoshop)?

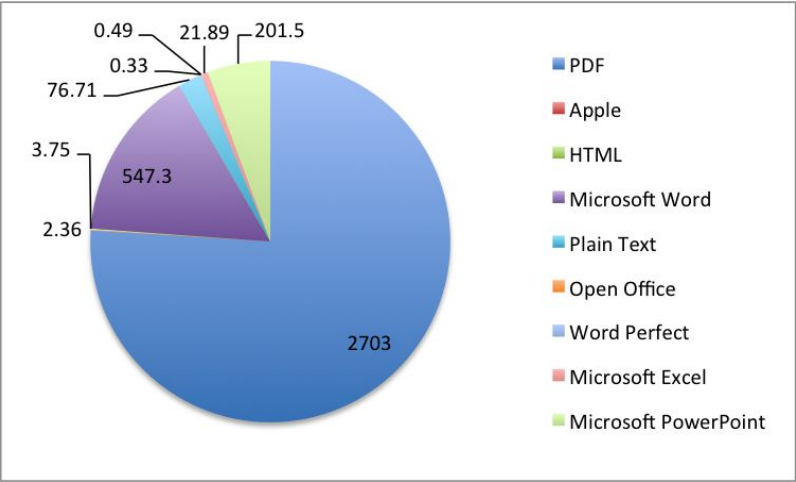
Organization:

- How are the materials ordered? Is there any organizational structure (folders, hierarchies)?
- How are files named?
- Is some kind of version control used (e.g. filename1, filename2 for 1st, 2nd draft)?

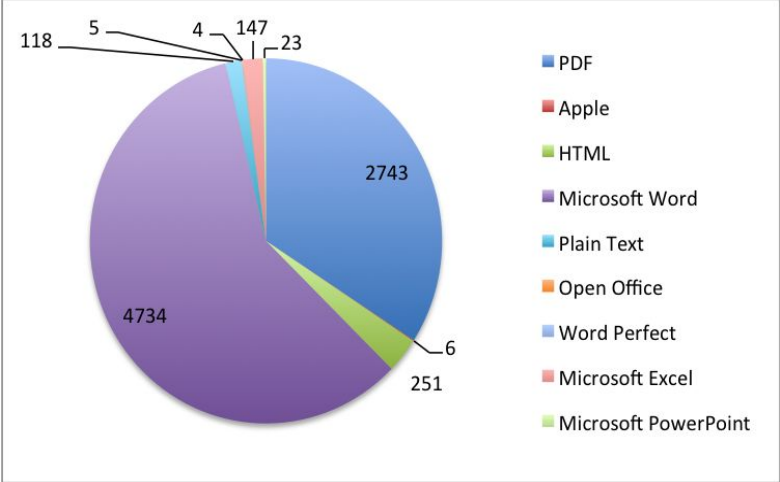
Collection Summaries

Documents

Size (MB)




Count



Collection Summaries

Access, Use, and Appraisal Considerations:

Intellectual property issues	Email messages from people other than donor
Sensitivity concerns	Financial documents such as tax records
Possible out-of-scope materials	Real-estate listings represent approximately 25% of email messages
Blank, damaged or unreadable materials	3 floppy disks were unreadable; 5 CD-Rs were blank. Recommendation to discard (carriers do not have significant markings on them).
Duplicative material	Multiple copies (same version) of 60 videos (i.e. 180 files for 60 performances)
Access considerations	Protocols for access to email messages still to be determined
Technological limitations	Not able to preserve revision history of Google Docs 

Collection Summaries

Material Transferred:

Object Type	Count	Size	Note
Laptop Computer	1	128 Gigabytes	1st Gen (2008) Macbook Air, Serial #W8811R87Y51. 128GB HDD

Acquisition Methodology

The acquisition of the computer presented several problems. While it was possible to boot the machine to a forensic distribution via external usb cdr, this version of the Macbook Air only a

Object Type	Count	Size	Note
CD-Rs	50	375 gigabytes	

Acquisition Methodology

ISO disk images were created for all optical disks in FTK imager and ingested into FTK

Collection Summaries

Imaging Notes

Note from imager to processing archivist/curators to assist with processing and provide technical context for the collection. Include date span when known.

The disks came wrapped in papers that contained detailed inventories as to their contents. The disks were used with an AT&T word processor and are likely a CP/M file system. These inventories were scanned and added to the electronic dossier. The inventories are listed by disk name with is recorded in the "Label" field in CMS.

Floppy disks were imaged using the KryoFlux. A valid sector image was created for [REDACTED]-0005, [REDACTED]-0006, and [REDACTED]-0015. The imager could not determine the sector format of the other disks. KryoFlux stream files were created for each disk. Valid sector images were loaded in FTK for analysis.

Collection Summaries

Problem Media

Item number	Notes
██████-0011	Disk casing and the magnetic disk visibly dented. All bad sectors.
██████-0016	Some bad sectors. Invalid Block Number Error. Cannot be read in CiderPress. Files copied out of MAME. Directory list created with MAME imgtool.
██████-0041	“
██████-0044	“
██████-0061	“
██████-0019	Some bad sectors. Invalid disk block. Cannot be read in CiderPress. directory list created in imgtool.

Conclusions

Transparency

Efficiency and Brevity

Multiple Use / Repurposability