Managing Digital Collections

Greg Colati, Digital Directions, 2014

Managing Digital Collections

Managing digital collections is the same as managing physical collections, except when it is different.

- Appraise
- Accession
- Arrange
- Describe
- Make available

Digital Collections…

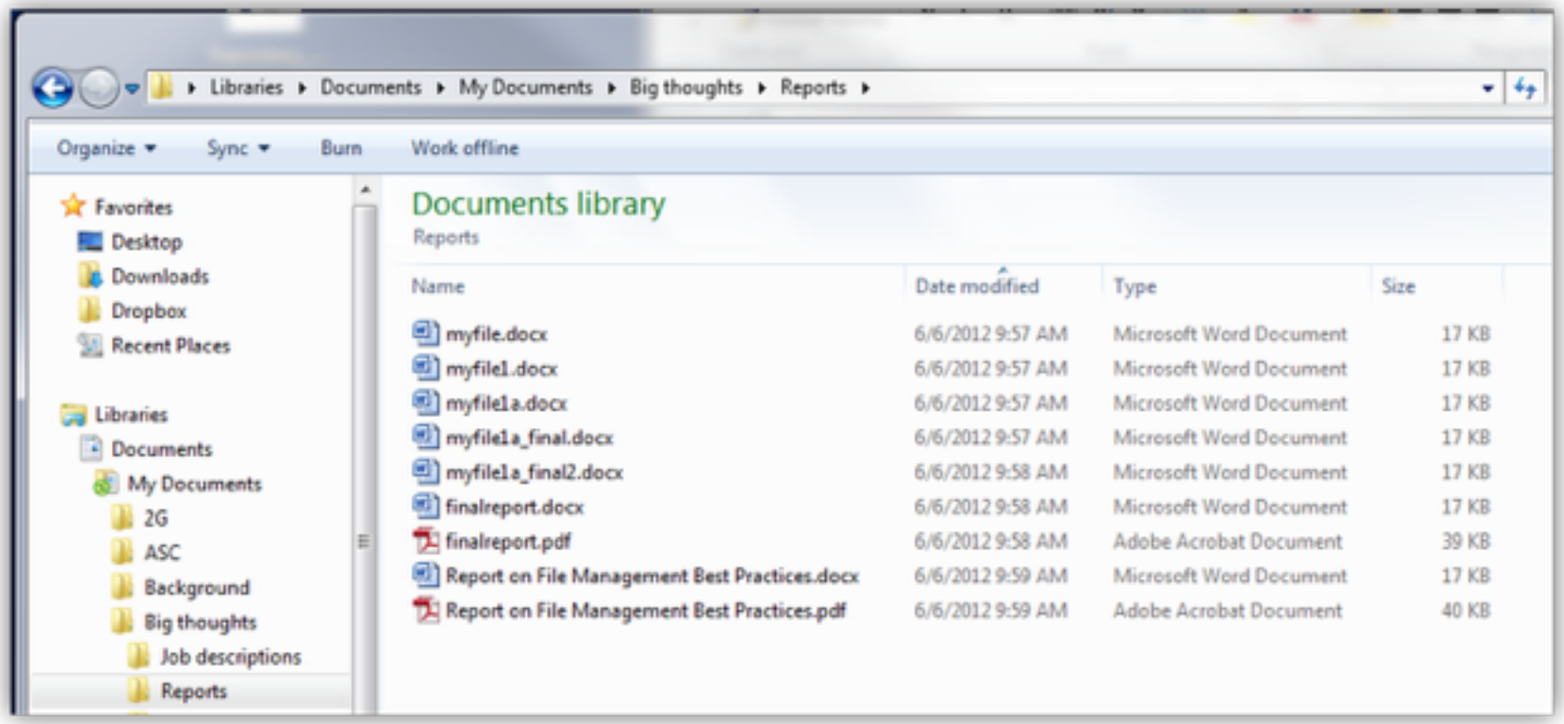
… are **fragile**

… require some type of **item-level control** at least for management, if not for access

... require an **intervening technology** at every stage.

Appraisal in a Digital World

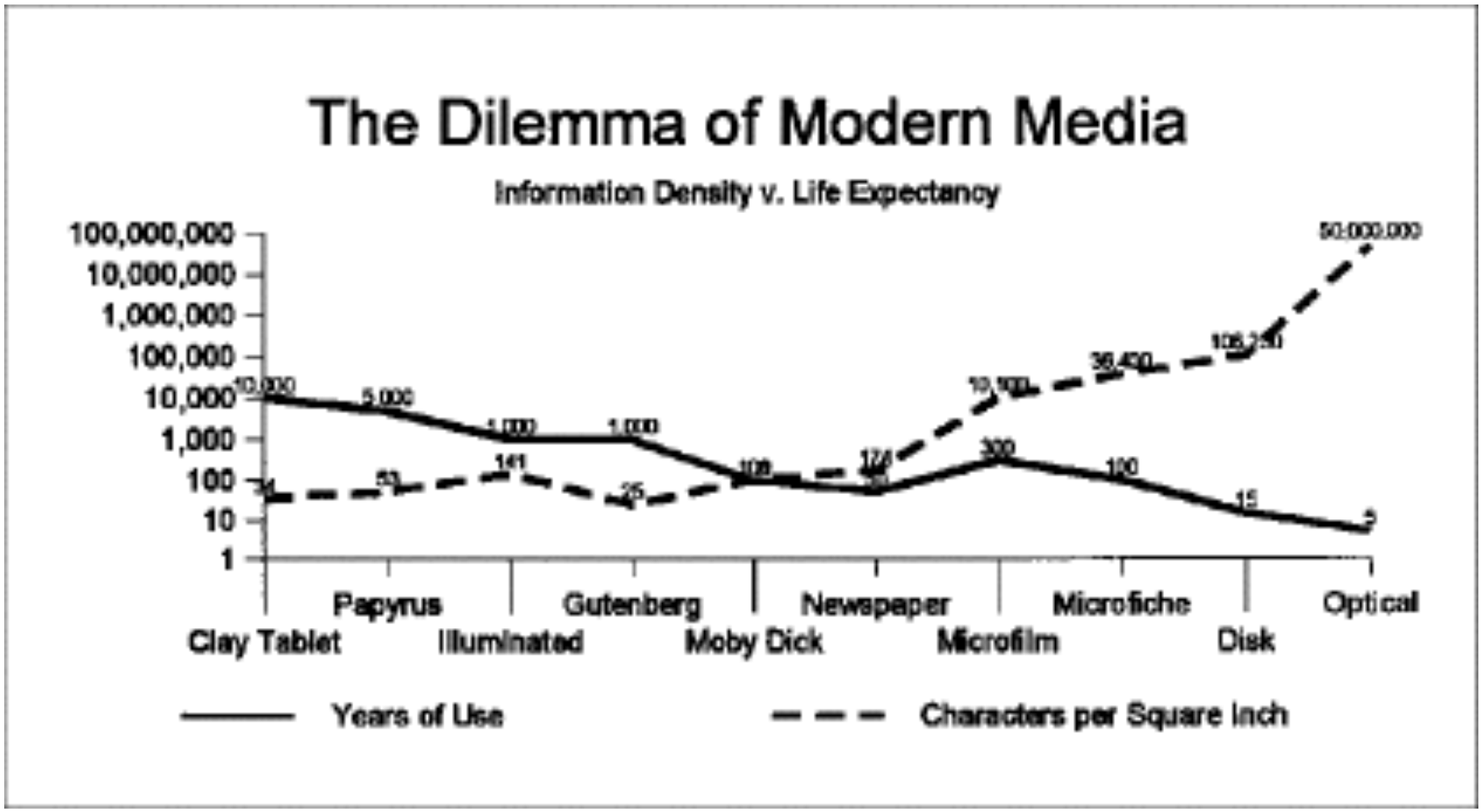
- Carrier media and file formats
- De-duplication
- Deleted files
- Disk images
- Virtual machines



What Are We Trying to Save?

- Original source file characteristics
- Normalized information objects
- Disk Images
- Operating Systems
- Software applications
- The "experience" of the original

The Dilemma of Modern Media, 1996



Dilemma of Modern Media, 1912

Cultural Armageddon, 1980s

Cultural Armageddon, 1990s

Cultural Armageddon, 2000s

CBSNEWS Video | US | World | Politics | Entertainment | Health | MoneyWatch

AP / *February 11, 2009, 8:51 PM*

Coming Soon: A Digital Dark Age?

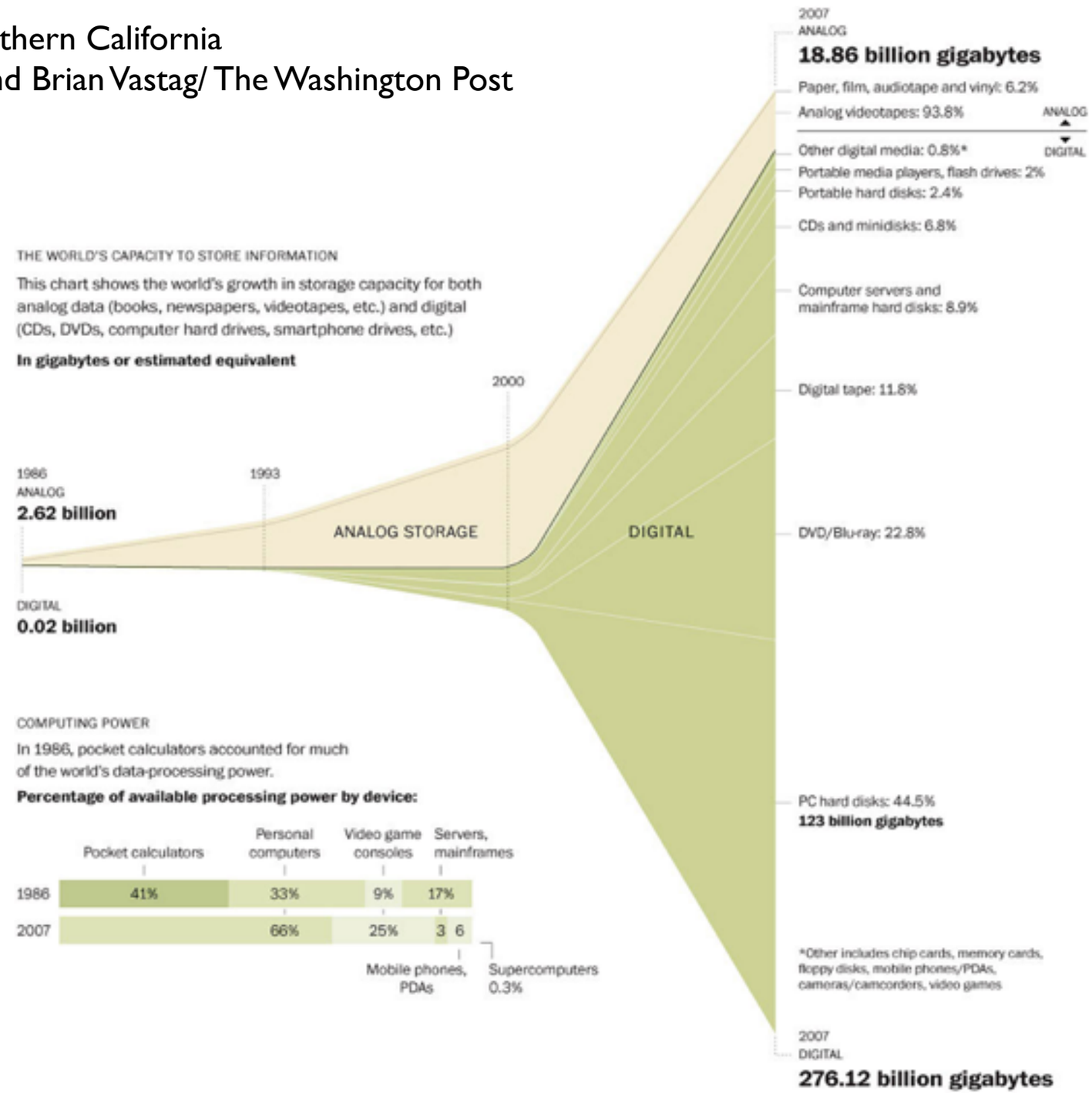
Comment / Shares / Tweets / Stumble / Email More +

You may have recently discovered priceless photographs of your childhood, yellowing but still tangible.

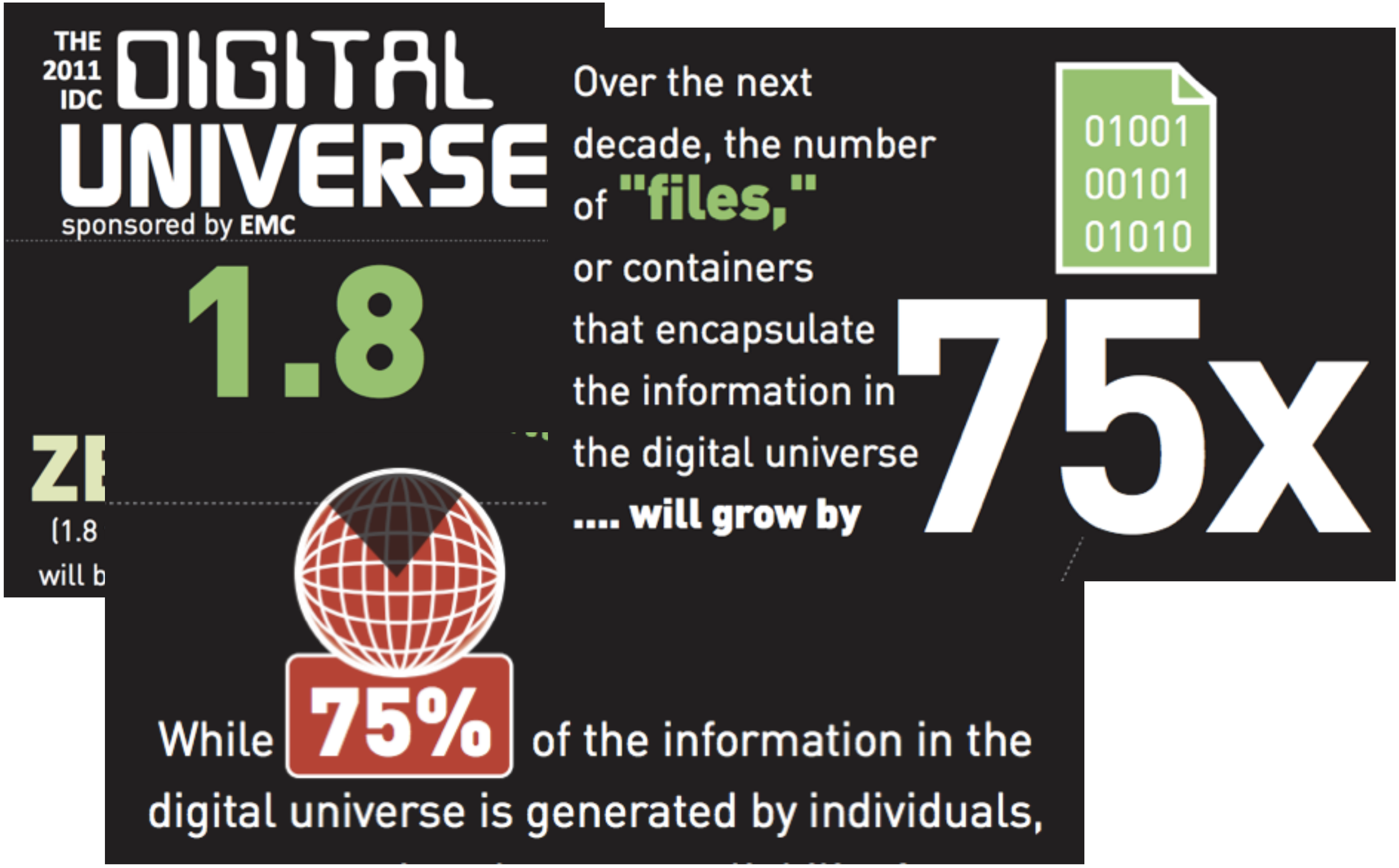
Your grandkids probably won't fare as well with your digital photos.

Cultural Armageddon, 2011

Source: University of Southern California
Credit: Todd Lindeman and Brian Vastag/ The Washington Post



Cultural Armageddon, 2012



Cultural Armageddon, 2013

Legacy.com®

Where life stories live on.®

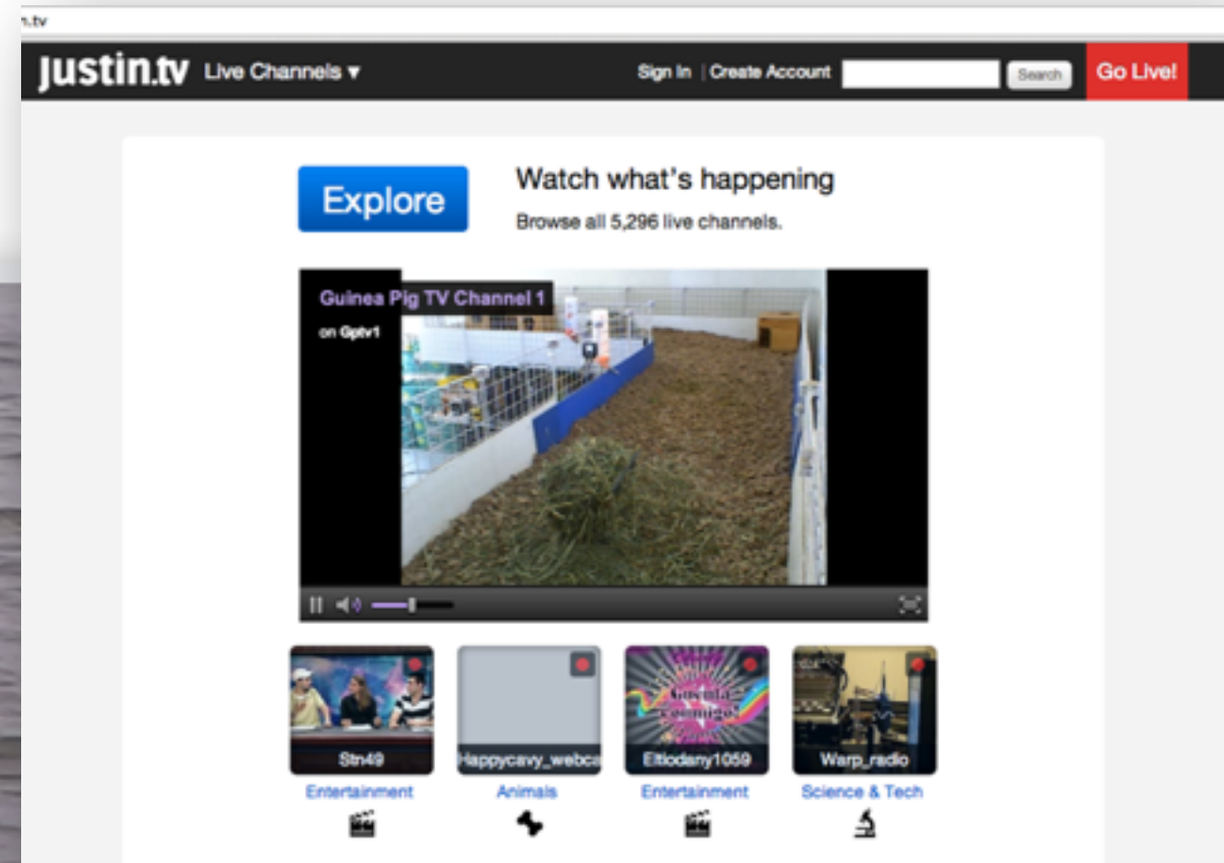
TERMS OF USE

Updated: March 2013

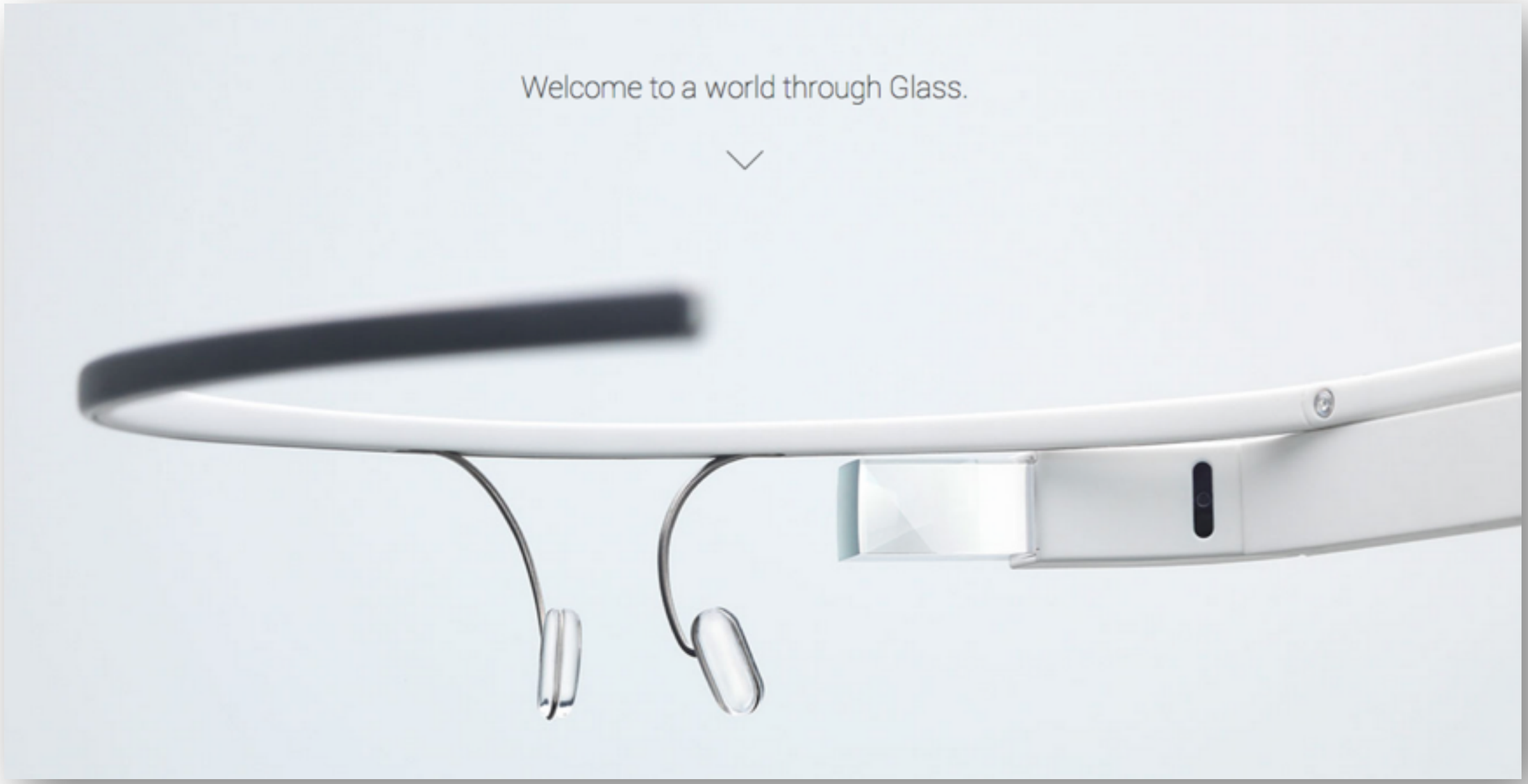
Scope. Legacy.com, Inc. grants you permission (which ma
or no reason) to view the Services and to download, integra
email, or print individual pages of the Services in accordance with these Terms of Use and solely for your own personal, non-commercial use, provided you do not remove any trademark, copyright or other notice contained on such pages. No other use is permitted. You may not, for example, incorporate the information, content, or other material from the Services in any database, compilation, archive or cache. You may not modify, copy, distribute, re-publish, transmit, display, perform, reproduce, publish, reuse, resell, license, create derivative works from, transfer, or sell any information, content, material, software, products or services obtained through the Services, except as specifically noted above.

"You may not incorporate the information… in any …archive"

Cultural Armageddon, 2013

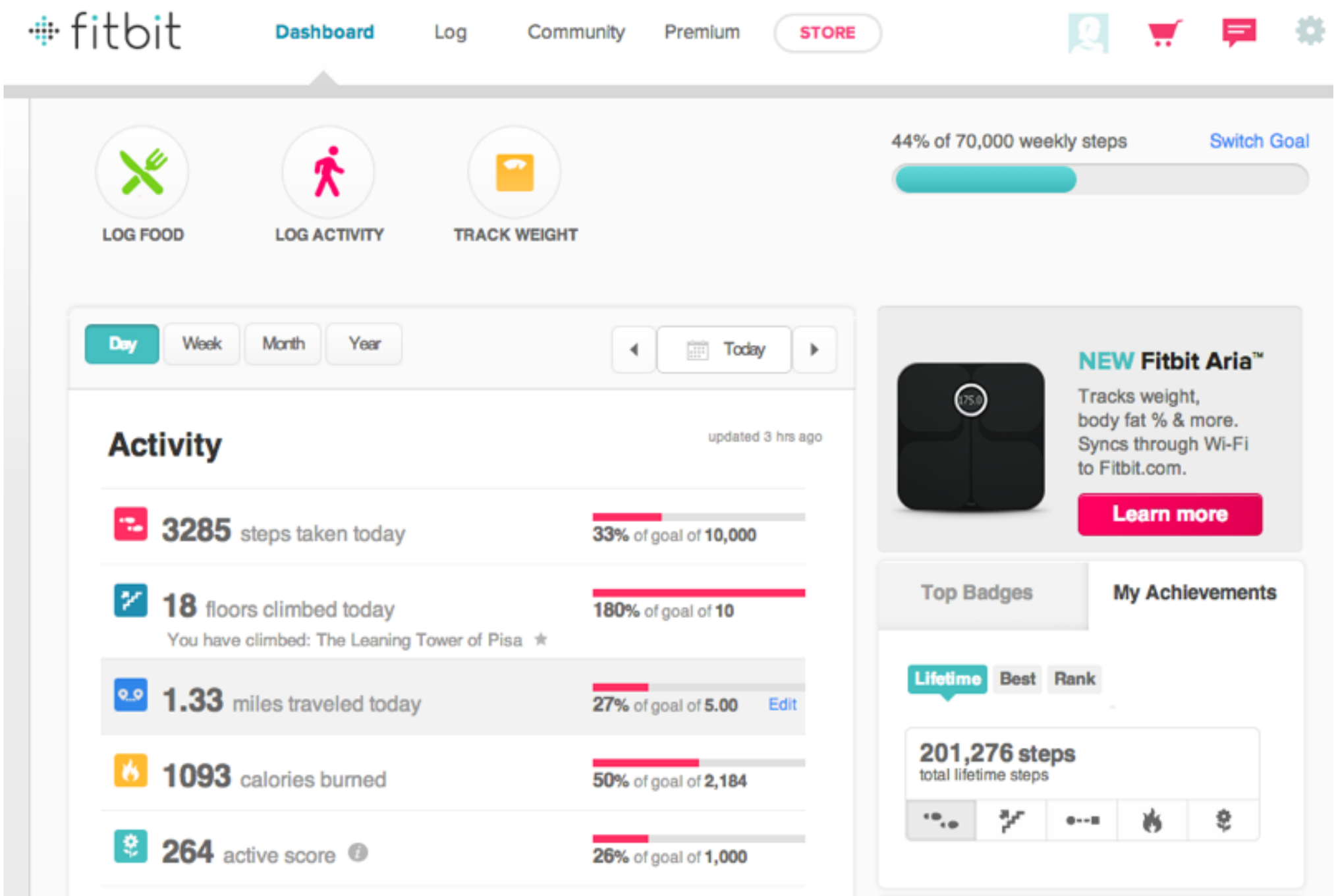


Cultural Armageddon, 2014?

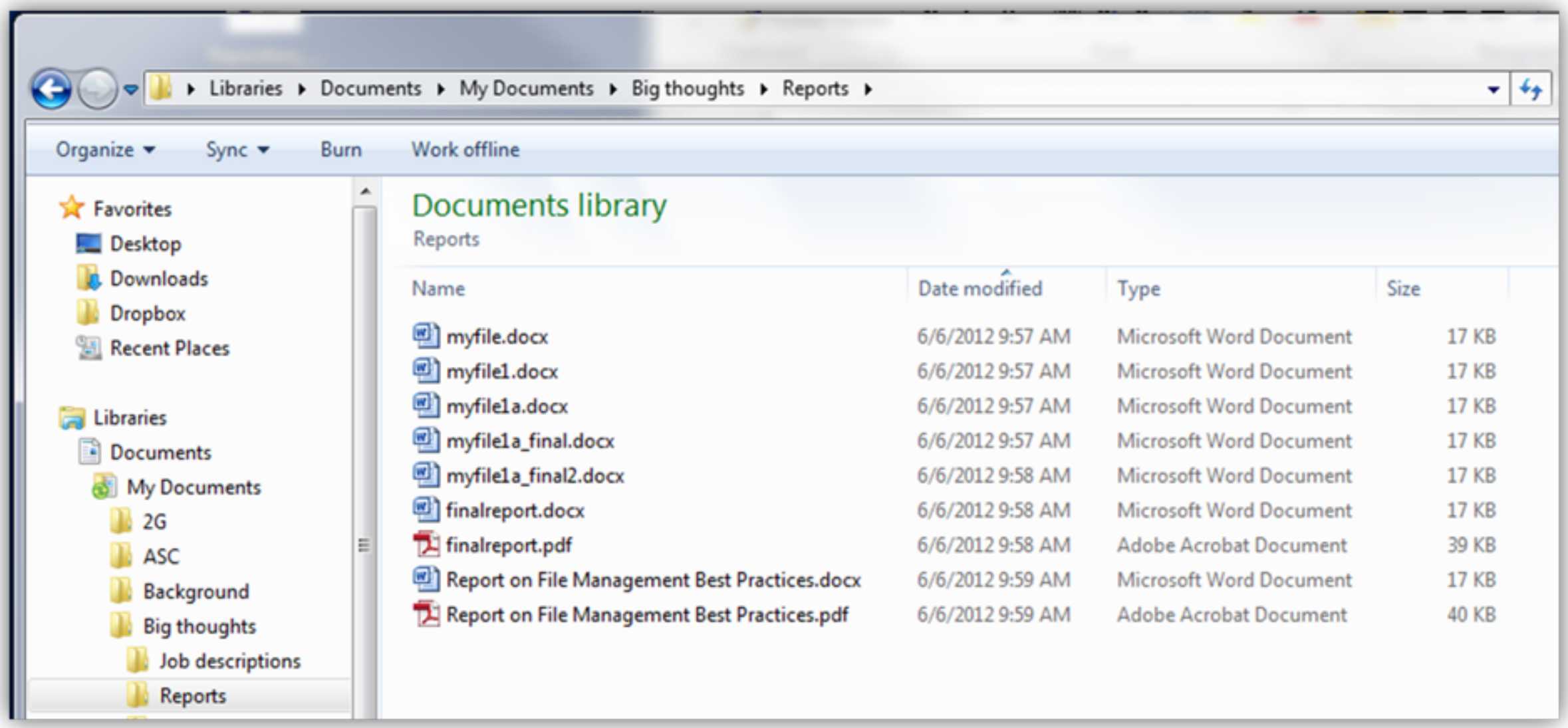


What Has Value?

What is Noise?



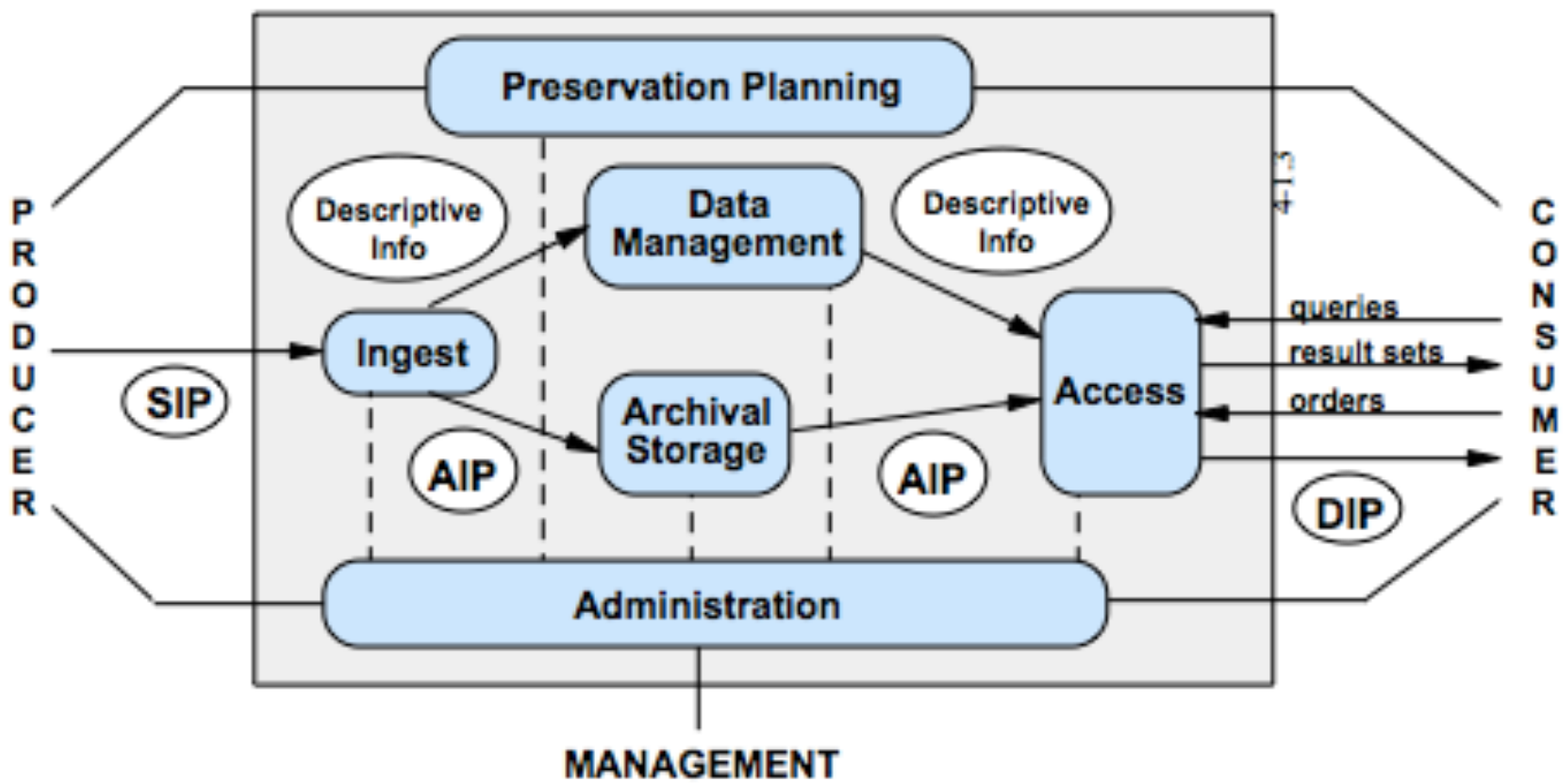
Cultural Armageddon: The Digital Attic



OAIS: It **IS** Rocket Science!

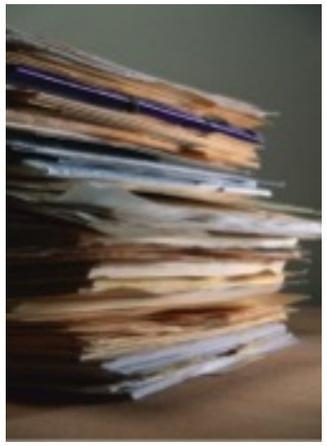
- Conceptual model of an information object that is self-contained and self-describing.
- A set of data elements (metadata of all types, primary content objects, associative information, system information) combined into a “package” that is internally coherent and can be managed in an electronic environment by applications and processes.

Gratuitous OAIS Slide

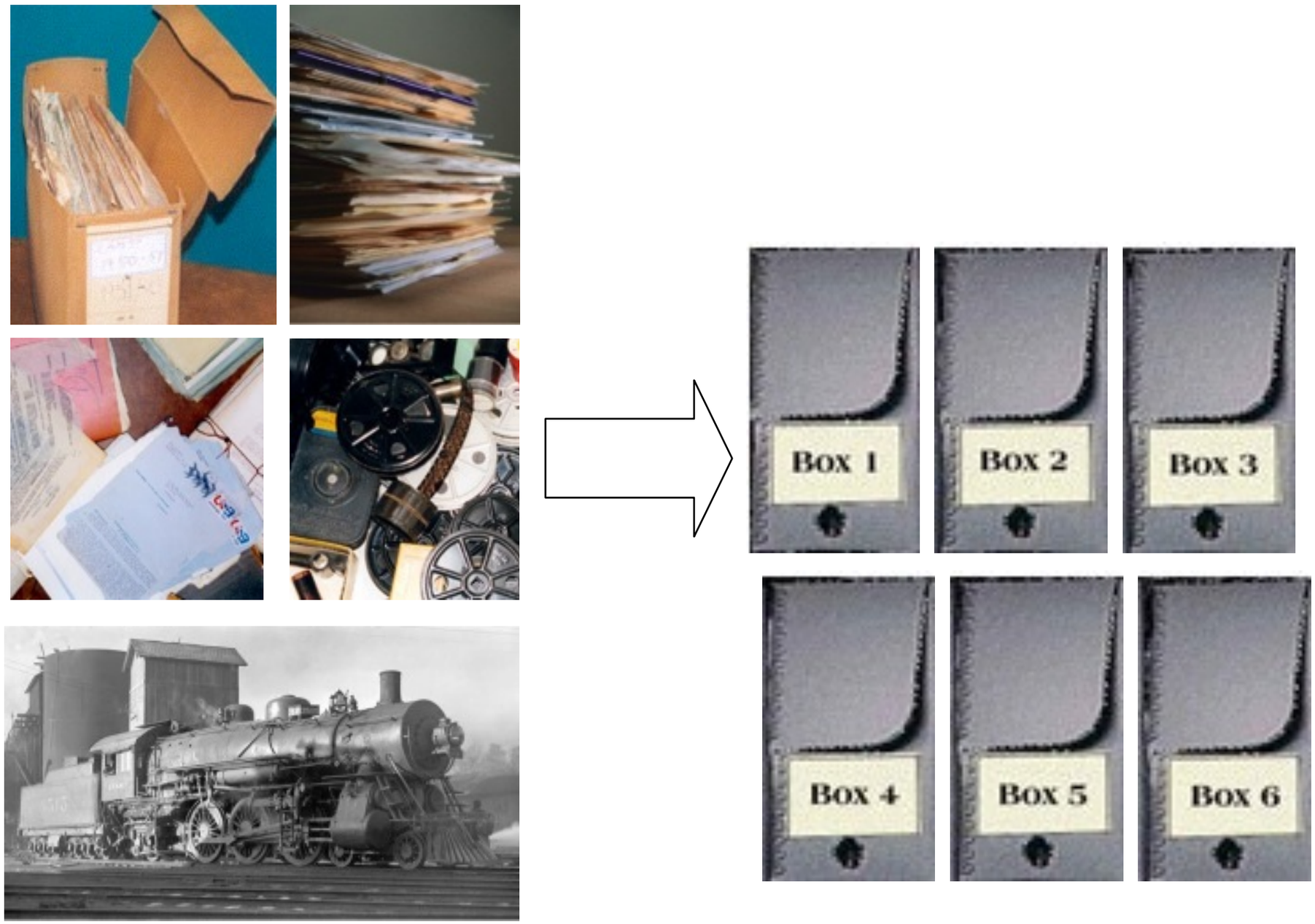


“Born Analog” Content…

- Records, ledgers, transcripts, photographs, film, diaries, maps, correspondence, paintings, artifacts, artwork, texts, audio, video, datasets, learning objects, institutional records, official documents, bulletins, directories and publications…



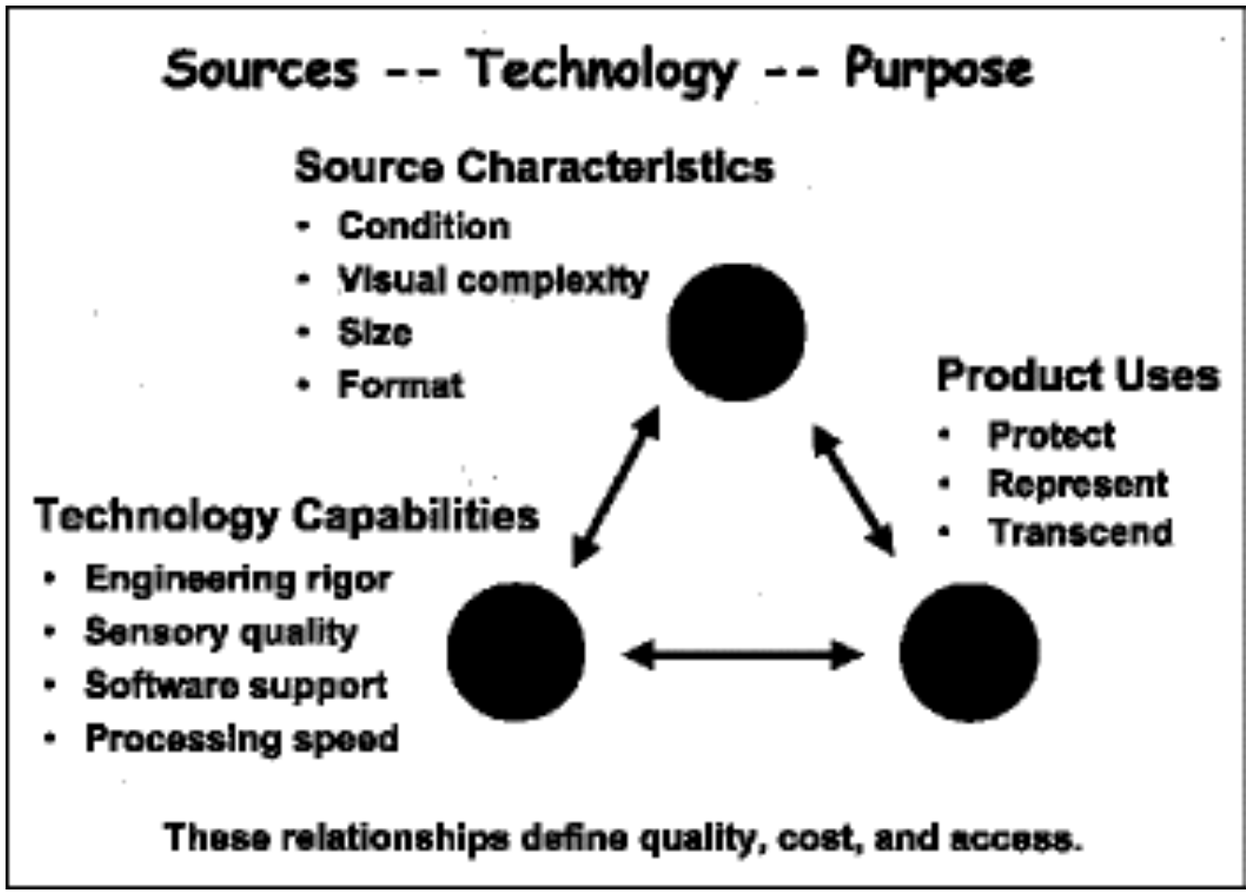
…Becomes Physical Items in Collections…



…And can be Transformed into Quality Digital Objects

- Images
  - Format, bit depth and resolution
- Audio
  - Format, sample rate and quantization (compression)
- Text
  - Format, levels of accuracy, levels of encoding,
- 3-D objects
  - Levels of detail, capture to support static vs. dynamic representation

How Born Analog Objects Acquire Digital Identities



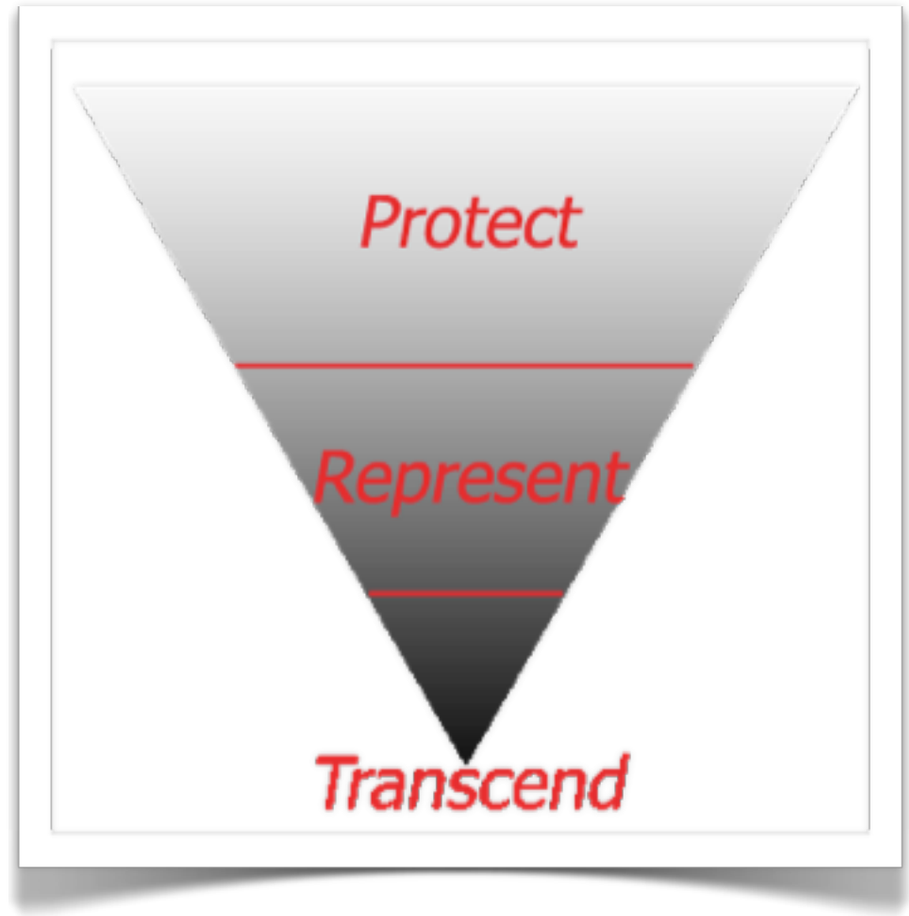
Paul Conway, in Handbook for Digital Projects, NEDCC, 2000

Source

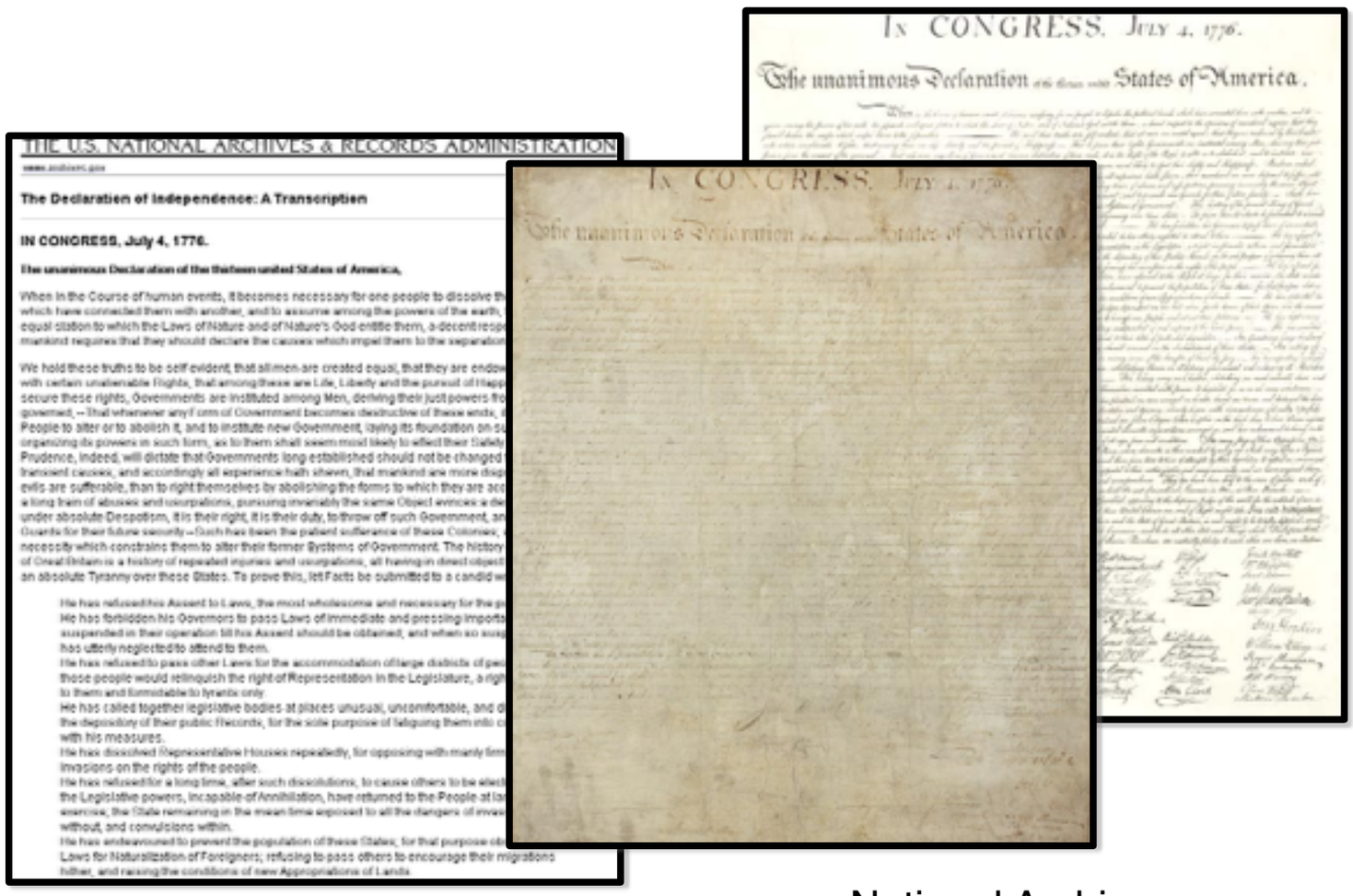
- Stable format
- Good condition
- Well documented--at the item level
- Unique
- Familiar

Purpose

- **Protect** original from handling
- **Represent** informational content
- **Transform** use



Practical vs Symbolic Value



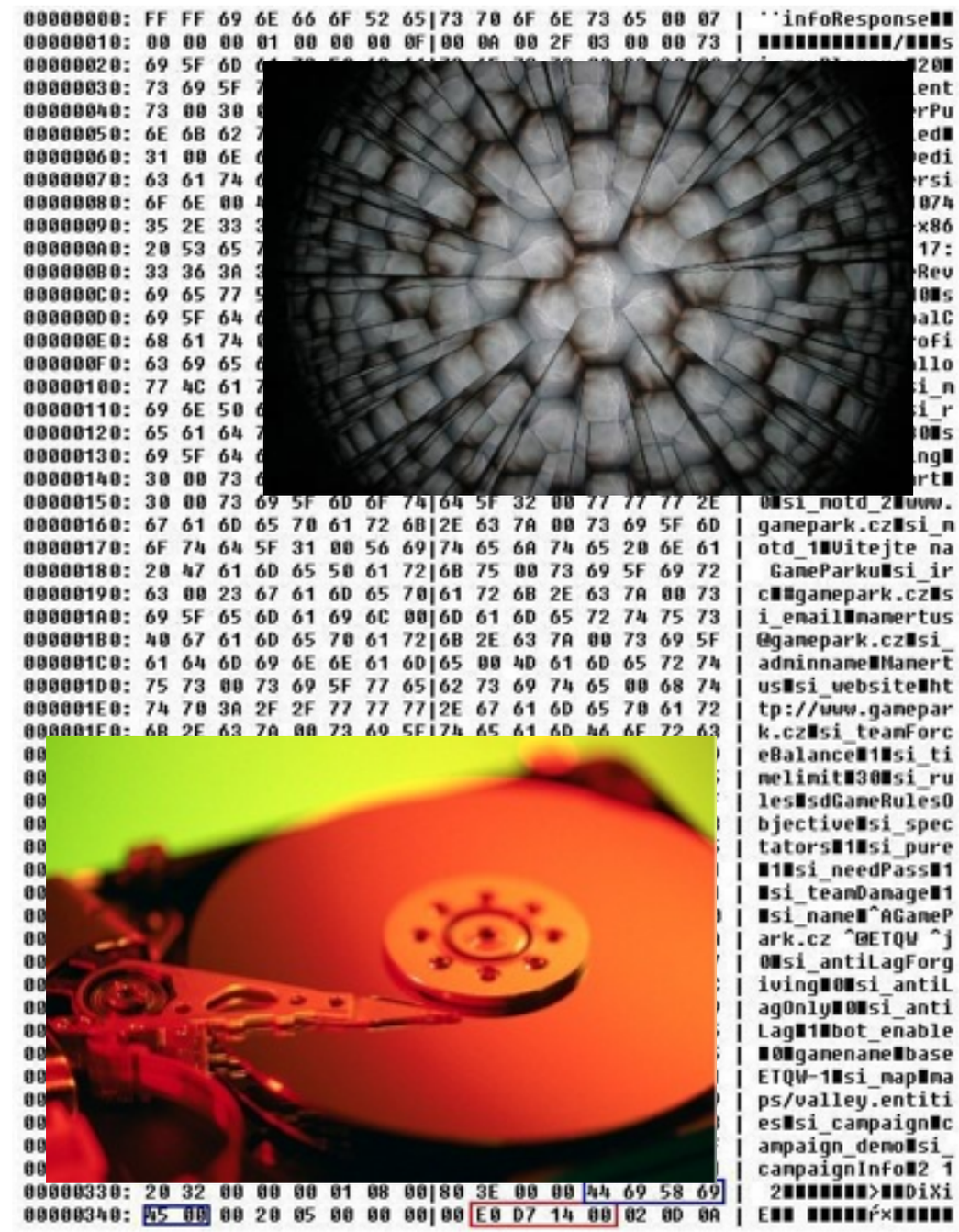
National Archives

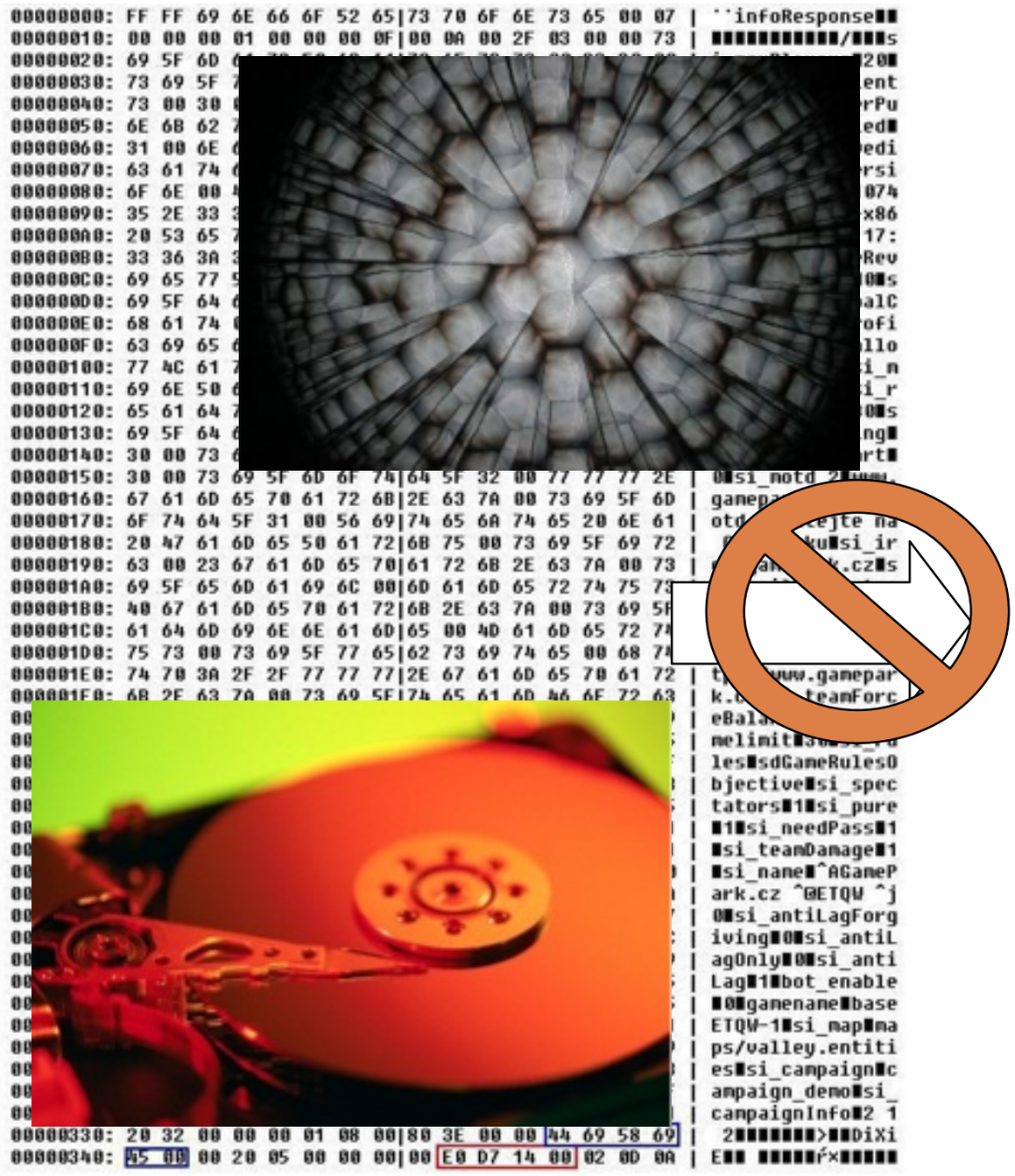
Technology

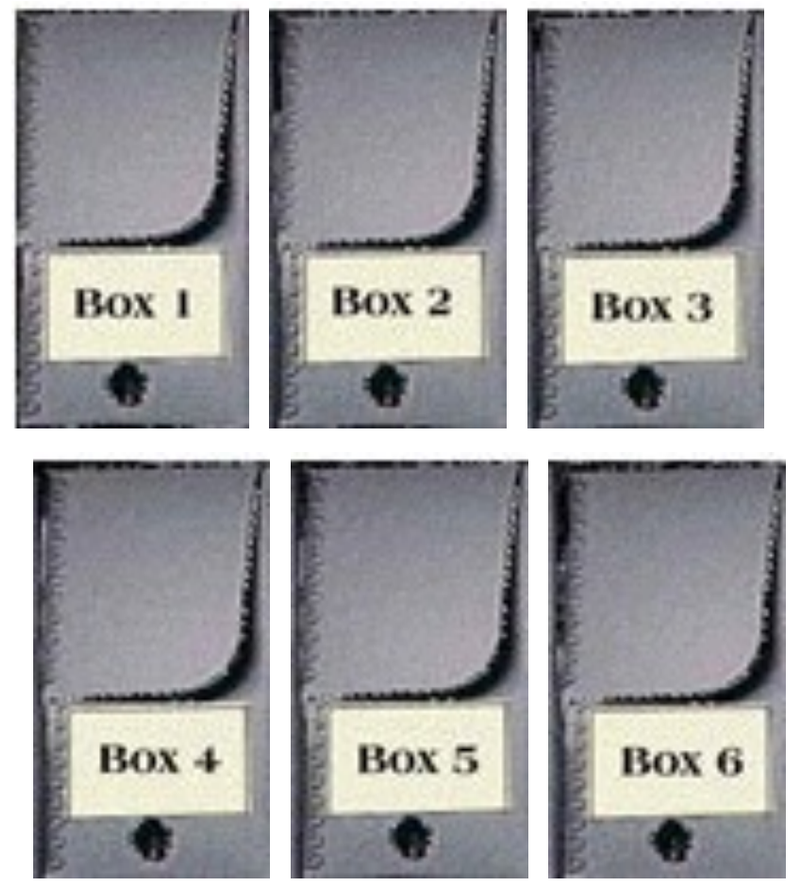
- Does the technology exist to do what you want?
  - Can you do it yourself?
  - Can you afford to outsource it?
- Can you manage and maintain the resulting output?

“Born Digital” Content Follows a Different Path

- Electronic records and documents, digital audio, digital video, digital still imagery, datasets, multi-media works, online publications, web sites, interactives, blogs, wikis…

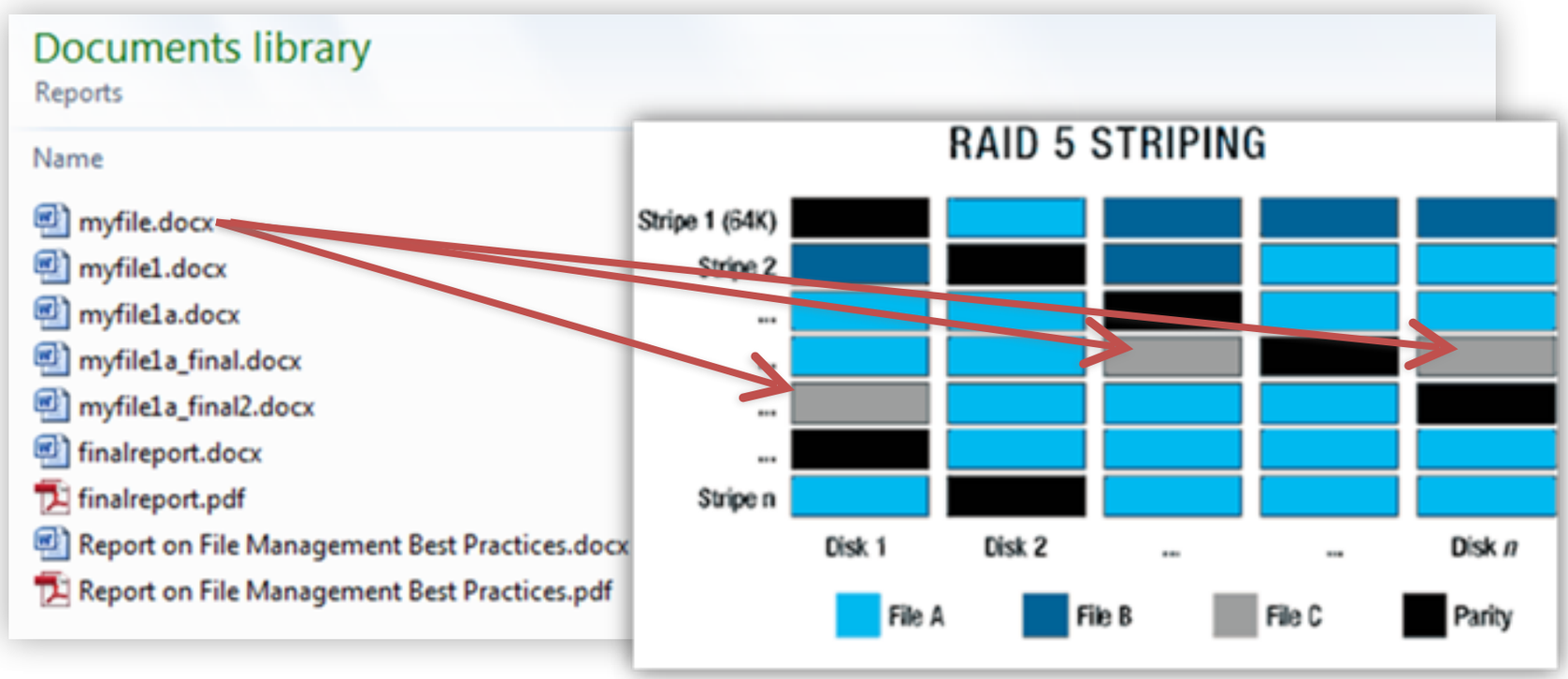





Box 1
Box 2
Box 3
Box 4
Box 5
Box 6

Arrangement and Description

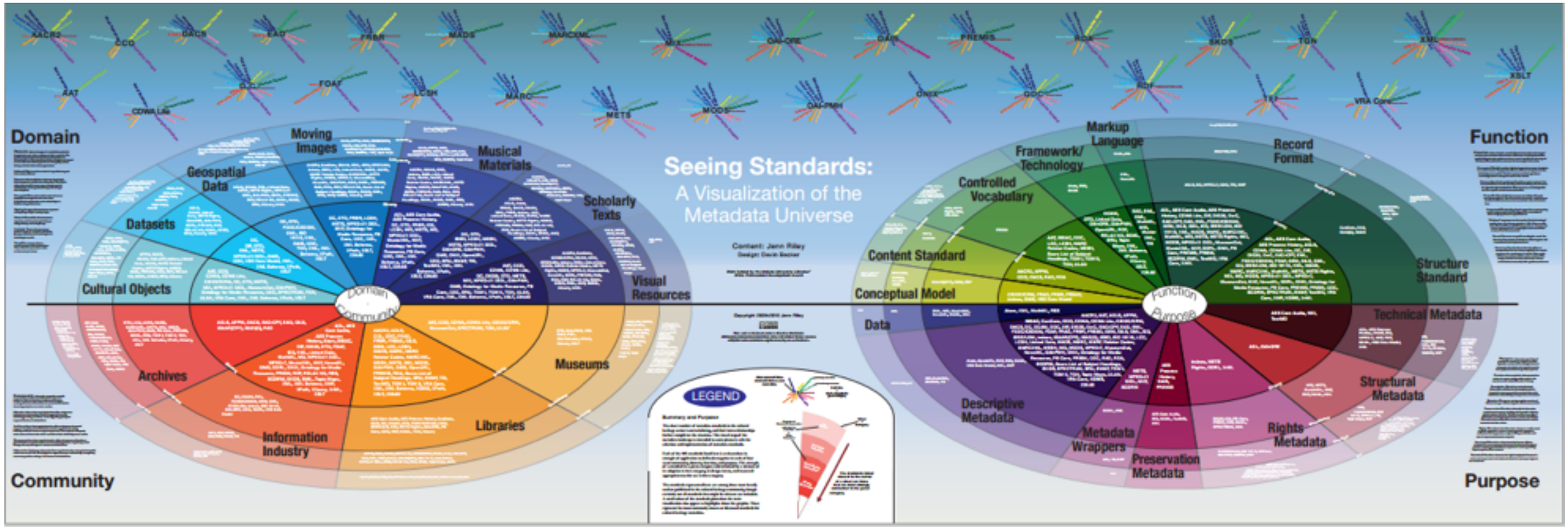
- There is no such thing as physical arrangement
- Intellectual arrangement is completely separate from physical management



Arrangement and Description

- All arrangement is intellectual—and may be unintentional!
- How you define “objects” affects management and access more than arrangement.

Content Brings Its Own Metadata Needs



- Metadata is often specific to a community, a content type, a format, or an audience.

Seeing Standards: A Visualization of the Metadata Universe, Jenn Riley.

…Or No Metadata at All!

How Much Metadata is Enough?

- Metadata maintenance is perpetual
  - To conform to emerging standards
  - To adjust to new technical environments
  - To add functionality and interoperability

Applications will come and go;
metadata (and digital objects) will live on!

How Do We Think About Our Work?

"When it was made simple, counted in bits, information was found to be everywhere"

-James Gleick, The Information: A History, A Theory, A Flood (2011)

"It is not just about the data, it is about the story"

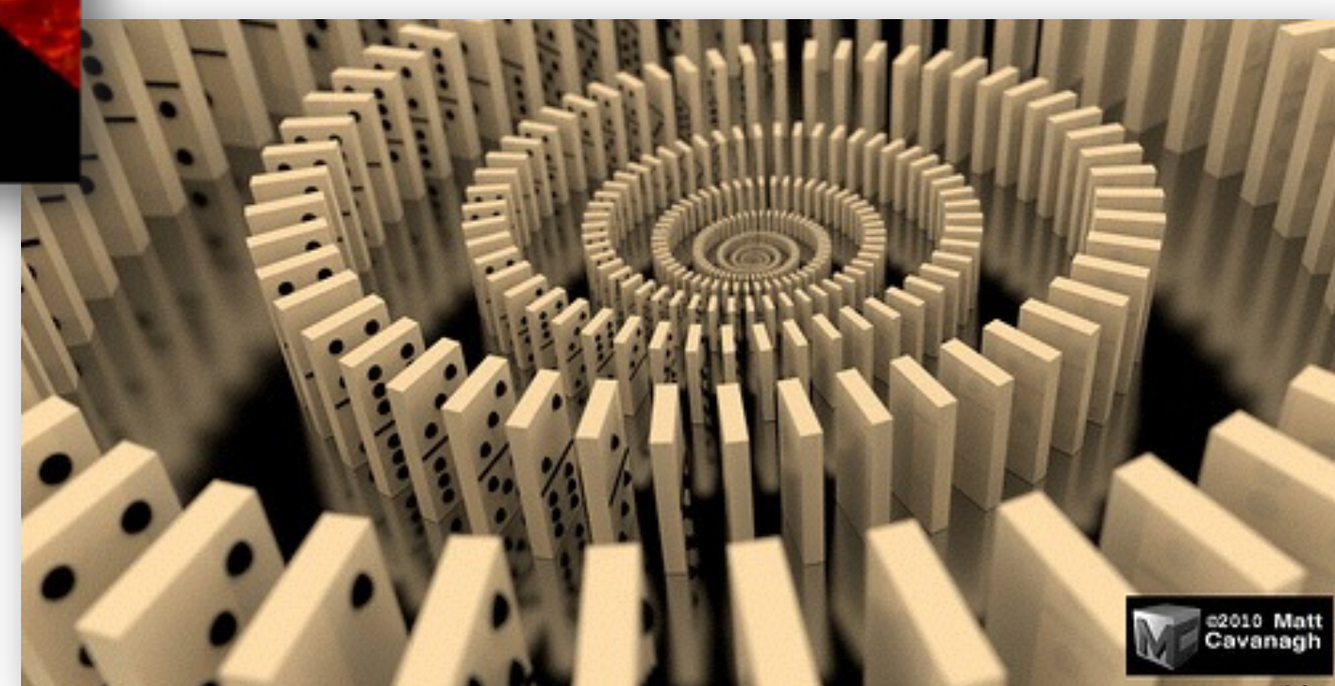
-Arianna Huffington (2012)

Managing Digital Content (Oh yeah... that)

- What is the smallest/largest information unit that we manage as an independent package?
- What quality decisions have we made about the primary content file and its metadata?
- What are the implications of these decisions?

What are We Trying to Manage?

Lumpers vs. Splitters



Lumping into a Simple Object

Elisha Fuller's Account Book, delivered as a PDF file.

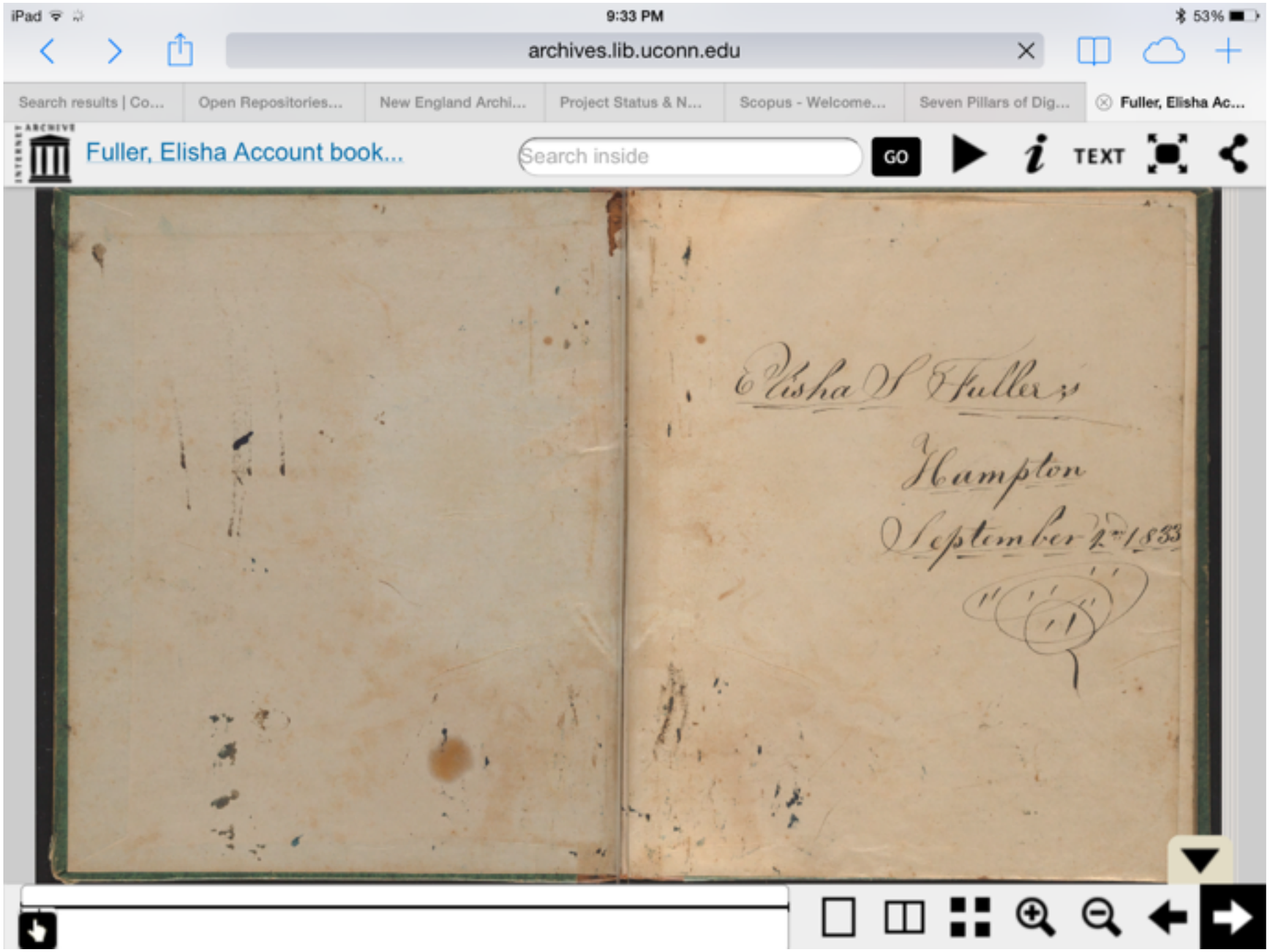
Is this best managed/ delivered as one file...

Splitting Objects

... a set of pages looked at individually ...

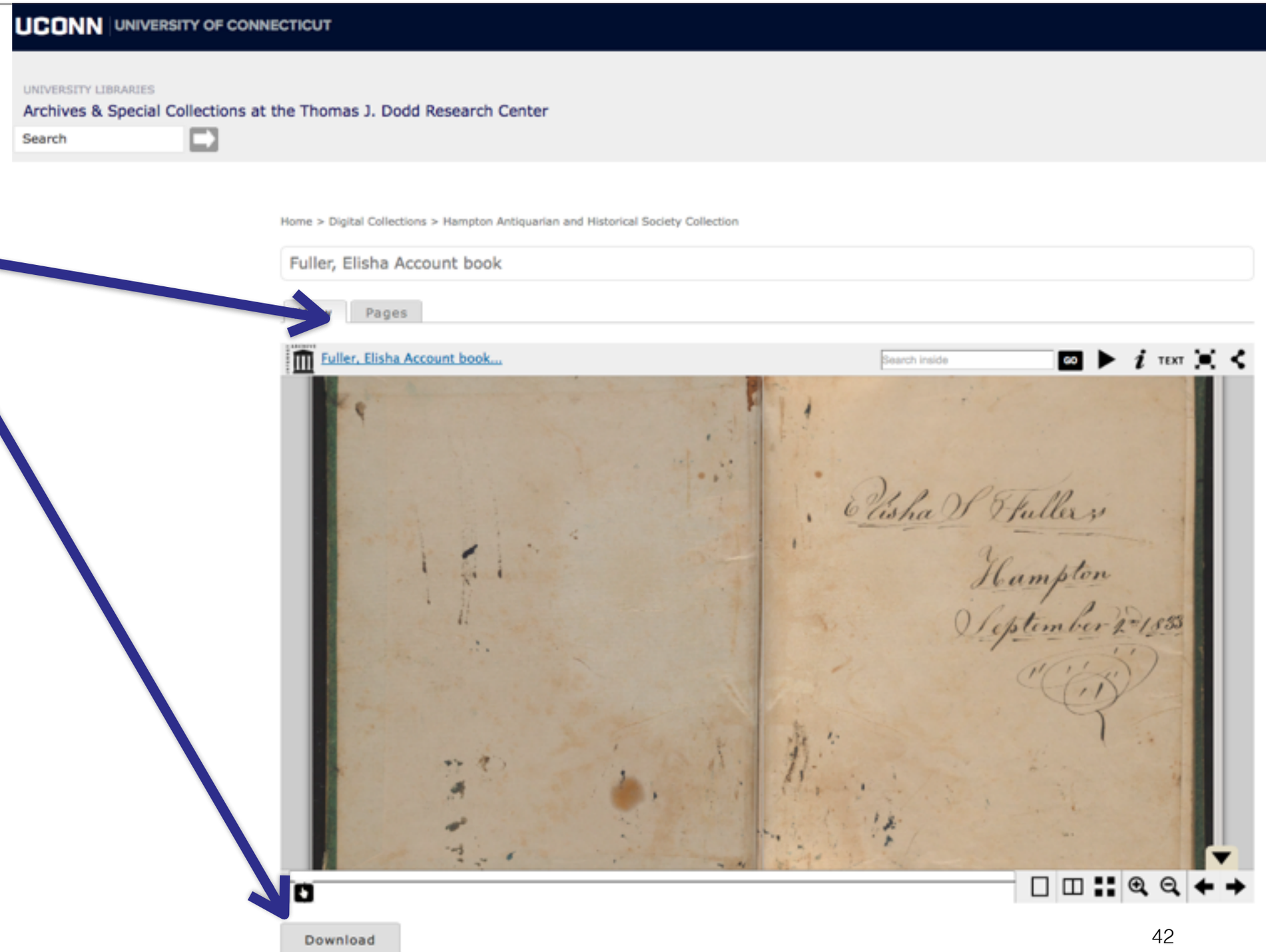
Creating Complex Objects

... a book or ...



... All Three?

Download the pdf,
read the book,
look at the pages.



Discovery and Presentation

- Options are the result of decisions made at the point of creation or ingest
- Is it easier to put things together or take them apart?
- Metadata is the grease that makes the system run smoothly.

How Much Context is Enough?

- Salman Rushdie's Desktop
- Hoagy Carmichael's Room

Management Requires Tools

- Because managing digital objects requires intervening technology at EVERY stage
- Translate functional needs into application services by evaluating and using
  - Vendor-supplied software and tools
    - Digital Commons, Digital Archive, Documentum, Cumulus, Artesia, DigiTool…
  - Open-source software and tools
    - DSpace, Fedora, EPrints, ArchivesSpace, Islandora, Collective Access, …

Software Myth

- “If only we had (fill in the blank)* everything would be fine”

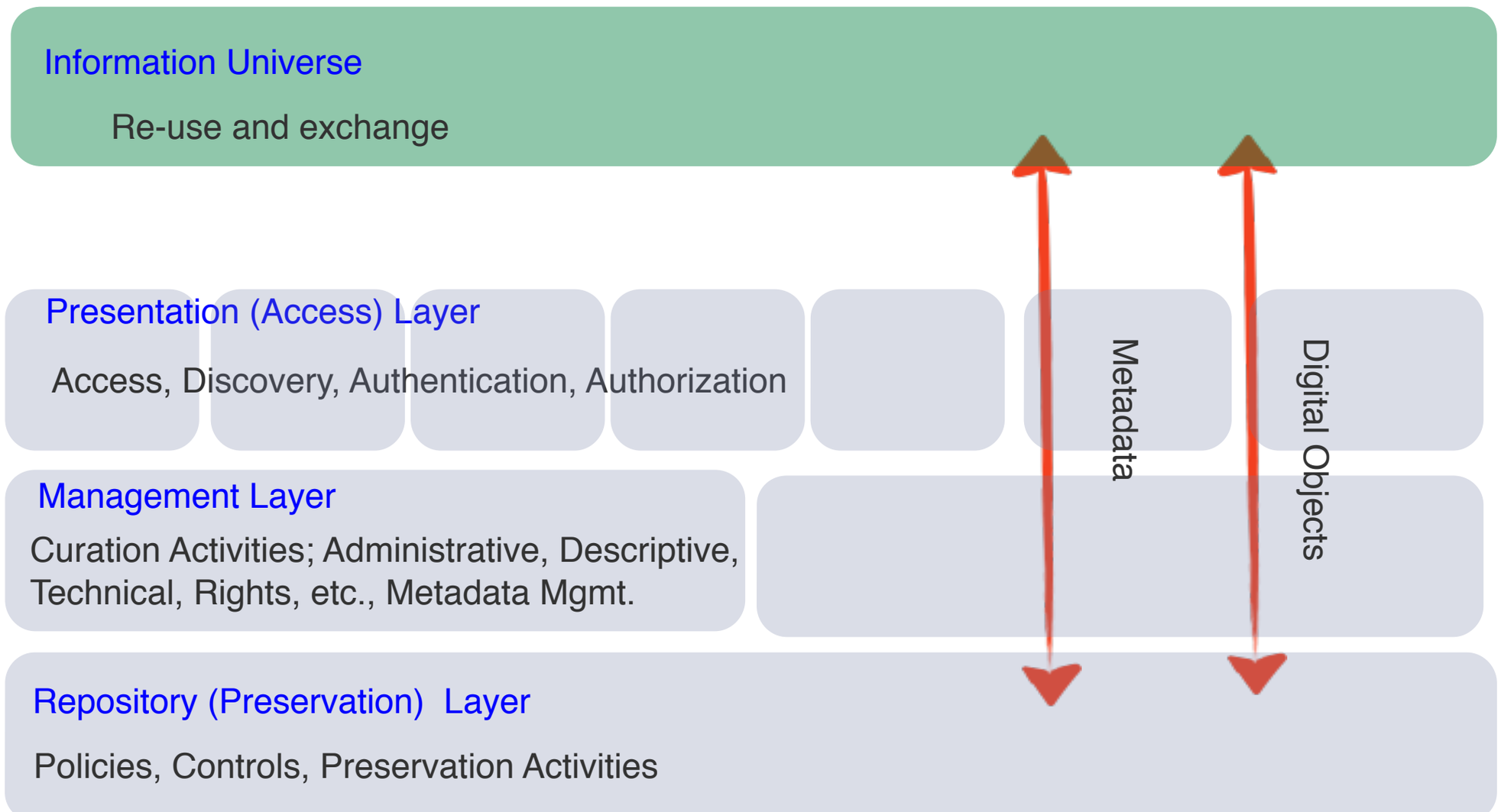
* Fedora, DSpace, ContentDM, Insight, Greenstone, Archivist Toolkit, DLXS, Documentum, Lightroom, Omeka, Islandora, Hydra, etc…

Software Reality

- No one software application will do everything well:
  - Manage digital objects and metadata
  - Provide public access
  - Permit easy re-purposing
  - Preserve digital assets
  - Apply and Manage policies
  - Make Quality choices
  - Determine Access controls

Software should be seen as a tool, not a “solution”

Layer view of a Digital Repository System



Let the Tools Do the Work

- Create it once, use it often
  - Central repository manages metadata, archival masters, and access derivatives for use in multiple presentation venues
  - Metadata is interoperable across multiple schemas, crosswalks are key
  - Automate activities as much as possible

Managing Digital Collections

Managing digital collections is the same as managing physical collections, except when it is different.

Respect the principles of the profession

Embrace the potential of technology

Thank you

Greg Colati

Director, Archives, Special Collections and Digital Curation

University of Connecticut

greg@uconn.edu