

Reconstructing the Digital Preservation & Archiving Practices of a Conservator: The Paul N. Banks Collection

Lorraine Dong, Elspeth Healey, Samantha Mueller, Sarah Ticer
August 2009

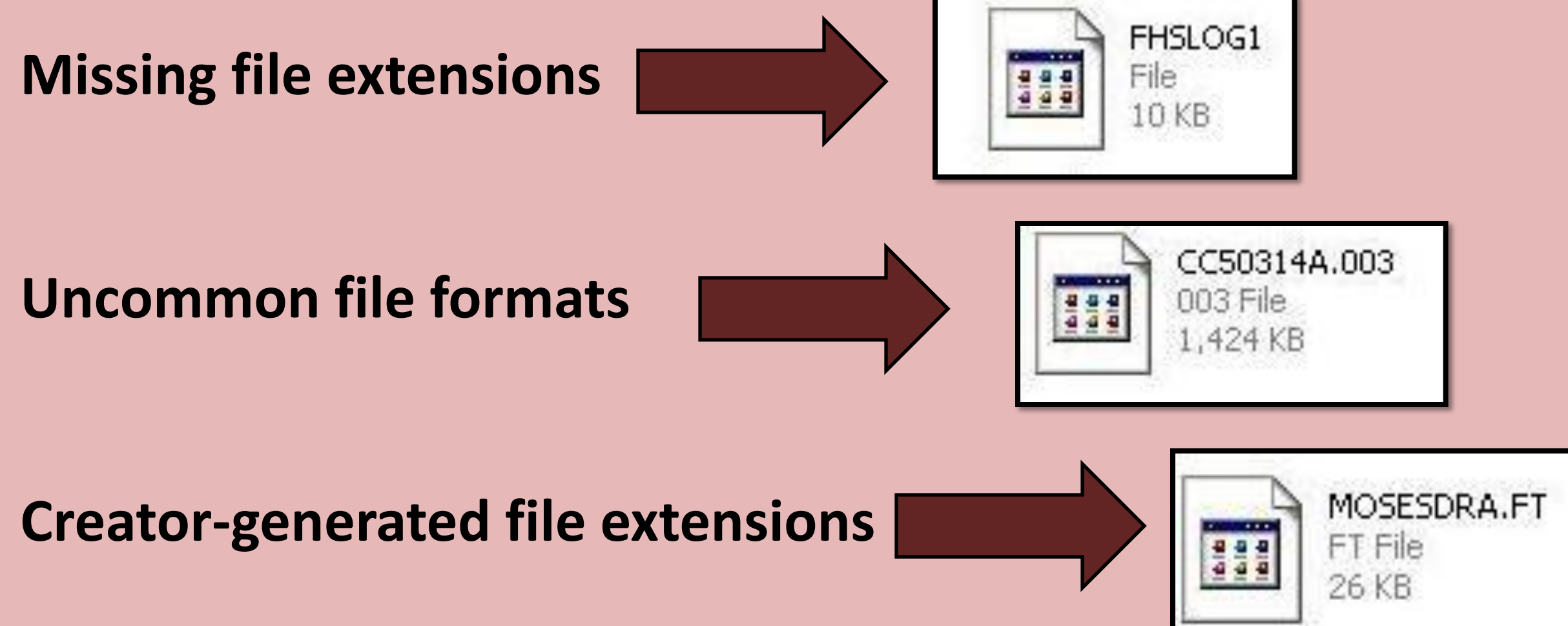
“Electronic, digital, and magnetic information will not survive passively, as librarians and archivists are used to assuming for older media; neglect can no longer be benign.”
- Banks, 2000

Who: Paul N. Banks (1934-2000) was a pioneer in the book and paper conservation field. His broad interests in the preservation of information included learning about and experimenting with new electronic media, as seen in his own personal digital archiving practices beginning circa 1983.

The Project: While concurrently migrating Banks’ electronic files to a digital repository and processing his paper collection, we became familiar with his digital preservation philosophy through both his practices and writings. Here we present the preservation obstacles that can arise even with a creator who was acutely aware of such issues, and our strategies to maintain and interpret Banks’ idiosyncratic digital practices.

Gaining intellectual control over artifacts with unusual file extensions

“There are three interrelated aspects to system instability: proliferation, obsolescence, and non-interchangeability.”
- Banks, 2000

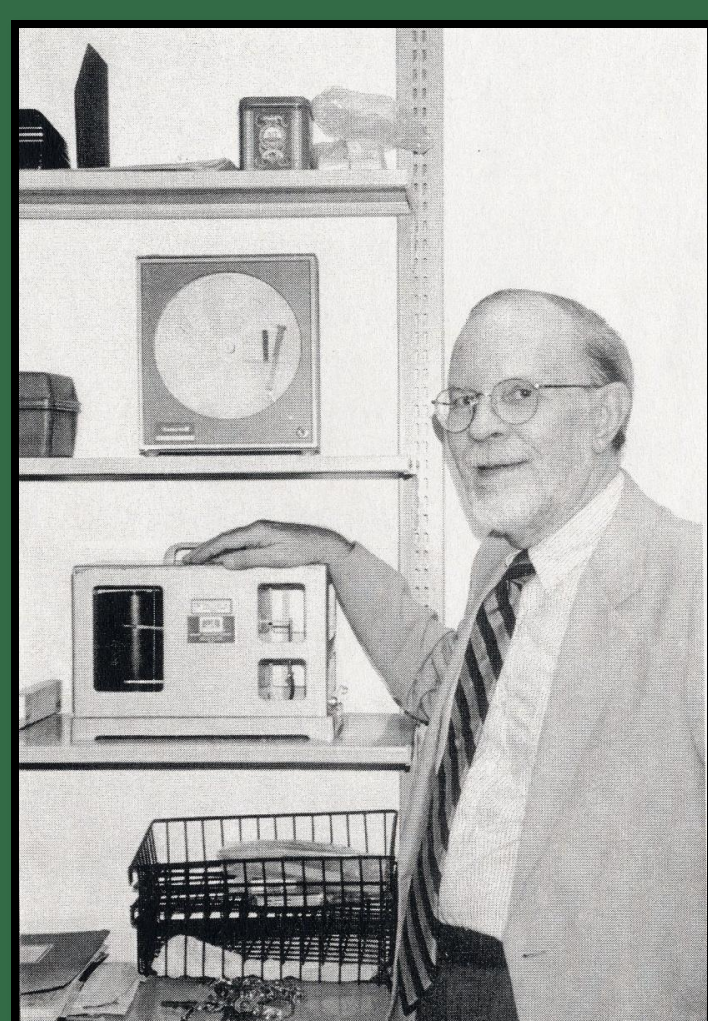
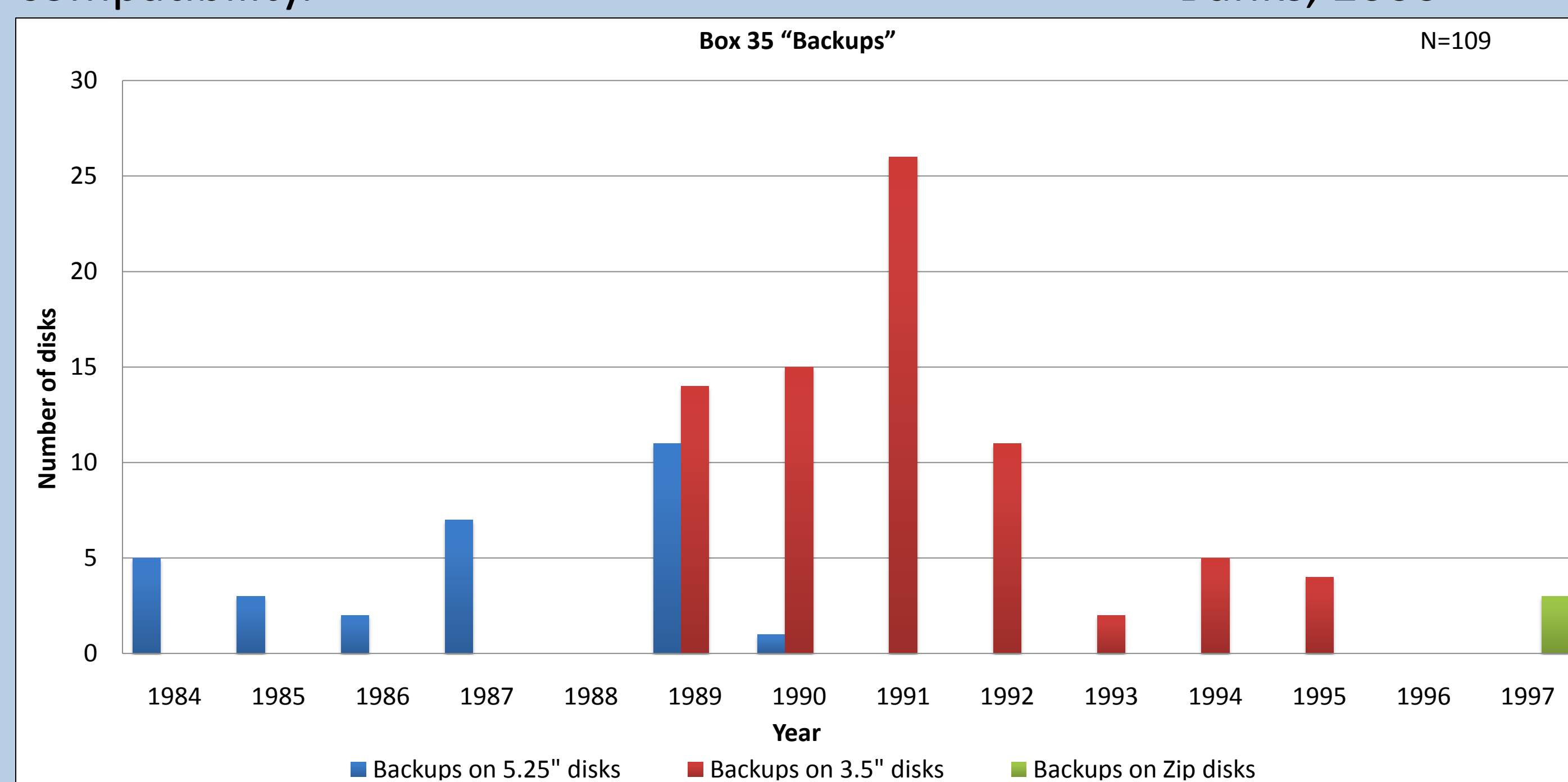


Strategies:

- Use metadata extractors to identify the programs associated with the more common file formats
- Use a hex editor to examine a file’s header information in order to identify the program that created the file
- Generate better file format registries to aid in distinguishing between uncommon file formats and creator-generated file extensions

Tracking refreshment activity

“Operating systems, application software, and even physical storage media evolve with stunning rapidity, and with only limited backward compatibility.”
- Banks, 2000

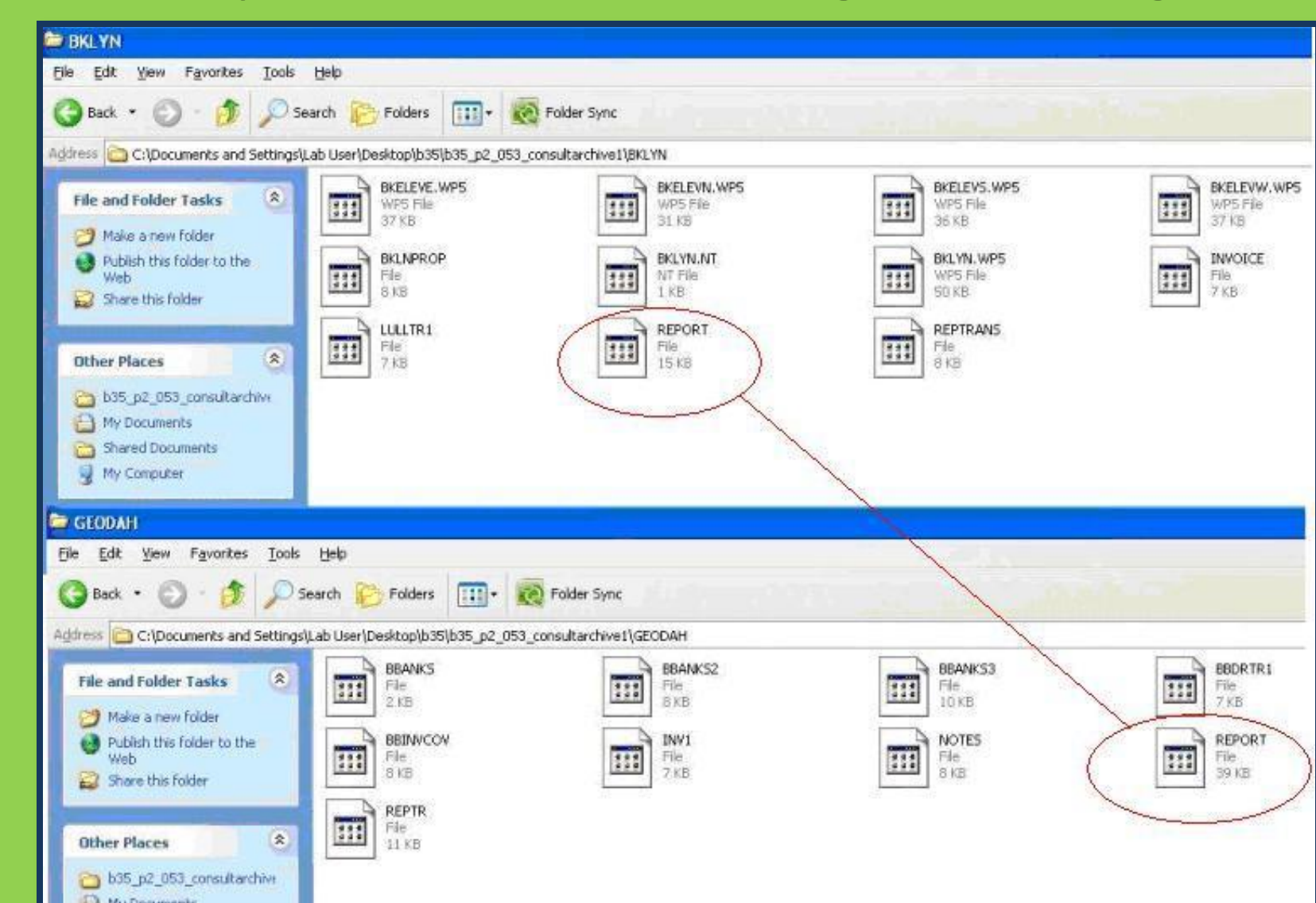


To learn more about Banks and our project, go to <https://pacer.ischool.utexas.edu/handle/2081/10682>

Preserving original folder structures

“[With migration,] there is always some alteration of the information in the process”
- Banks, 2000

Ingesting into Dspace required flattening Banks’ folder structure. Identically-named files in different folders within the same parent directory and the newly created metadata xml files were at risk of being inadvertently overwritten during flattening.

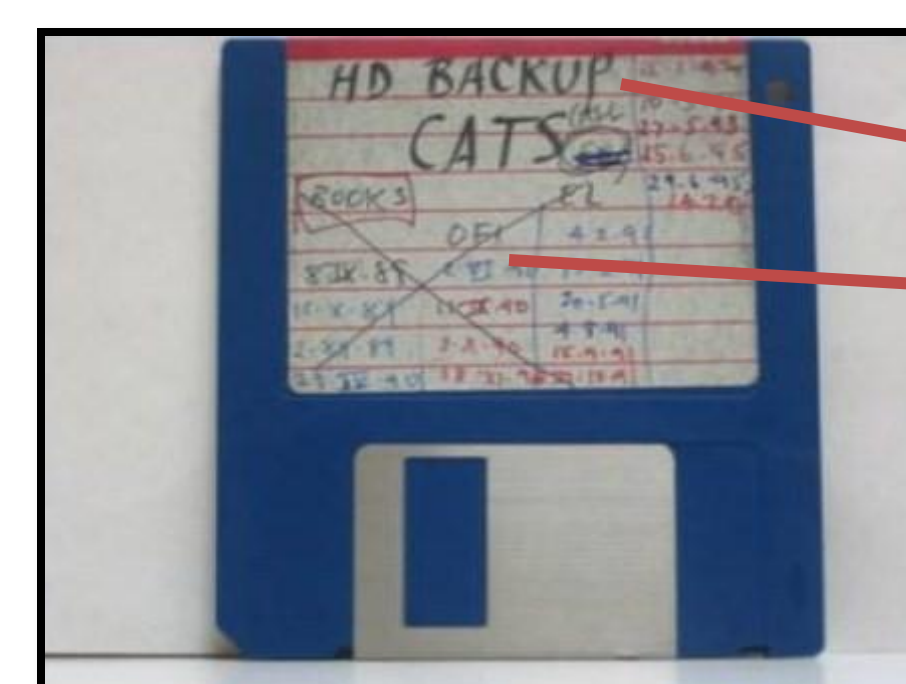


Strategies:

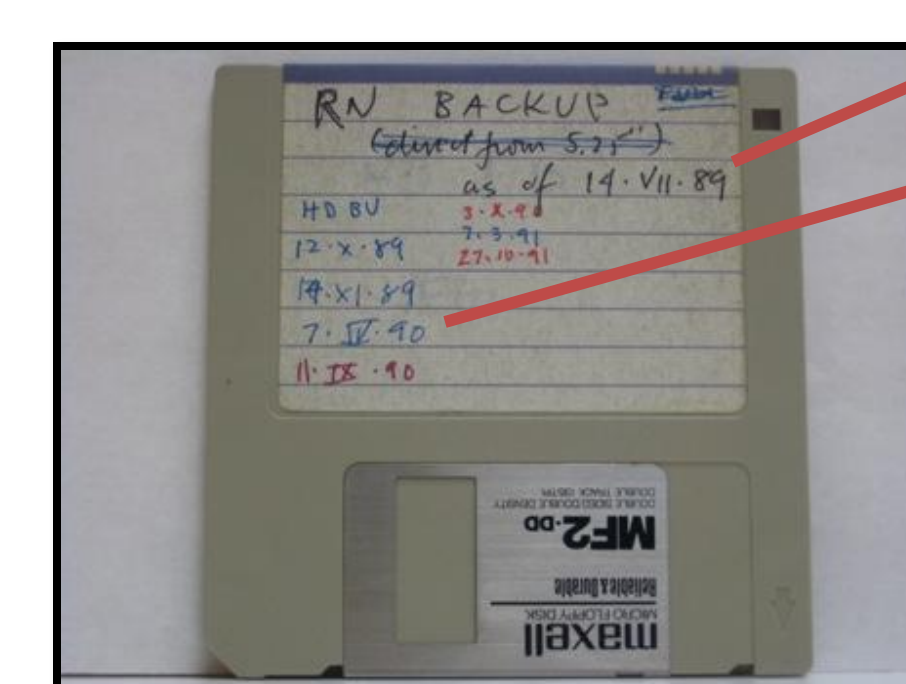
- Preserve a record of the creator’s folder structure using a file cataloging tool
- Identify all identically-named bitstreams within separate folders under the same parent directory
- Run metadata extraction tools separately on the identically-named bitstreams, using temporary folders when necessary to keep the bitstreams and xml metadata files distinct during the process

Identifying creator practices through physical media

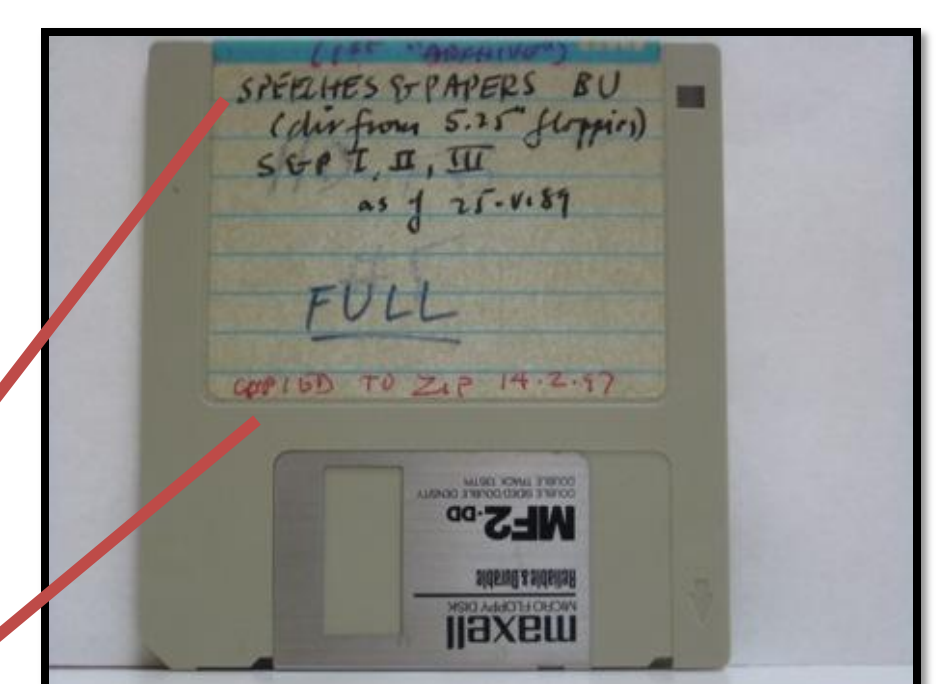
“There is always information in the artifact.”
- Banks, 1998



Possible title and contents
Frequent disk reuse



Banks’ idiosyncratic dating conventions
Periodic backup



Connections to paper collection
Refreshing to newer media