Computer Forensics and Born-Digital Content
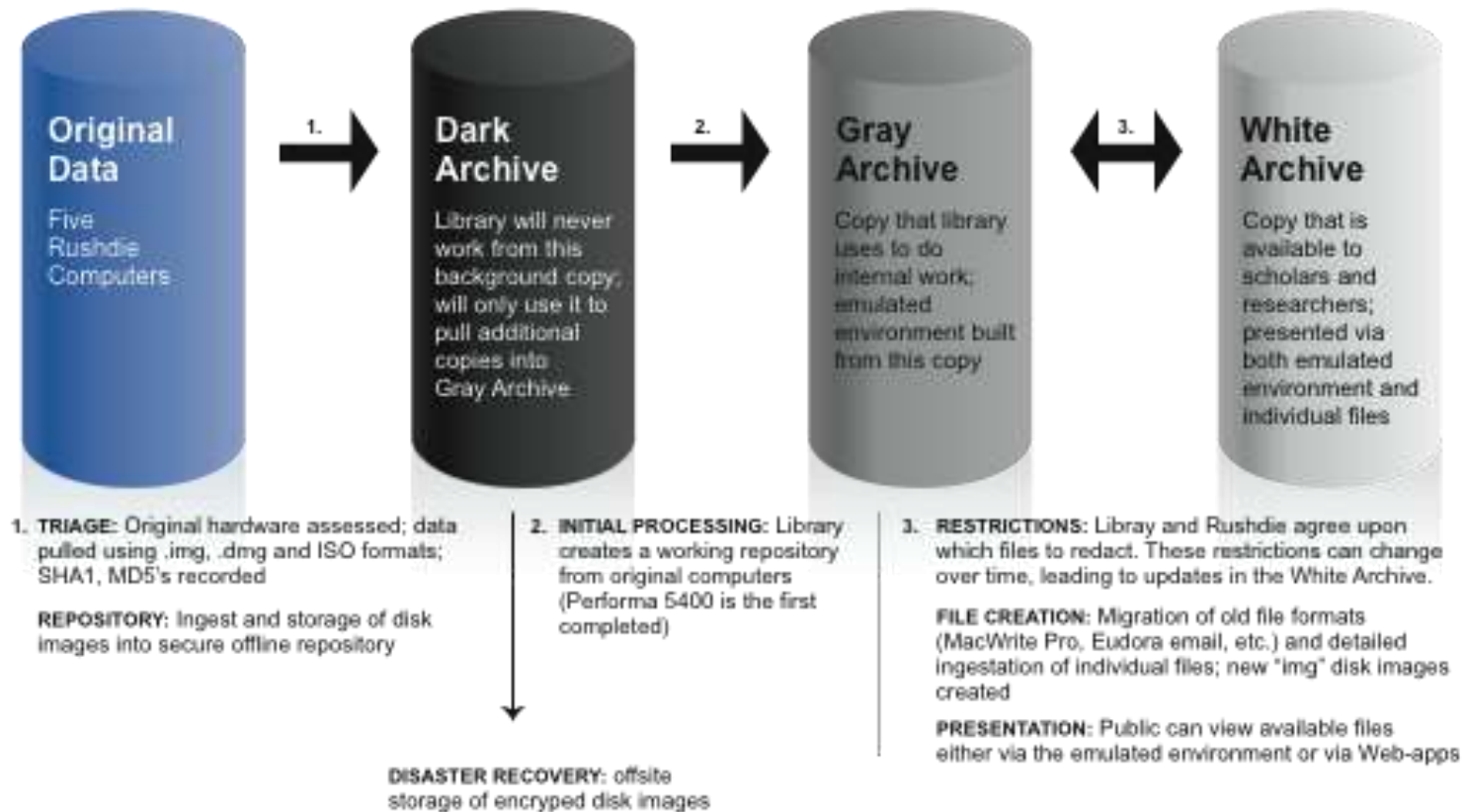in Cultural Heritage Collections

Forensics and
The Salman Rushdie Papers

Peter Hornsby, Software Engineer, Emory University Libraries

The Rushdie content model

Content Model: stages the born-digital data must go through before being made available to the public

Original Data

Five Rushdie Computers

1.

Dark Archive

Library will never work from this background copy; will only use it to pull additional copies into Gray Archive

2.

Gray Archive

Copy that library uses to do internal work; emulated environment built from this copy

3.

White Archive

Copy that is available to scholars and researchers; presented via both emulated environment and individual files

1. **TRIAGE**: Original hardware assessed; data pulled using .img, .dmg and ISO formats; SHA1, MD5's recorded

   REPOSITORY: Ingest and storage of disk images into secure offline repository

2. **INITIAL PROCESSING**: Library creates a working repository from original computers (Performa 5400 is the first completed)

3. **RESTRICTIONS**: Library and Rushdie agree upon which files to redact. These restrictions can change over time, leading to updates in the White Archive.

   FILE CREATION: Migration of old file formats (MacWrite Pro, Eudora email, etc.) and detailed ingestation of individual files; new "img" disk images created

   PRESENTATION: Public can view available files either via the emulated environment or via Web-apps

DISASTER RECOVERY: offsite storage of encryped disk images

Triage is …

Hardware assessment

Software evaluation

Storage device analysis

Data acquisition

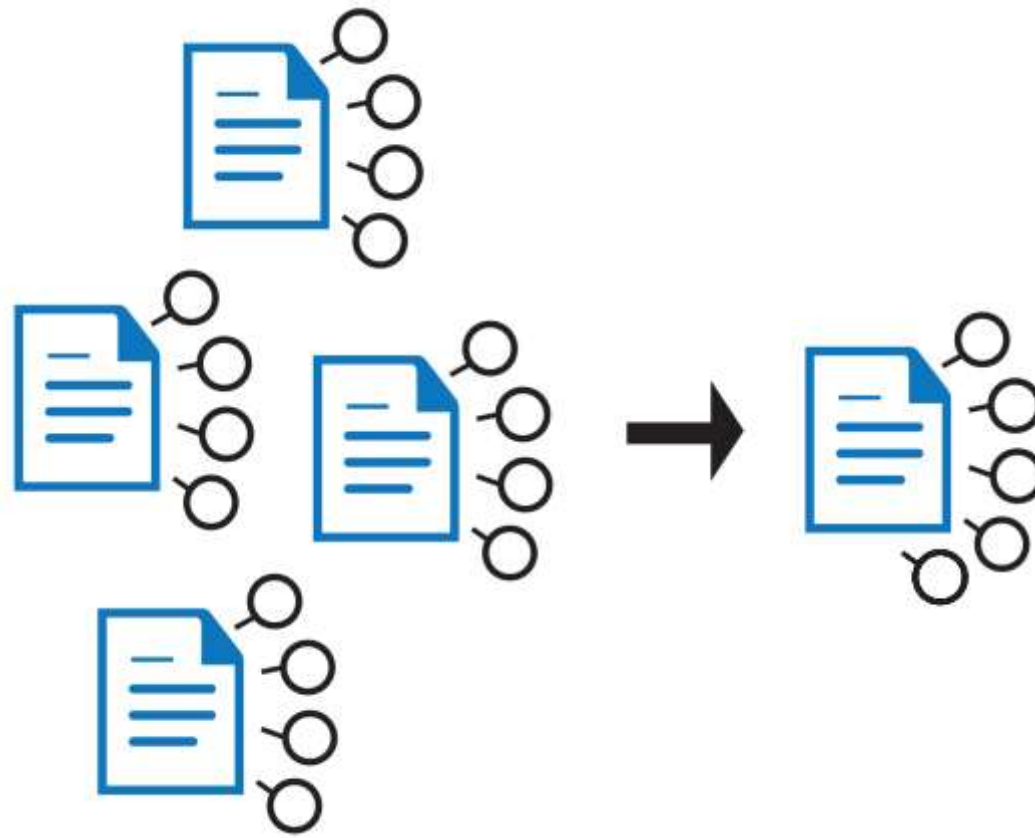# Computer forensic methods

Disk imaging

Hashes

Metadata

Files and checksums

Future gains

Frees software engineers' time

Safe data capture
& image creation

Quick, more cost-effective
discovery process