

Identification and Collection

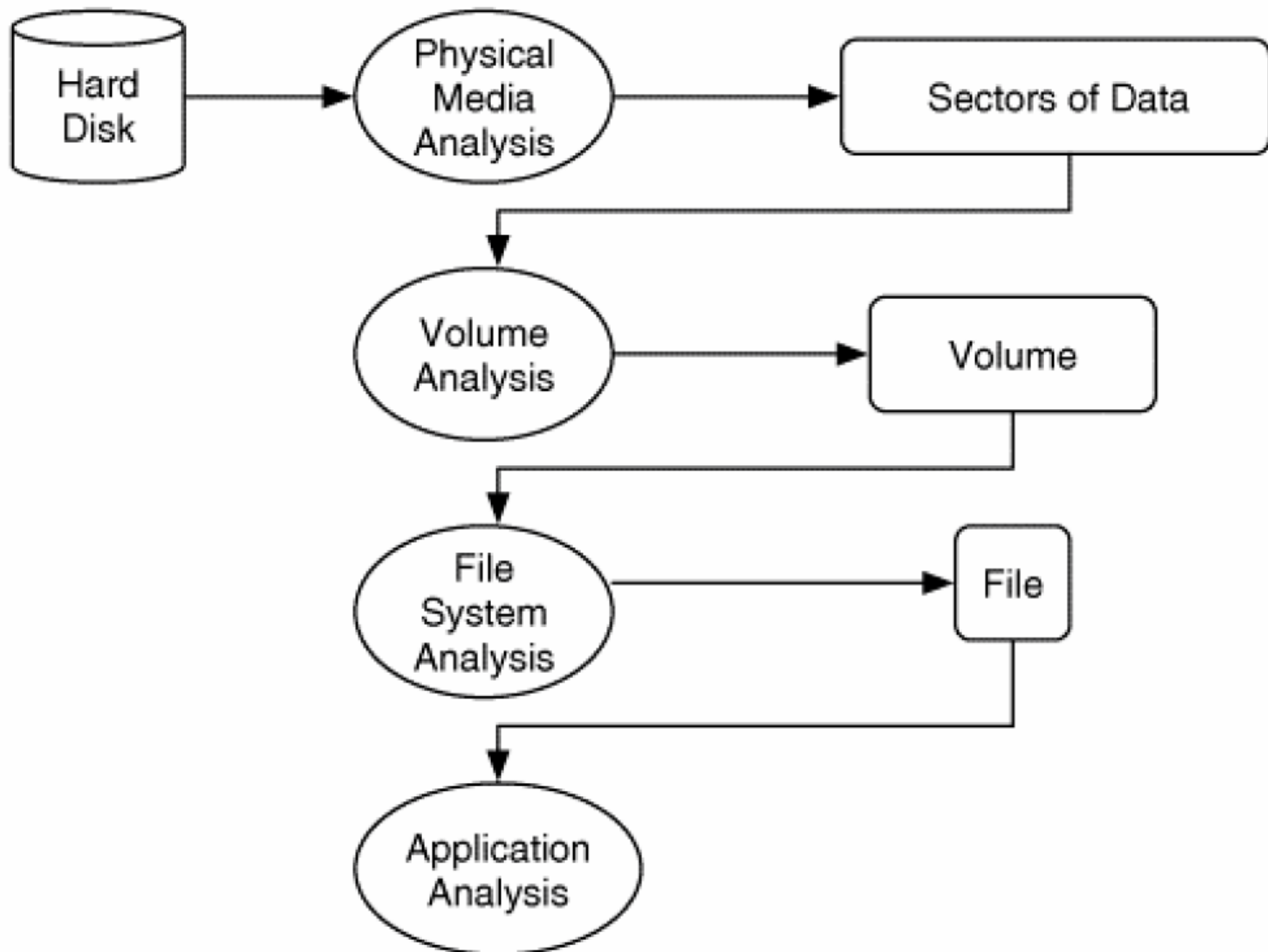
INFM 718X/LBSC 708X

Douglas W. Oard

“Data” Mapping

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

Levels of Analysis



How Disks Work

Step 1:

The circuit board controls the movement of the head actuator and a small motor.

Step 2:

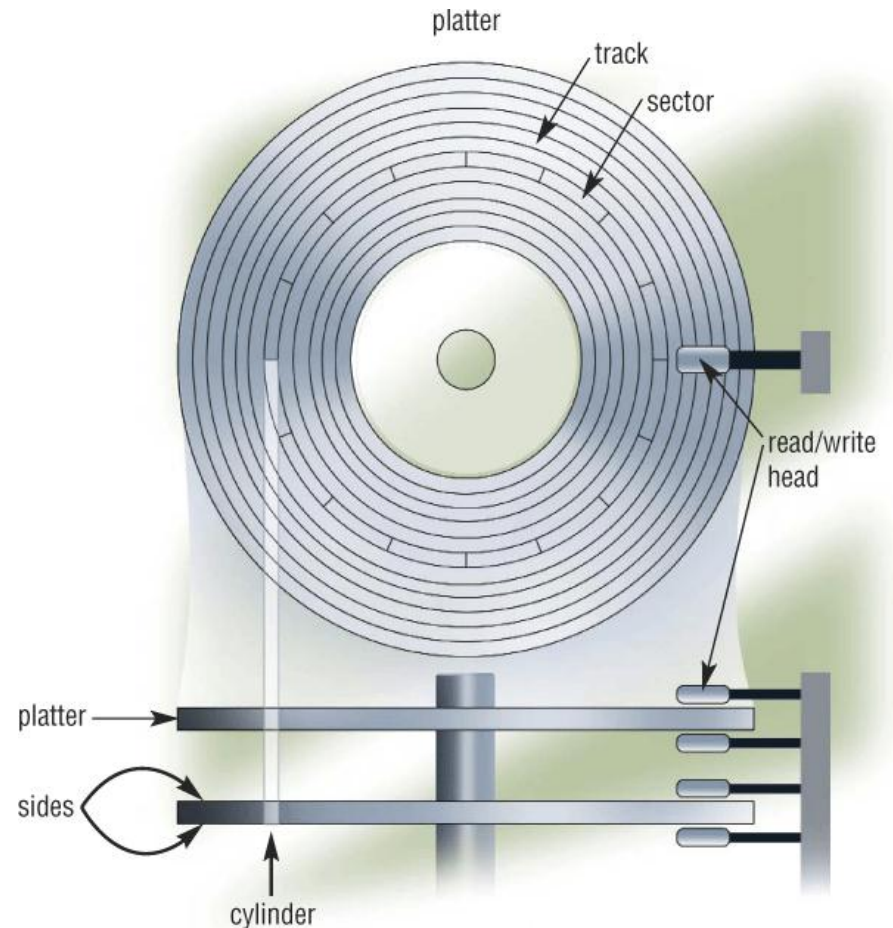
A small motor spins the platters while the computer is running.

Step 3:

When software requests a disk access, the read/write heads determine the current or new location of the data.

Step 4:

The head actuator positions the read/write head arms over the correct location on the platters to read or write data.



Windows “NTFS” File Metadata

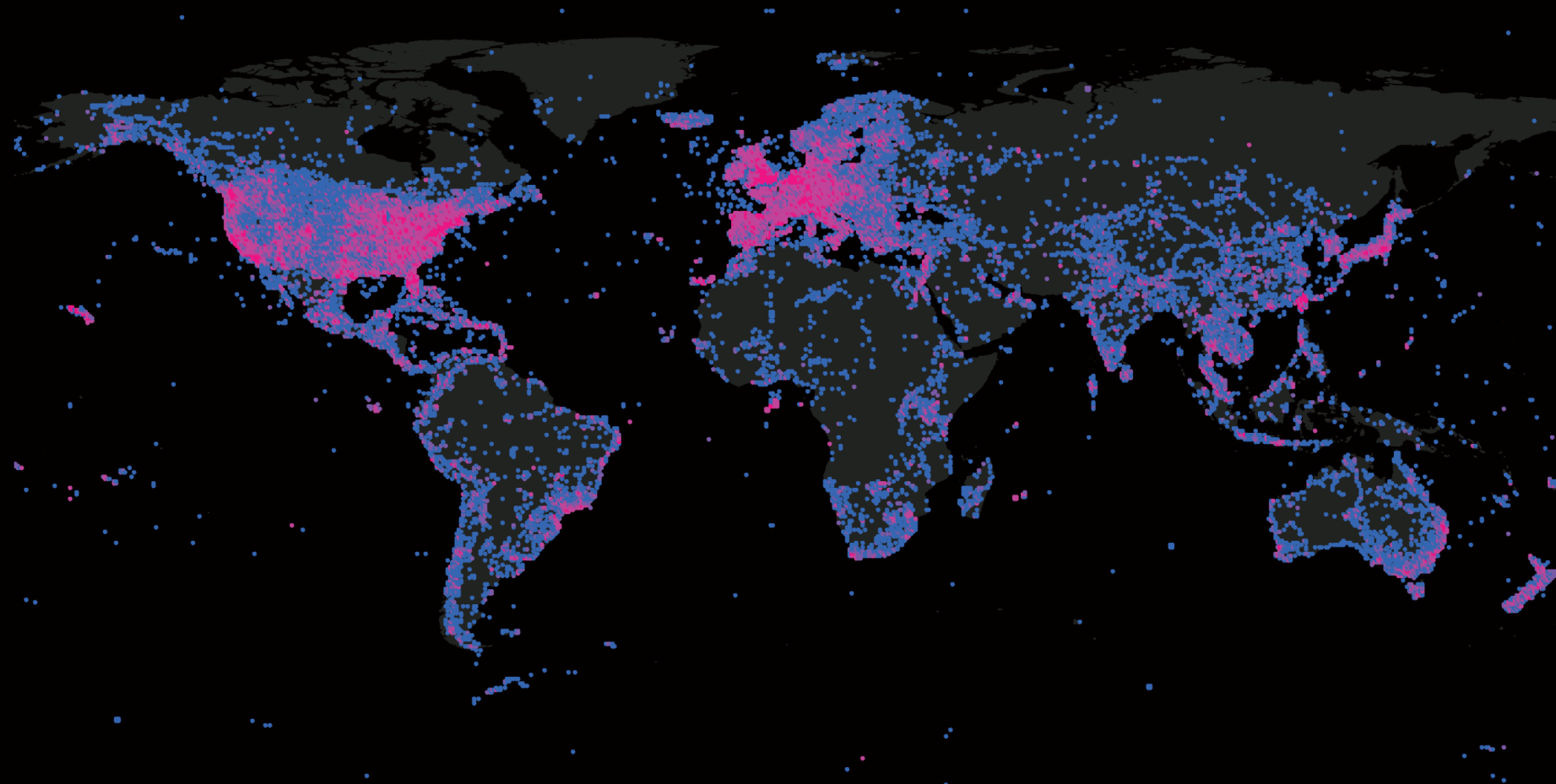
- Time file created (or copied)
 - Most recent one; optionally “journaled”
- Time file content changed (or made changeable)
 - Most recent one; optionally “journaled”
- Time file renamed (or moved)
 - Most recent one
- Time file metadata created or changed
 - Most recent one
- Time file accessed (content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



Number of photographs

- 5 - 100
- 501 - 10,000
- 101 - 500
- 10,001 - 1,000,000

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/>

Copyright © Oxford Internet Institute in cooperation with Dr. Corinne M. Flick and the Convoco Foundation 2011

This publication is released under the Creative Commons Attribution-NonCommercial-NoDerivs [CC BY-NC-ND] license.



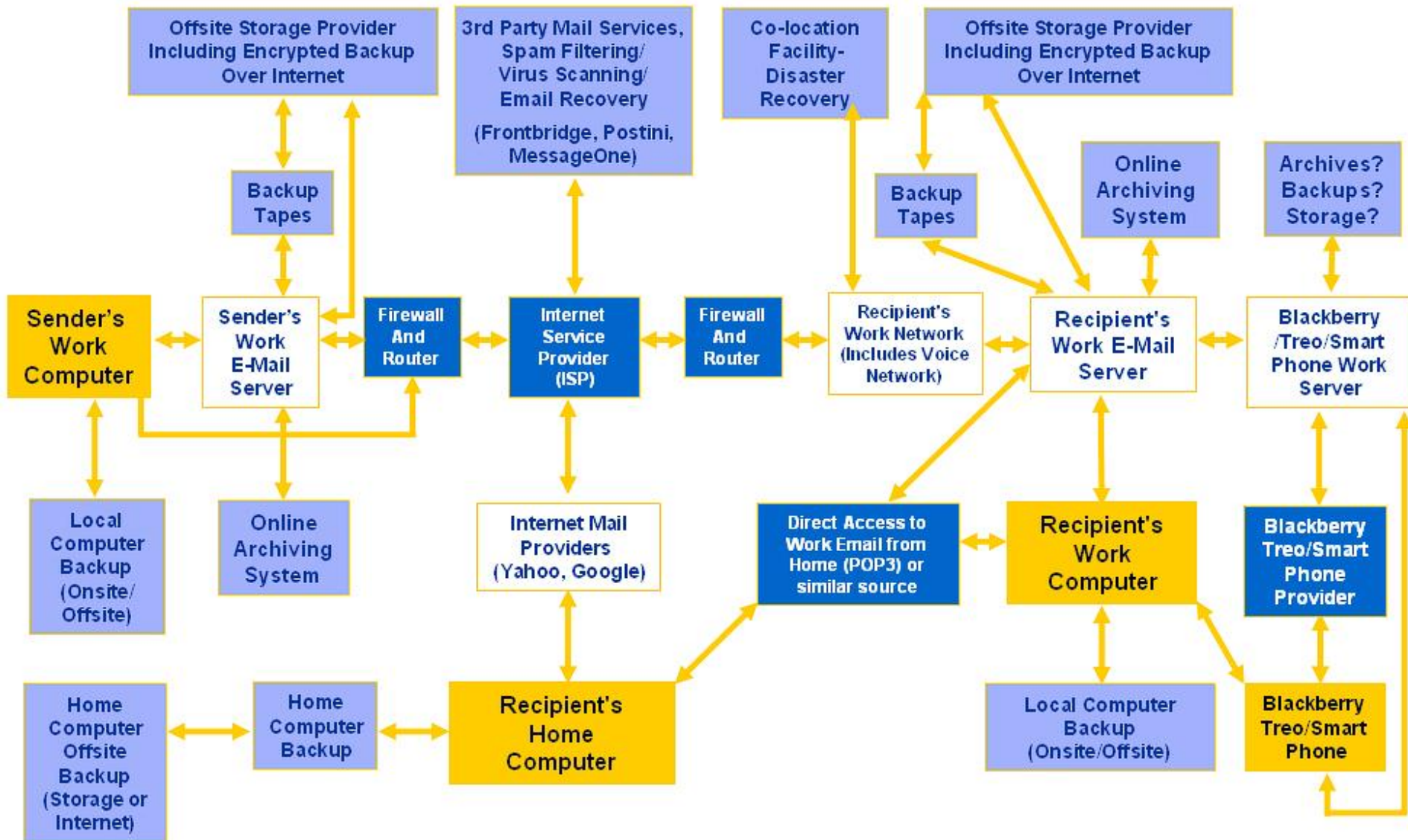
Email Metadata

- 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