Identification and Collection

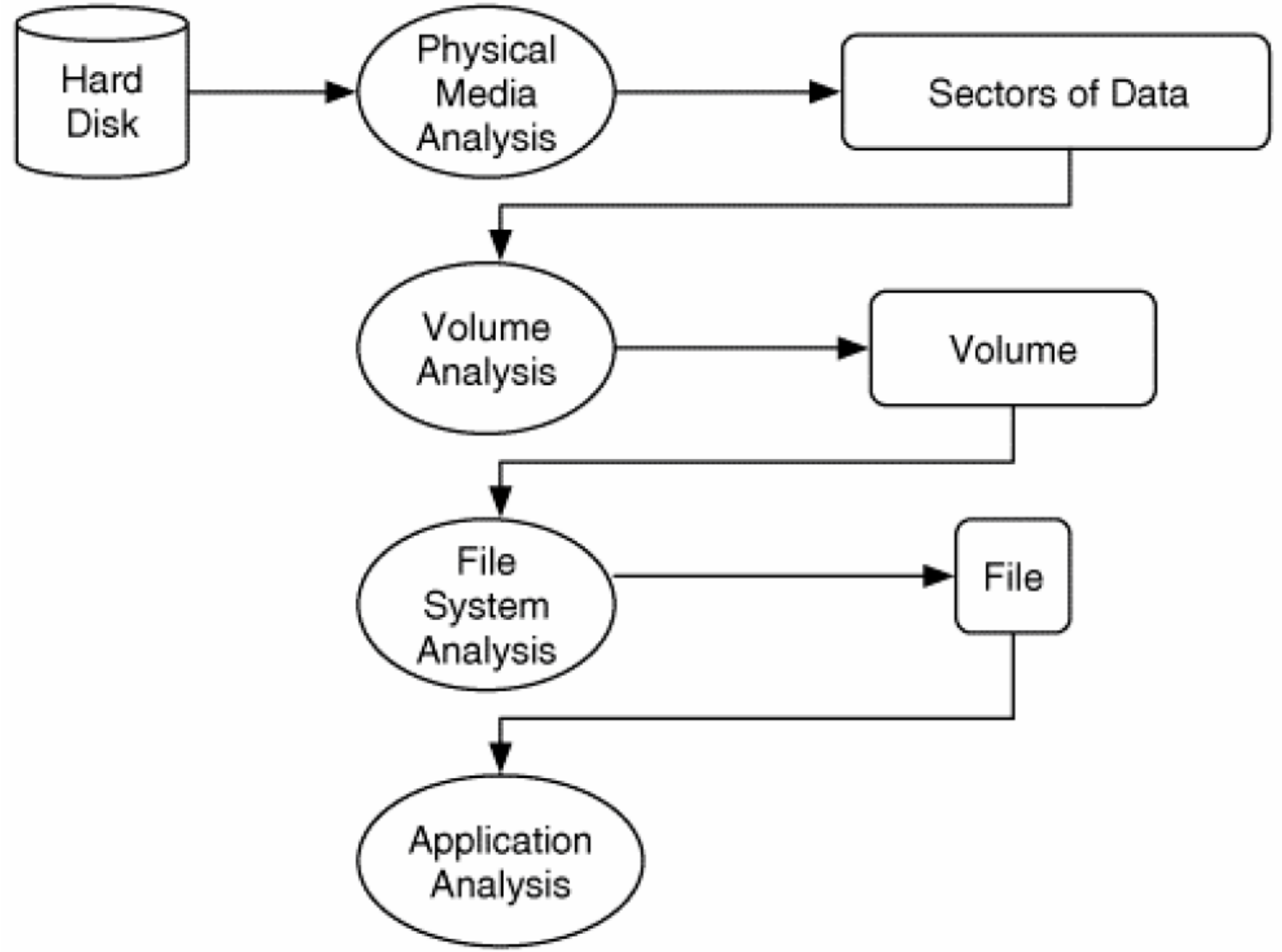
INFM 718X/LBSC 708X

Douglas W. Oard

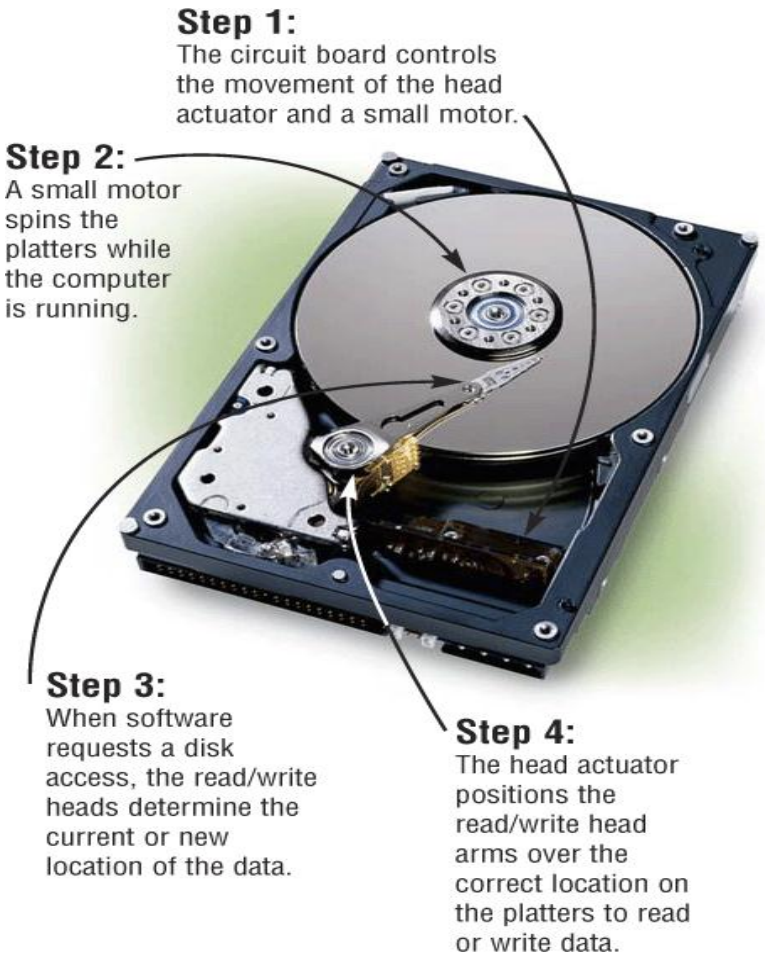
“Data” Mapping

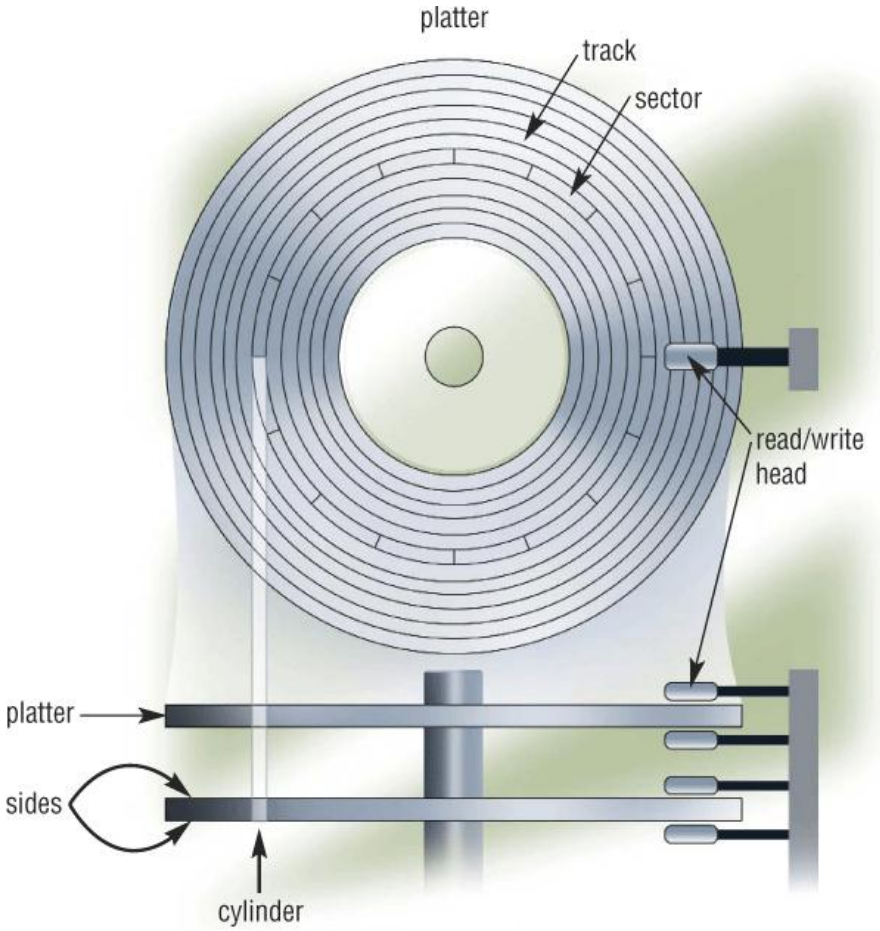
- Organizational
- Application-al
- Logical
- Physical
- Geographic

Levels of Analysis



How Disks Work





Extracted From Shelly Cashman Vermatt's Discovering Computers 2004

Windows “NTFS” File Metadata

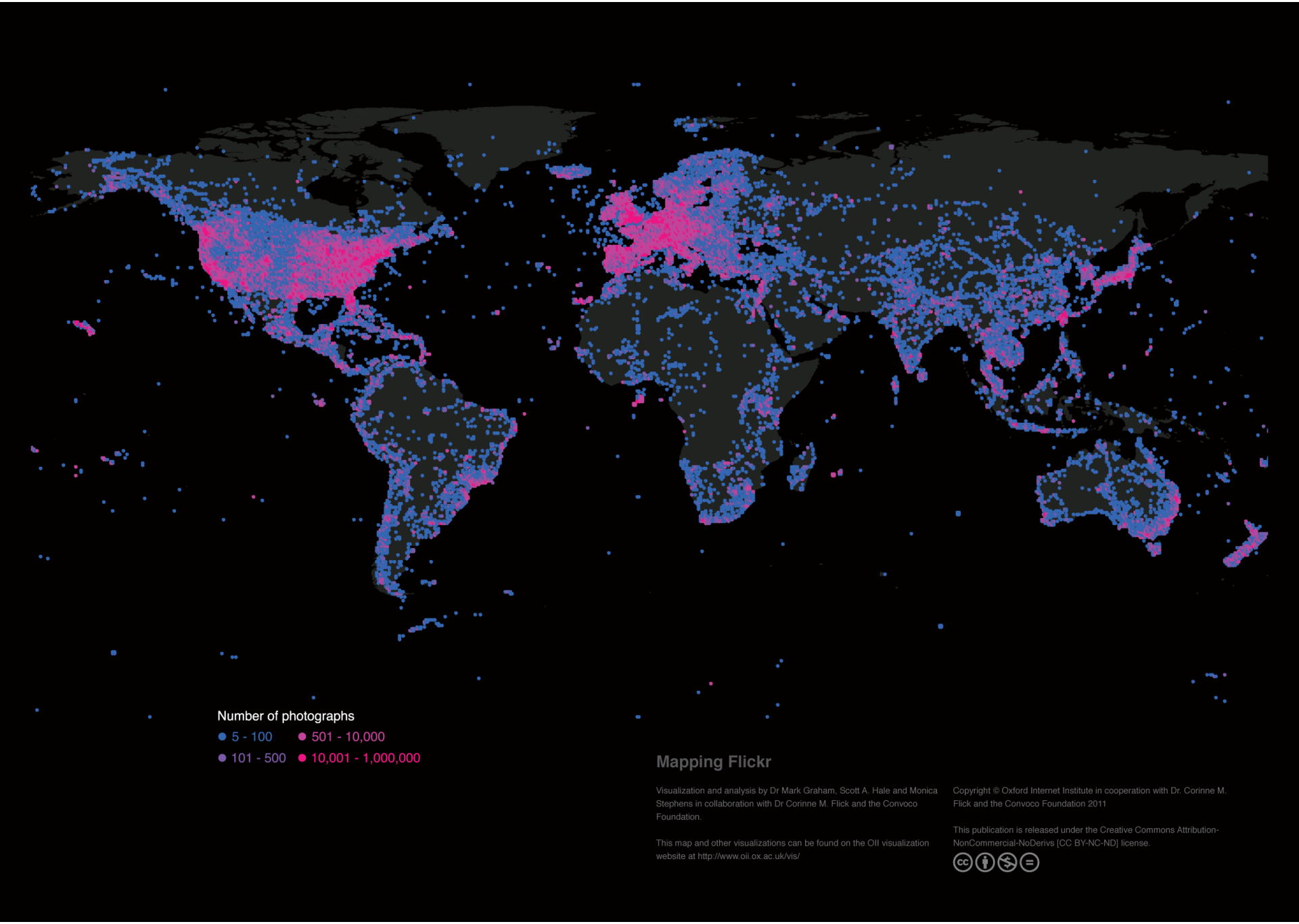
- Time file created (or copied)
  - Most recent one; optionally “journaled”
- Time file content changed (or made changeable)
  - Most recent one; optionally “journaled”
- Time file renamed (or moved)
  - Most recent one
- Time file metadata created or changed
  - Most recent one
- Time file accessed (content or metadata)
  - Most recent one; optionally disabled

Microsoft Word Metadata

- Author
- Title
- Dates (may not agree with NTFS!)
  - Created
  - Modified
  - Accessed
  - Printed
  - Each tracked change

EXIF Image Metadata

- Time
- Location
- Camera manufacturer and model
- Camera orientation
- Exposure information (shutter speed, f stop)
- Thumbnail versions
  - Altering the image may not change the thumbnail!



Mapping Flickr

Visualization and analysis by Dr Mark Graham, Scott A. Hale and Monica Stephens in collaboration with Dr Corinne M. Flick and the Convoco Foundation.

This map and other visualizations can be found on the OII visualization website at http://www.oii.ox.ac.uk/vis/

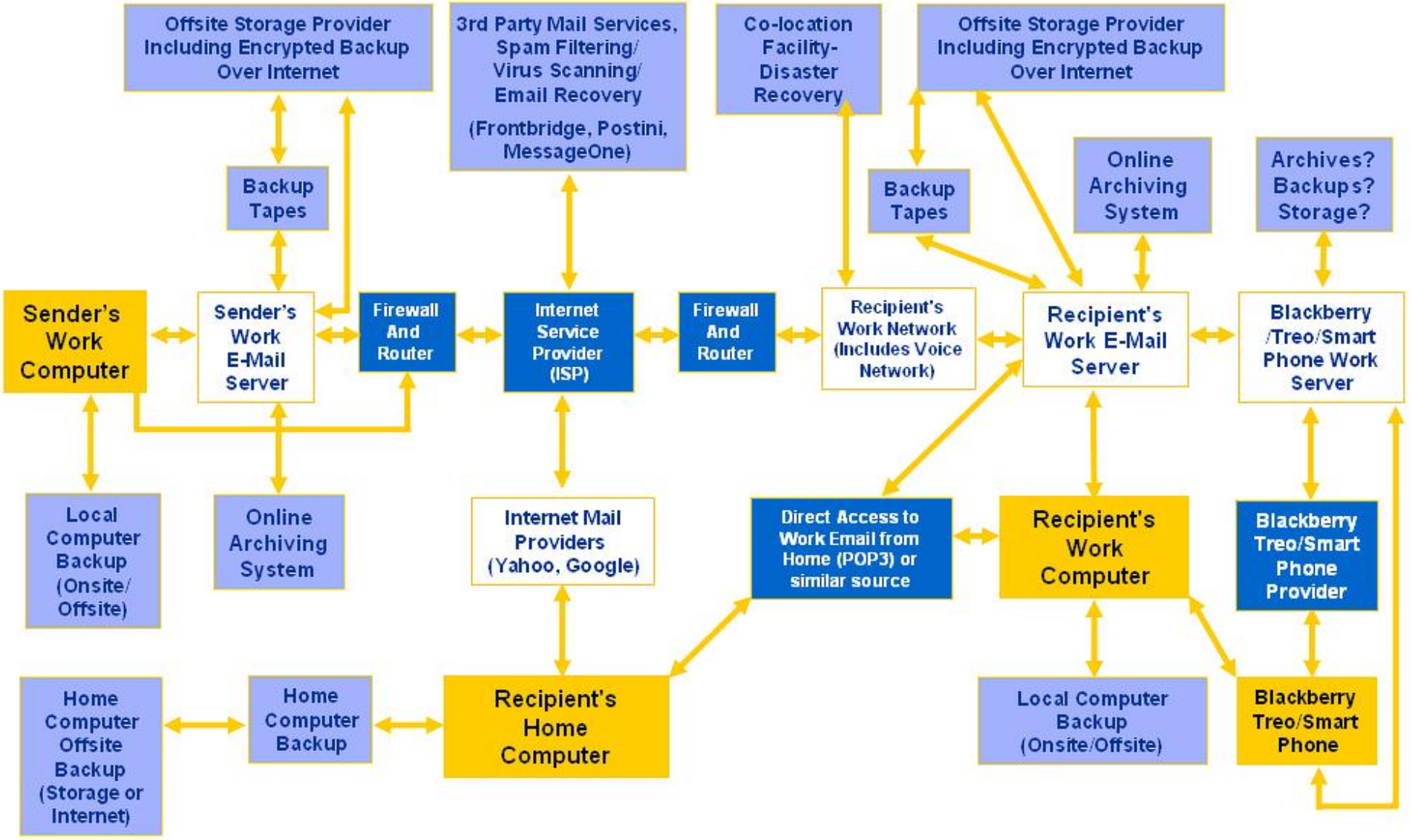
Email Metadata

- Message metadata
  - Times
    - Sent
    - Resent
    - Received
  - Route
  - In-reply-to
  - Attachment file type
- System metadata
  - Folder

File Types

- Extensions
  - MyDocument.xls
- MIME type
- Magic bytes
- Supervised machine learning

Potential Email Routes and Locations



Capture

- Imaging
  - Tape copy
  - Disk image
- Active file capture
  - Hardware write block
  - Software write blocking
- File system copy

Culling

- Custodian
- De-NISTing
  - Based on NIST list of known program hashes
- Date range

Preservation

- Future accessibility
  - Replication
  - Service copies
- Authenticity
  - Documented traceable process
  - Separately stored hashes