

# Overview of Digital Preservation Challenges and Opportunities

**Cal Lee**

**School of Information and Library Science  
University of North Carolina, Chapel Hill**

**SERI Institute**

**July 8-12, 2013**

**Indianapolis, Indiana**



**UNC**  
SCHOOL OF INFORMATION  
AND LIBRARY SCIENCE

# Outline

- Issues of Preserving Meaningful Information
- Nature of Digital Objects
- Layers and Abstraction
- Technology Obsolescence
- Significant Properties
- Technical Strategies
- Concluding Thoughts

# Origins

- Documentation of activities allows us to know about them without having to be there.
- Historically, contributions to this process have included oral communication, physically fixed artifacts and now digital systems.

# The Hermeneutic Gap

- All conveying of memory (even to ourselves) runs into a hermeneutic gap.
- Context is never captured or perpetuated completely.
- We use current understanding & place in the world to fill in gaps of previous contexts in order to make sense of memories.
- This is one of our greatest strengths as humans but also raises many issues related to concepts we cherish (e.g. truth, tradition, accuracy, accountability).

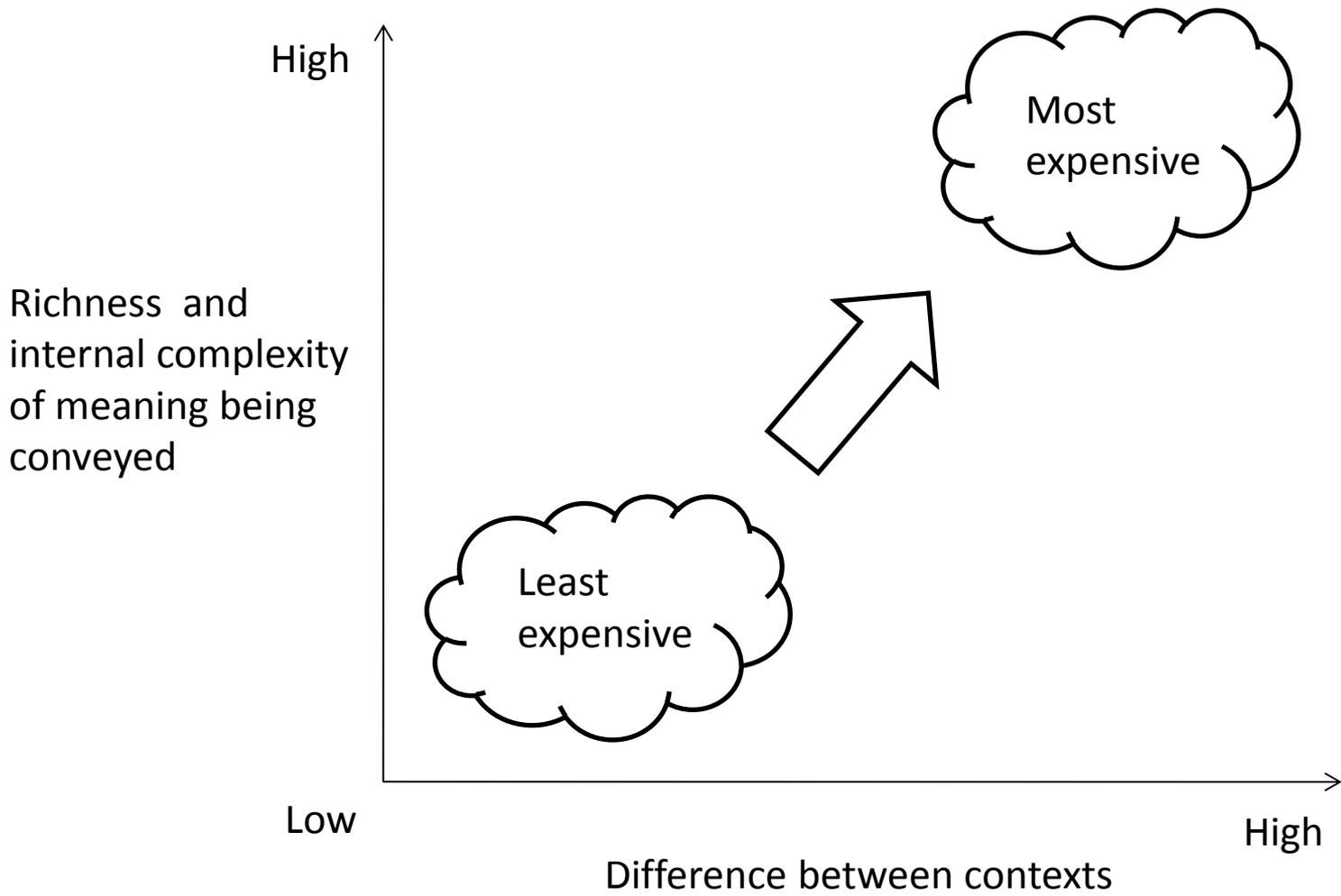
# Bridging the Gap (Sort of)

Information professionals work to bridge the gap through:

- adding metadata into the system (filing cabinets, policies) & at point of creation (naming, filing, genre conventions)
- selection
- retention scheduling
- disposition actions
- transfer of custody to trusted third parties
- labor-intensive arrangement & description
- controlled custody environments
- one-on-one reference services

# Resources are Limited, Meaning is Expensive

- Always true, but increasingly important in a digital environment
- Two often competing demands:
  - more heterogeneous access (any type of client can access any type of object)
  - more functionality (each object becomes increasingly complex, thus carrying more dependencies)



# Bits will be Bits (But not for Long)

- Physical media should be stored in appropriate environmental conditions.
- Take care in handling of media.
- Maintain integrity of bit stream through security, checksums, periodic sampling and other validation
- Bit rot and advantages of newer media both call for **periodic refresh and reformatting**.
- Ensuring the **integrity of the bit stream** in such transfers is extremely important.

# Digital objects are sets of **instructions for future interaction**

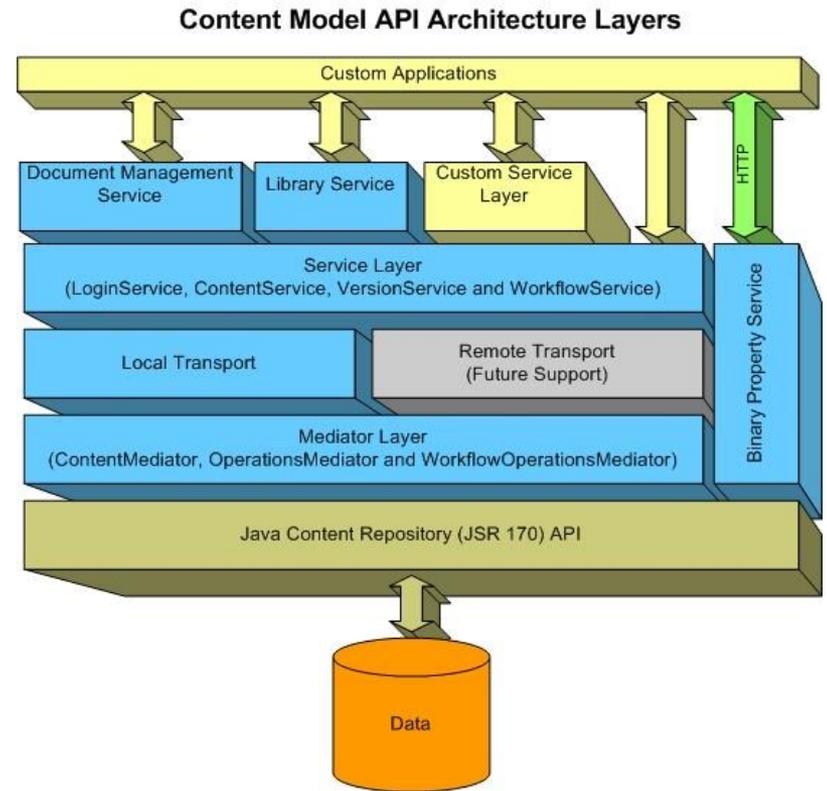
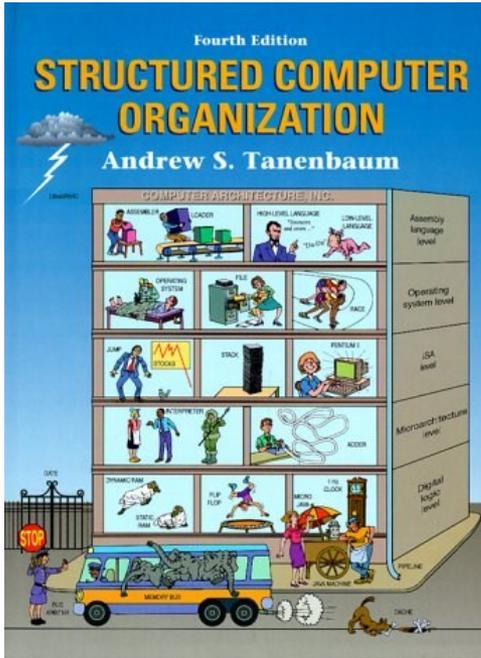
- Digital objects are useless (and don't even exist) if no one can interact with them
- Interactions depend on numerous technical components
- Only a small part of preservation work is about treating them like physical artifacts.

# Layers and Abstraction

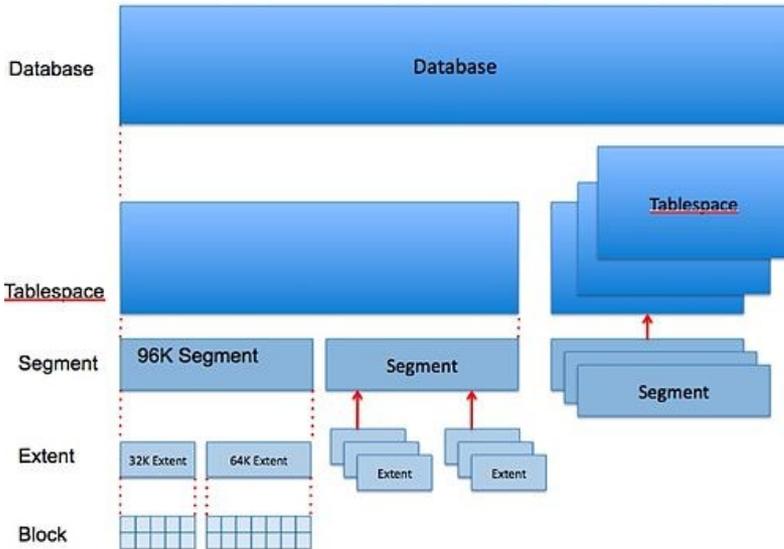
"Computer science is largely a matter of **abstraction**: identifying a wide range of applications that include some overlapping functionality, and then working to **abstract out** that shared functionality into a distinct service layer (or module, or language, or whatever). That new service layer then becomes a platform on top of which many other functionalities can be built that had previously been impractical or even unimagined. How does this activity of abstraction work as a practical matter? It's technical work, of course, but it's also **social work**. It is unlikely that any one computer scientist will be an expert in every one of the important applications areas that may benefit from the abstract service. So **collaboration** will be required." (emphasis added)

- Phil Agre, Red Rock Eater, March 25, 2000

# Layers, Layers Everywhere

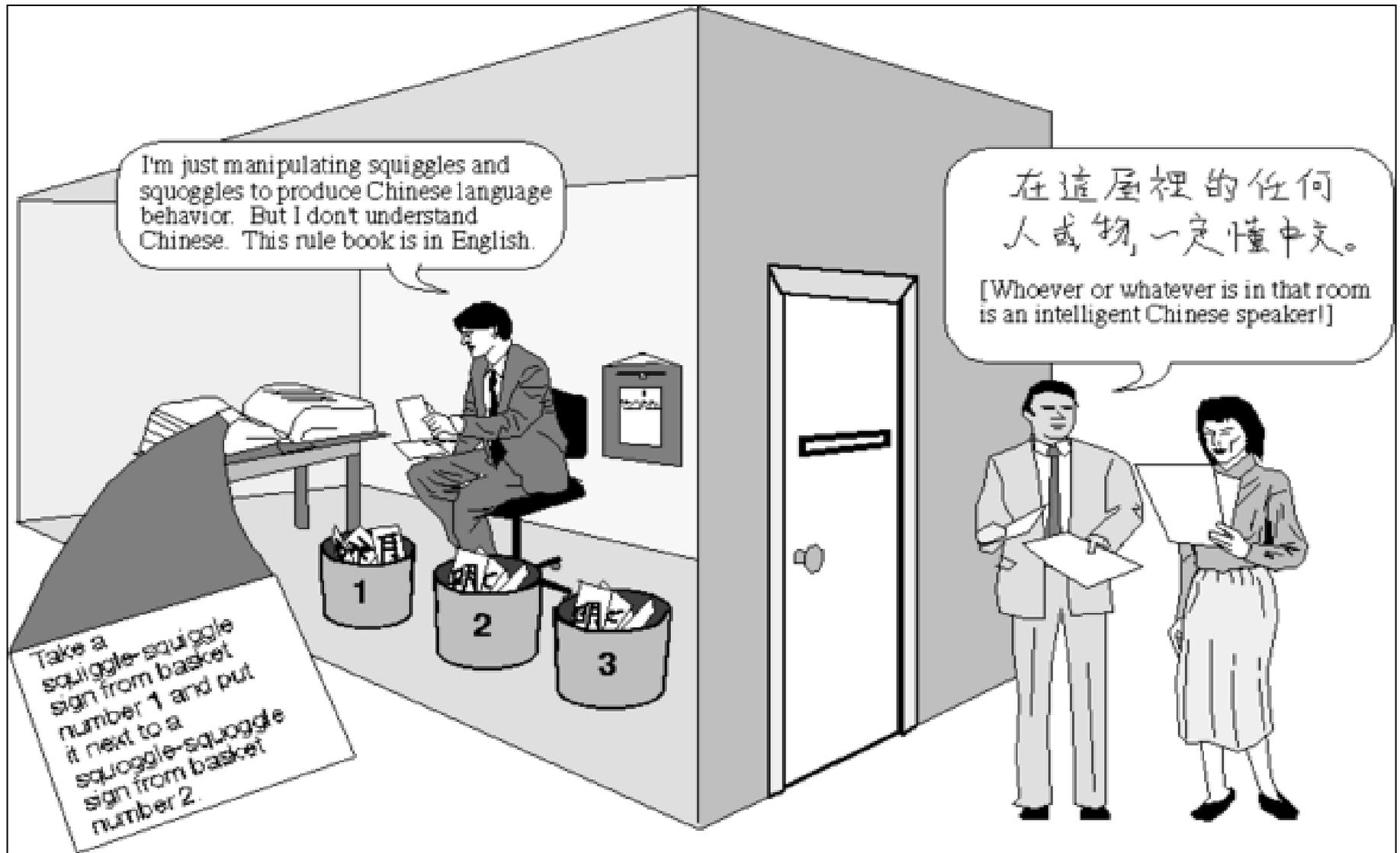


[http://www.ibm.com/developerworks/websphere/techjournal/0607\\_kubik/0607\\_kubik.html](http://www.ibm.com/developerworks/websphere/techjournal/0607_kubik/0607_kubik.html)



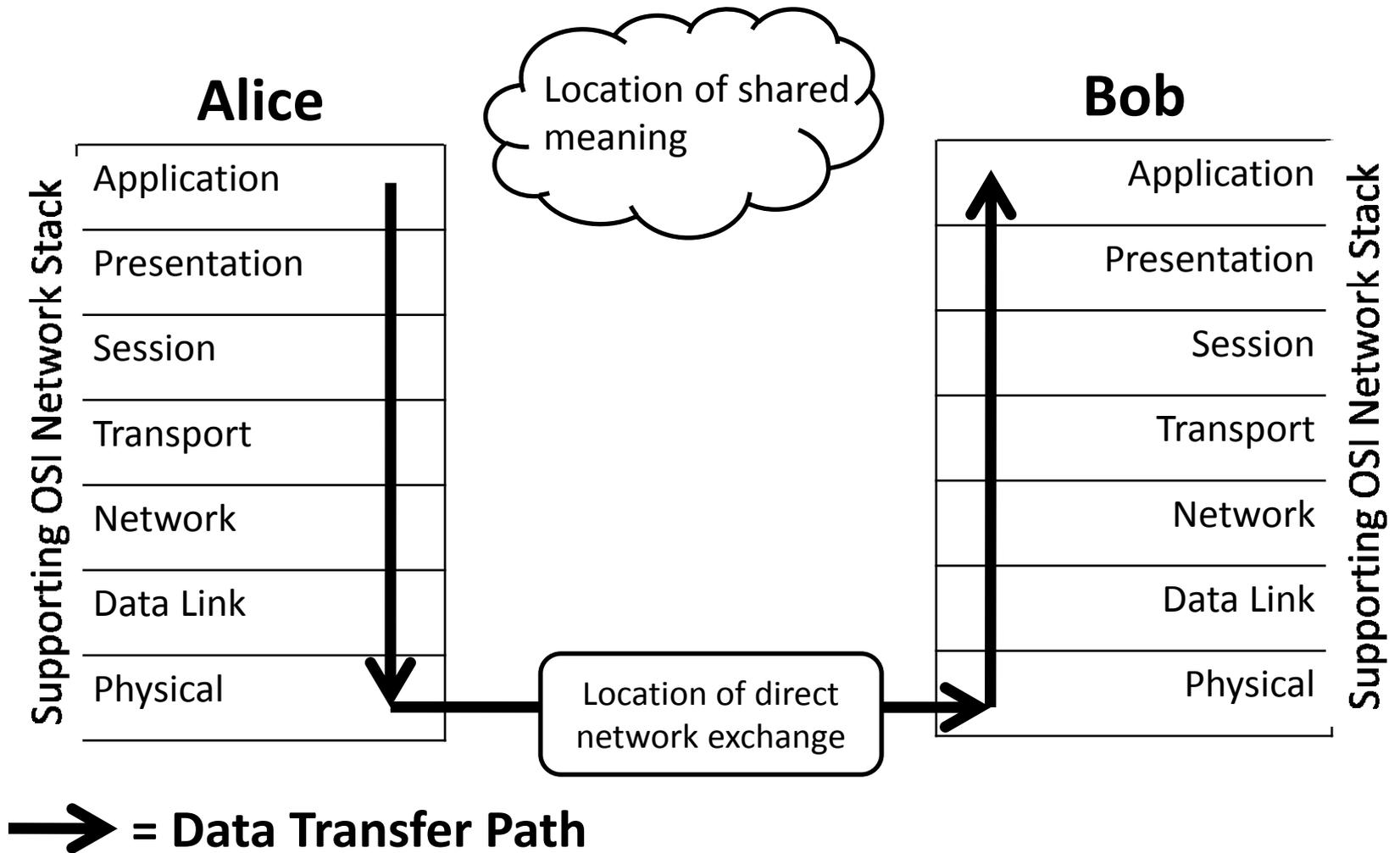
[http://upload.wikimedia.org/wikipedia/commons/thumb/c/c7/Oracle\\_Storage\\_Hierarchy.jpg/500px-Oracle\\_Storage\\_Hierarchy.jpg](http://upload.wikimedia.org/wikipedia/commons/thumb/c/c7/Oracle_Storage_Hierarchy.jpg/500px-Oracle_Storage_Hierarchy.jpg)

# John Searle's "Chinese Room" – An Input-Output Scenario Involving Abstraction



Source: MacroVU, Inc. Mapping Great Debates: Can Computers Think?  
<http://www.macrovu.com/CCTMap4ChineseRm.html>

# Implied Communication Dynamics of Open Systems Interconnection (OSI) Network Model



# Digital Resources - Levels of Representation

Level	Label	Explanation
8	Aggregation of objects	Set of objects that form an aggregation that is meaningful encountered as an entity
7	Object or package	Object composed of multiple files, each of which could also be encountered as individual files
6	In-application rendering	As rendered and encountered within a specific application
5	File through filesystem	Files encountered as discrete set of items with associate paths and file names
4	File as “raw” bitstream	Bitstream encountered as a continuous series of binary values
3	Sub-file data structure	Discrete “chunk” of data that is part of a larger file
2	Bitstream through I/O equipment	Series of 1s and 0s as accessed from the storage media using input/output hardware and software (e.g. controllers, drivers, ports, connectors)
1	Raw signal stream through I/O equipment	Stream of magnetic flux transitions or other analog electronic output read from the drive without yet interpreting the signal stream as a set of discrete values (i.e. not treated as a digital bitstream that can be directly read by the host computer)
0	Bitstream on physical medium	Physical properties of the storage medium that are interpreted as bitstreams at Level 1

# Interaction Examples

Level	Examples
Aggregation of objects	Browsing the contents of an archival collection using a finding aid
Object or package	Viewing a web page that contains several files, including HTML, a style sheet and several images
In-application rendering	Using Microsoft Excel to view an .xls file, watching an online video by using a Flash viewer
File through filesystem	Viewing contents of a folder using Windows Explorer, typing "ls" at the Unix command prompt to show the contents of a directory
File as "raw" bitstream	Opening an individual file in a hex editor
Sub-file data structure	Extracting a tagged data element in an XML document or value of a field in a relational database
Bitstream through I/O equipment	Connecting a hard drive to a host computer and then generating a sector-by-sector image of the disk using Unix dd command
Raw signal stream through I/O equipment	Connecting a floppy drive to a host computer and then generating a magnetic flux transition image of the disk
Bitstream on physical medium	Using a high-power microscope and camera to take a picture of the patterns of magnetic charges on the surface of a hard drive or pits and lands on an optical disk

*NOTE: More about access and use on Thursday*

# Interaction Examples

## Level

## Examples

Aggregation of objects

Object or package

In-application rendering

File through filesystem

File as "raw" bitstream

Sub-file data structure

Bitstream through I/O equipment

Raw signal stream through I/O equipment

Bitstream on physical medium



[\[Home\]](#)[\[Publications\]](#)[\[Reports\]](#)[\[Add\]](#)[\[View\]](#)[\[Search\]](#)[\[Profile\]](#)[\[Visualize\]](#)[\[Monitor\]](#)[\[Tools\]](#)[\[Developer\]](#)

This page lists all the seed queries that are used for monitoring videos related to elections on YouTube. Clicking on a query will show all the results collected over several crawls. Total number of these results are also listed here for each query. The last column in the following table shows how many total results YouTube had for a given query during our latest crawl. Clicking on 'Setup' associated with a query will bring up an interface where the curator can specify what constitutes as a "significant" change for a video of that query.

#	Query	Setup	Total results so far	Max results on last crawl
1	<a href="#">election 2008</a>	<a href="#">Setup</a>	574	6150
2	<a href="#">US election 2008</a>	<a href="#">Setup</a>	349	795
3	<a href="#">United States election 2008</a>	<a href="#">Setup</a>	216	257
4	<a href="#">presidential election 2008</a>	<a href="#">Setup</a>	206	1820
5	<a href="#">campaign 2008</a>	<a href="#">Setup</a>	273	2530
6	<a href="#">decision 2008</a>	<a href="#">Setup</a>	168	142
7	<a href="#">Joe Biden</a>	<a href="#">Setup</a>	209	1080
8	<a href="#">Hillary Rodham Clinton</a>	<a href="#">Setup</a>	193	353
9	<a href="#">Christopher Dodd</a>	<a href="#">Setup</a>	267	815
10	<a href="#">John Edwards</a>	<a href="#">Setup</a>	902	7540
11	<a href="#">Mike Gravel</a>	<a href="#">Setup</a>	301	1210
12	<a href="#">Dennis Kucinich</a>	<a href="#">Setup</a>	229	1600
13	<a href="#">Barack Obama</a>	<a href="#">Setup</a>	861	9140
14	<a href="#">Bill Richardson</a>	<a href="#">Setup</a>	287	1100
15	<a href="#">Wesley Clark</a>	<a href="#">Setup</a>	191	375
16	<a href="#">Al Gore</a>	<a href="#">Setup</a>	613	4910
17	<a href="#">Tom Vilsack</a>	<a href="#">Setup</a>	89	68
18	<a href="#">Sam Brownback</a>	<a href="#">Setup</a>	254	404
19	<a href="#">John McCain</a>	<a href="#">Setup</a>	22	16

of the patterns of magnetic charges on the surface of a hard drive or pits and lands on an optical disk

# Interaction Examples

**Level**

**Examples**

Aggregation of objects

Browsing the contents of an archival collection using a finding

**Object or package**

In-application rendering

File through filesystem

File as "raw" bitstream

Sub-file data structure

Bitstream through I/O equipment

Raw signal stream through equipment

Bitstream on physical media



[\[Home\]](#)
[\[Publications\]](#)
[\[Reports\]](#)
[\[Add\]](#)
[\[View\]](#)
[\[Search\]](#)
[\[Profile\]](#)
[\[Visualize\]](#)
[\[Monitor\]](#)
[\[Tools\]](#)
[\[Developer\]](#)

This page presents contextual information for a video captured over a number of days. Contextual information is defined as the information about a video that may change with time. Usually this information is contributed by the visitors of the video page. [See](#) the metadata information for this video. Description of various attributes displayed is given [here](#).



Query: *Rudy Giuliani*

[I Got A Crush On.... Giuliani](#)

Collaboration with the very talented JackDanyells, who came up with the concept for this video. Check out his channel at: <http://www.youtube.com/jackdanyells> -Lyrics by JackDanyells -Vocal melody composed and sung by me -Royalty free background music from sounddogs.com

Comedy

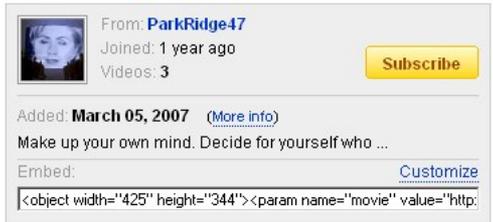
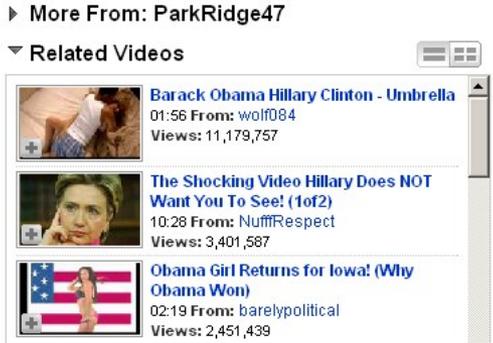
Crawling since 2007-07-19

Color coding for % changes

< 0.05 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0 5.0 >

Crawl #	Crawl date	Rank	Views	Ratings	Avg Rating	Comments	Links	Favorited	Honors	Change
1	2007-07-31	5	27357	301	3.74	288	5	44	0	--
2	2007-08-01	5	27452	303	3.73	290	5	44	0	--
3	2007-08-02	5	27780	307	3.72	291	5	45	0	--
4	2007-08-03	5	28048	309	3.71	291	5	45	0	--
5	2007-08-04	2	28398	310	3.71	291	5	45	0	--
6	2007-08-05	2	28443	314	3.69	294	5	45	0	--
7	2007-08-06	3	28980	314	3.69	296	5	45	0	--
8	2007-08-07	3	29265	318	3.65	298	5	45	0	--
9	2007-08-08	3	29551	319	3.65	299	5	46	0	--
10	2007-08-09	3	30094	320	3.64	300	5	47	0	--
11	2007-08-10	3	30384	323	3.61	302	5	47	0	--
12	2007-08-10	5	30419	324	3.62	303	5	48	0	--
13	2007-08-11	3	30540	324	3.62	305	5	49	0	--
14	2007-08-12	3	30697	326	3.61	306	5	49	0	--
15	2007-08-13	3	30848	326	3.61	306	5	49	0	--
16	2007-08-14	3	31036	326	3.61	306	5	49	0	--
17	2007-08-15	2	31181	326	3.61	306	5	49	0	--
18	2007-08-16	2	31321	326	3.61	307	5	51	0	--
19	2007-08-17	2	31459	327	3.61	307	5	51	0	--
20	2007-08-18	2	31662	331	3.59	308	5	51	0	--
21	2007-08-19	2	31792	332	3.58	308	5	51	0	--
22	2007-08-20	2	31937	335	3.57	310	5	51	0	--
23	2007-08-21	2	32135	335	3.57	311	5	52	0	--
24	2007-08-22	2	32304	335	3.57	311	5	52	0	--

# Interaction Examples

Level	Examples
Aggregation of objects	Browsing the contents of an archival collection using a finding
Object or package	
<b>In-application rendering</b>	
File through filesystem	
File as "raw" bitstream	
Sub-file data structure	
Bitstream through I/O equipment	
Raw signal stream through I/O equipment	Connecting a floppy drive to a host computer and then generating a magnetic flux transition image of the disk
Bitstream on physical medium	Using a high-power microscope and camera to take a picture of the patterns of magnetic charges on the surface of a hard drive or pits and lands on an optical disk

```

C:\WINDOWS\system32\cmd.exe
Microsoft Windows XP [Version 5.1.2600]
(C) Copyright 1985-2001 Microsoft Corp.

G:\>dir /a
Volume in drive G is KINGSTON
Volume Serial Number is 17E9-242F

Directory of G:\

03/12/2009  08:54 AM                4,096  ._.Trashes
03/12/2009  08:54 AM          <DIR>      .Trashes
03/12/2009  08:54 AM          <DIR>      .Spotlight-V100
03/11/2009  07:07 PM    1,023,213  nc-busmodels-jpw2009.pptx
03/12/2009  08:55 AM                4,096  .nc-busmodels-jpw2009.pptx
03/31/2009  01:23 PM    6,442,496  EMSS Meeting.ppt
              4 File(s)          7,473,901 bytes
              2 Dir(s)          120,145,920 bytes free

G:\>

```

**Level**

Aggregation of objects

Object or package

In-application rendering

**File through filesystem**

“ls” at directory

Opening

Extracting value

Connecting general communication

Connecting general communication

Using of the

Name
 .Spotlight-V100
 .Trashes
 ._.Trashes
 .nc-busmodels-jpw2009.pptx
 EMSS Meeting.ppt
 nc-busmodels-jpw2009.pptx

drive or pits and lands on an optical disk

# Interaction Examples

**Level**

Aggregation of objects

Object or package

In-application rendering

File through filesystem

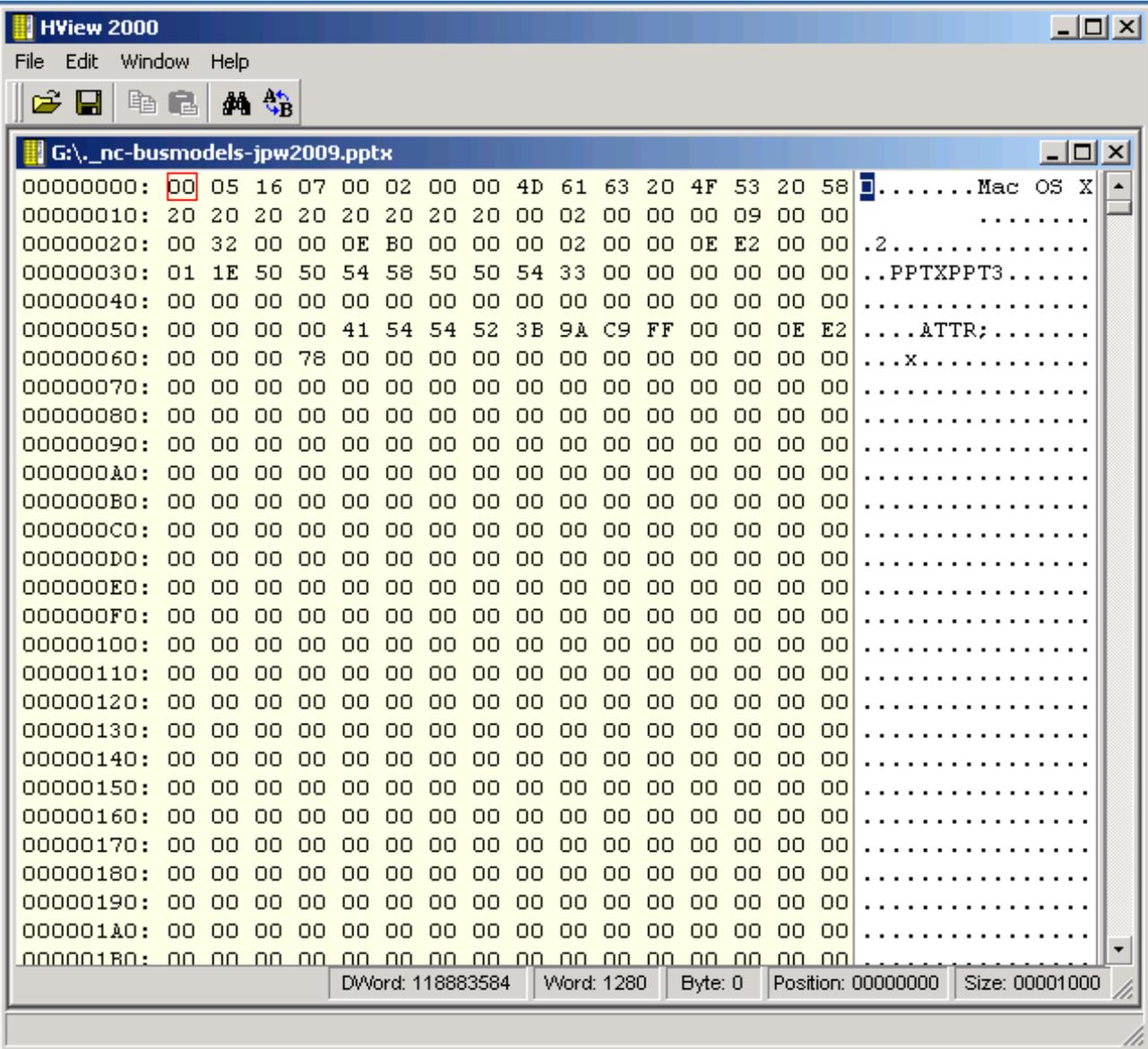
**File as “raw” bitstream**

Sub-file data structure

Bitstream through I/O equipment

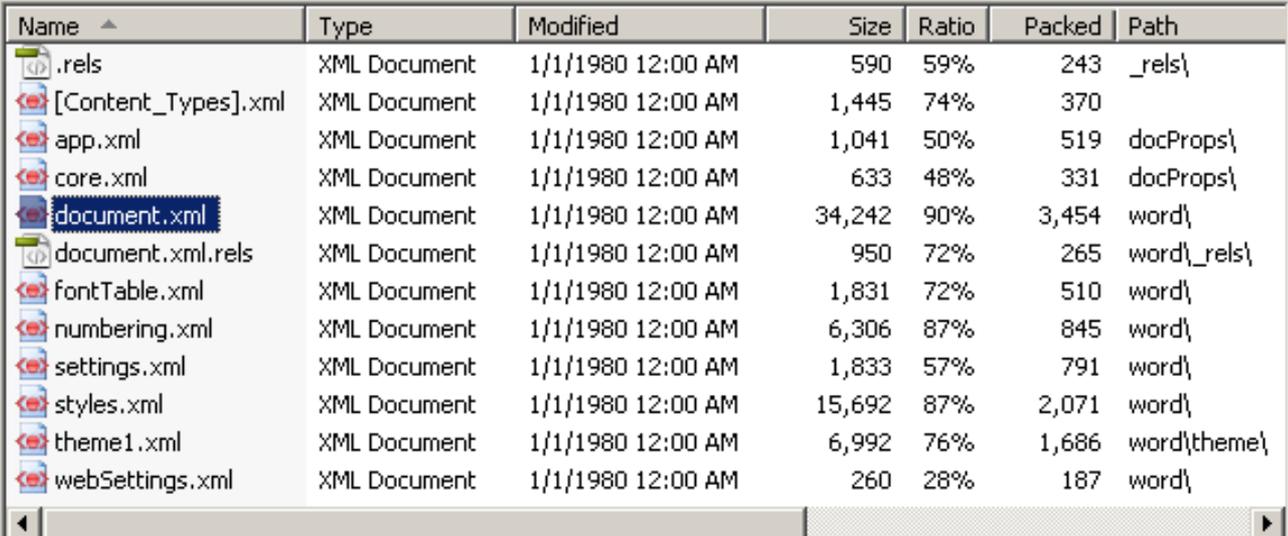
Raw signal stream through I/O equipment

Bitstream on physical medium



drive or pits and lands on an optical disk

# Interaction Examples

Level	Examples
Aggregation of objects	Browsing the contents of an archival collection using a finding aid
Object or package	
In-application rendering	
File through filesystem	
File as "raw" bitstream	
<b>Sub-file data structure</b>	
Bitstream through I/O equipment	
Raw signal stream through equipment	generating a magnetic flux transition image of the disk
Bitstream on physical medium	Using a high-power microscope and camera to take a picture of the patterns of magnetic charges on the surface of a hard drive or pits and lands on an optical disk

# Interaction Examples

**Level**

Aggregation of objects

Object or package

In-application rendering

File through filesystem

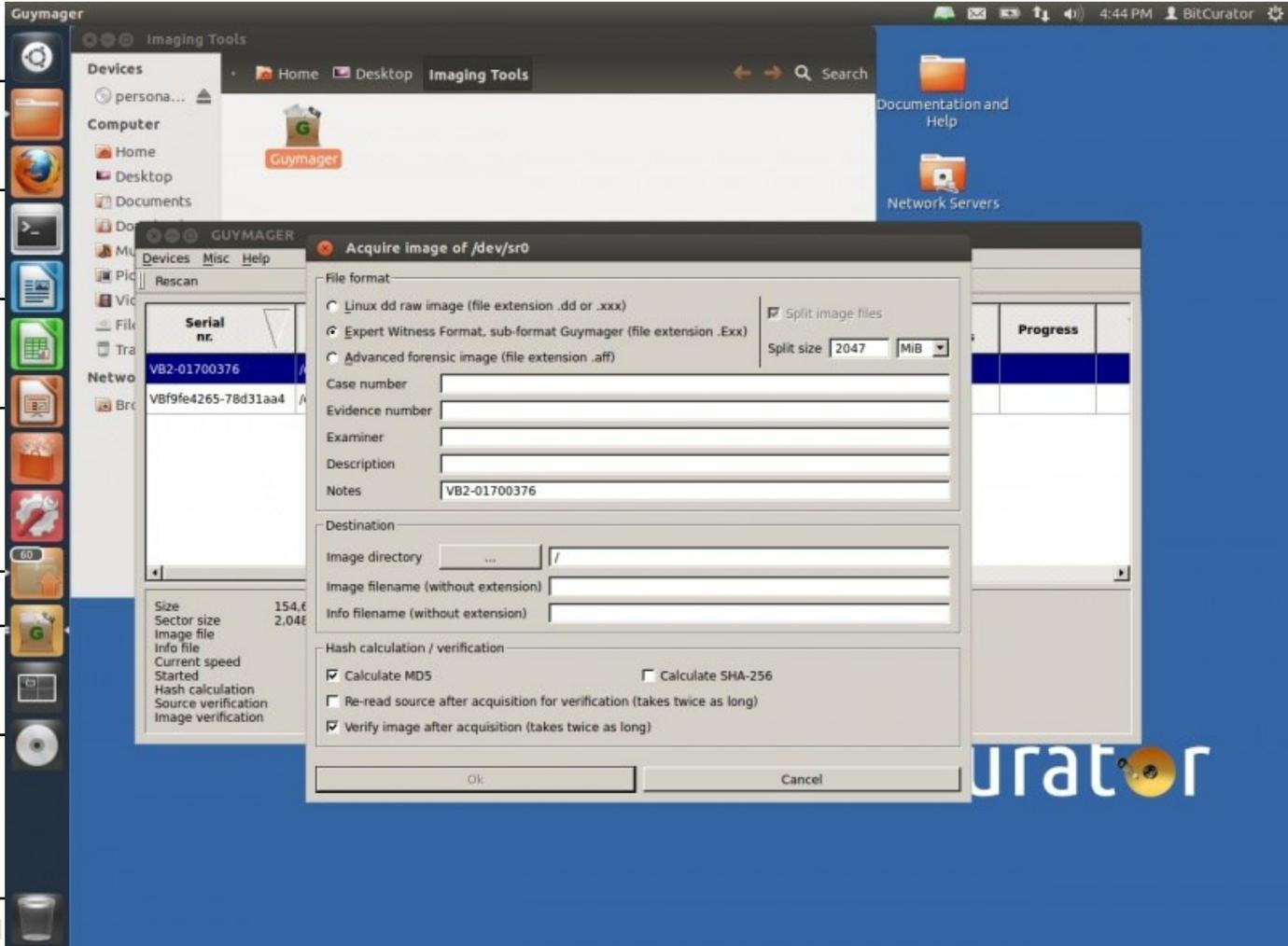
File as "raw" bitstream

Sub-file data structure

**Bitstream through I/O equipment**

Raw signal stream through equipment

Bitstream on physical medium



generating a magnetic flux transition image of the disk

Using a high-power microscope and camera to take a picture of the patterns of magnetic charges on the surface of a hard drive or pits and lands on an optical disk

**Level**

Aggregat

Object or

In-applic

File throu

File as "r

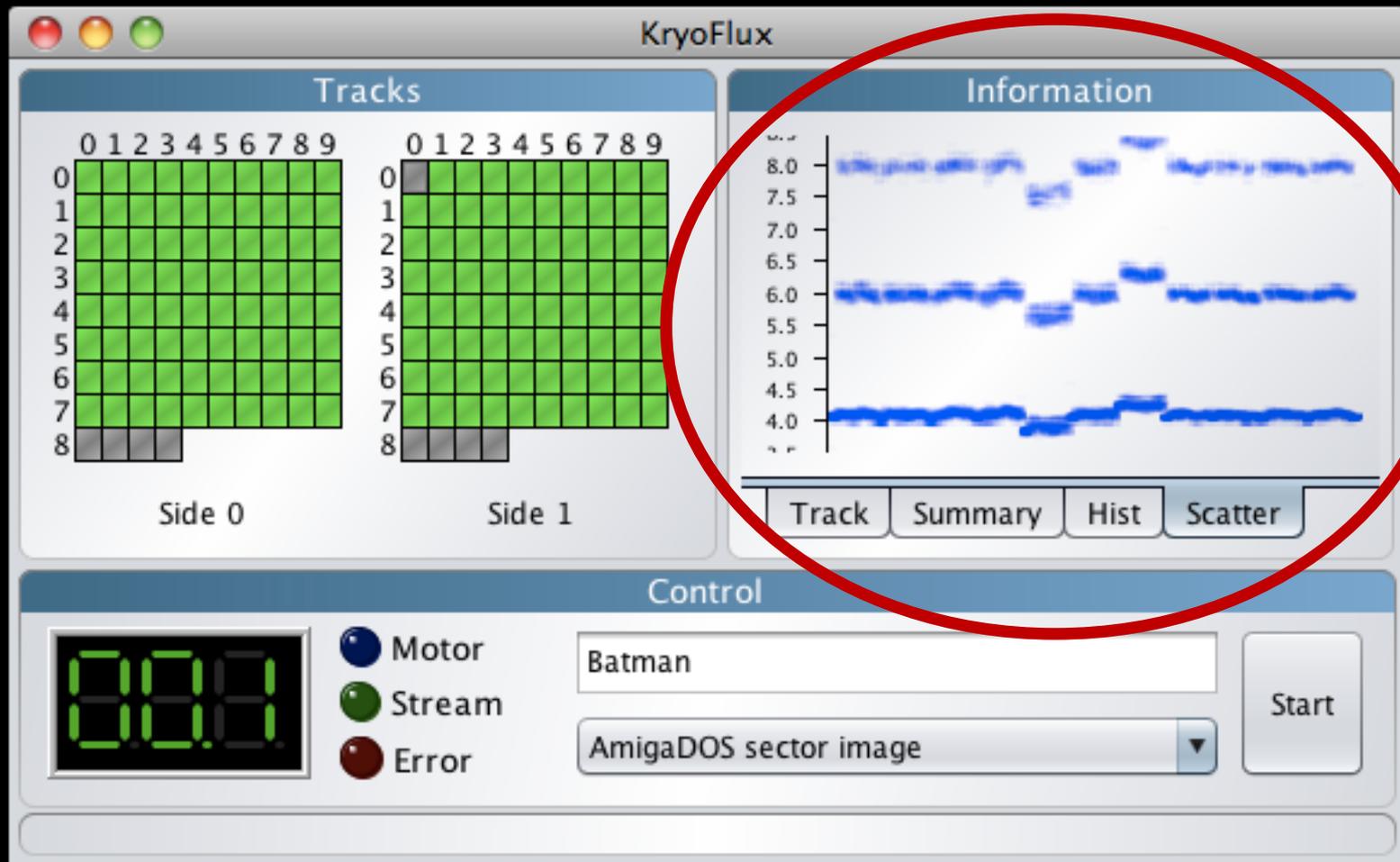
Sub-file c

Bitstream  
equipme

**Raw sign**

**I/O equipment**

Bitstream on physical medium



generating a magnetic flux transition image of the disk

Using a high-power microscope and camera to take a picture of the patterns of magnetic charges on the surface of a hard drive or pits and lands on an optical disk

# Interaction Examples

## Level

Aggregation of objects

Object or package

In-application rendering

File through filesystem

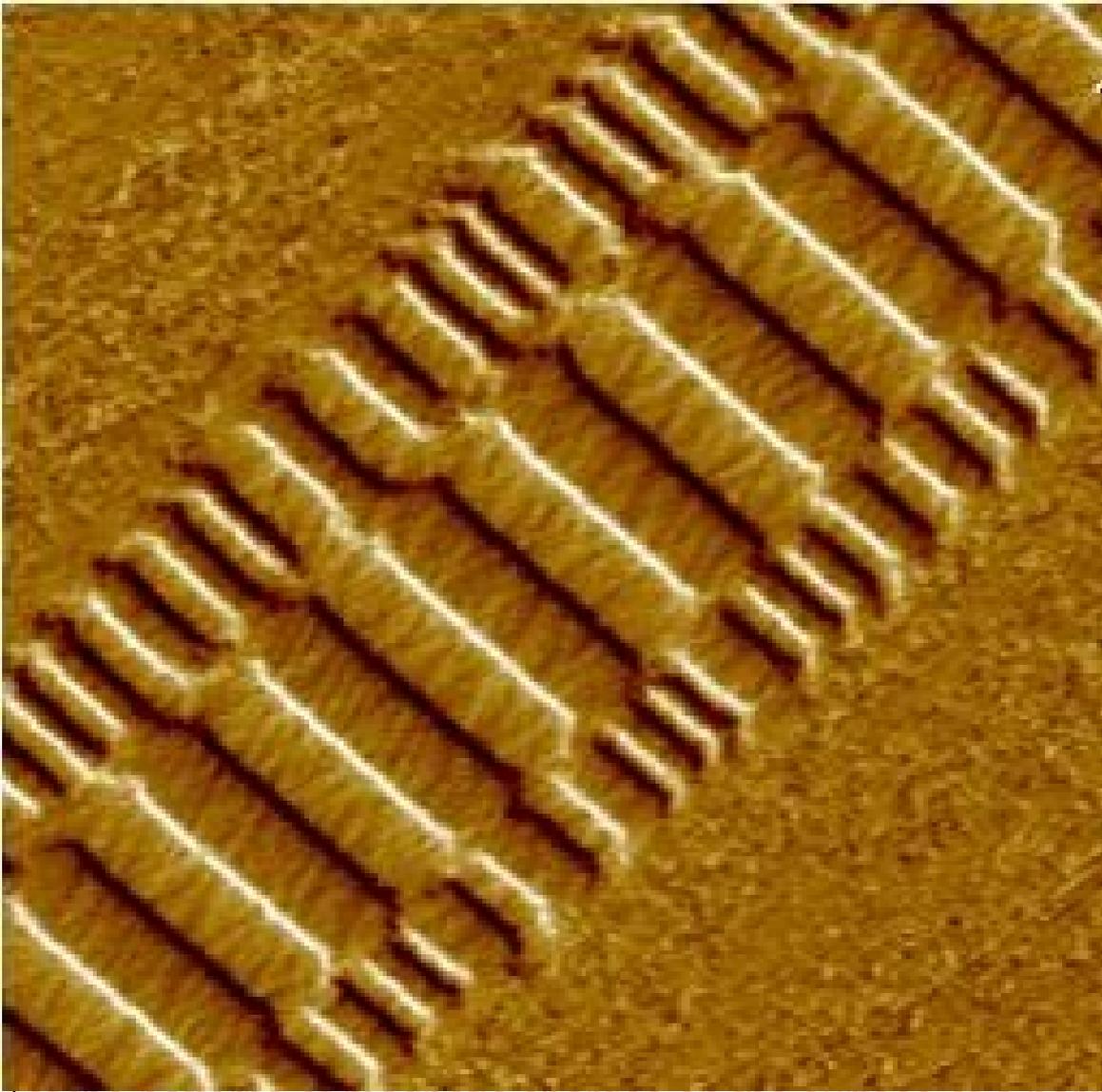
File as "raw" bitstream

Sub-file data structure

Bitstream through I/O equipment

Raw signal stream through I/O equipment

**Bitstream on physical medium**



Microscopic image of the patterns of magnetic charges on the surface of a hard disk platter.

# Stepping Above the Bits – Representation Information

# Representation Considerations

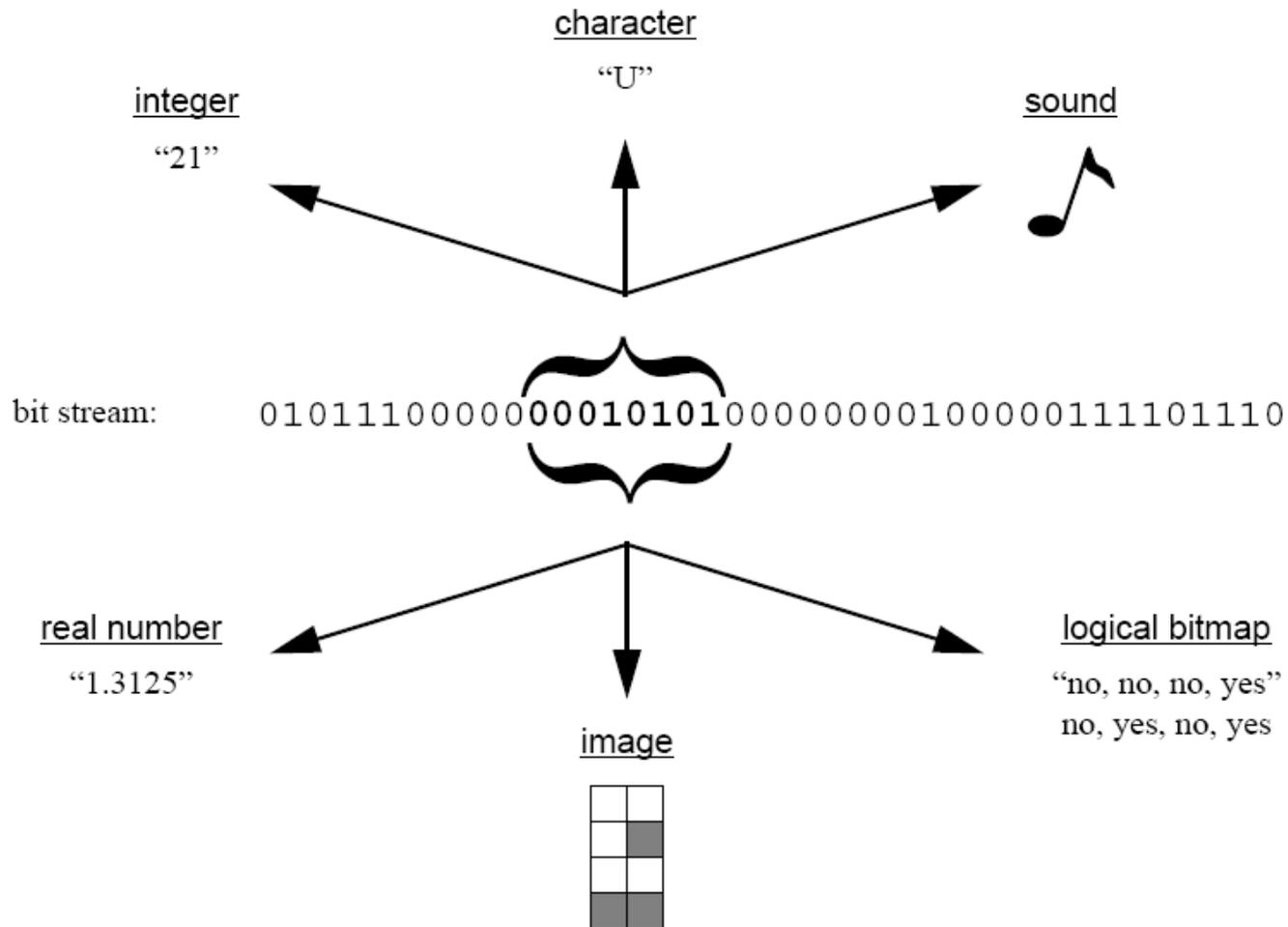
- Representation and interpretation are complementary – interpretation depends on representation
- Can have multiple interpretations of same representation, but...
- Some representation schemes make certain transformations and uses easier than others

# Representation at Multiple Levels

Every digital object is concurrently:

- *Physical* object – “inscription of signs on some physical medium”
- *Logical* object – “recognized and processed by software”
- *Conceptual* object – “recognized and understood by a person, or in some cases recognized and processed by a computer application capable of executing business transactions”

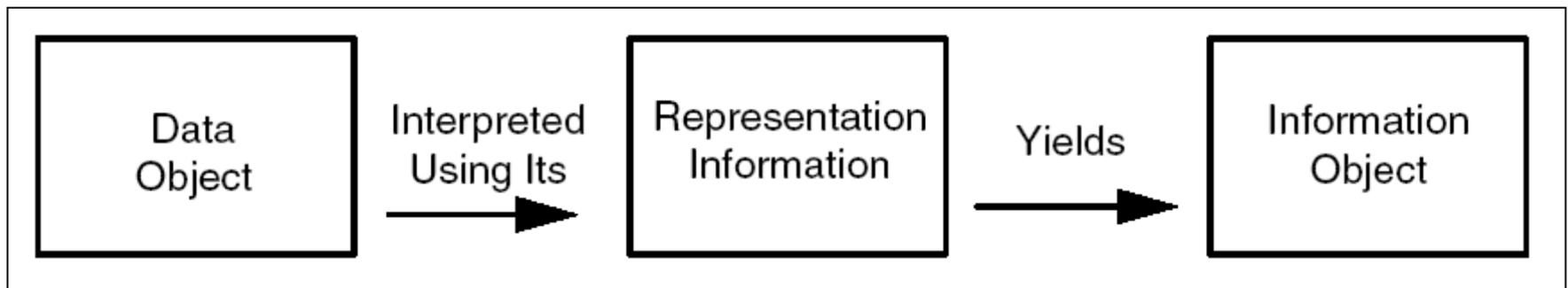
The 8 bits highlighted in the bit stream shown below can be interpreted in many ways, e.g., as an integer, a simple character code, a sound, a floating point number, an image, a logical bitmap, etc.



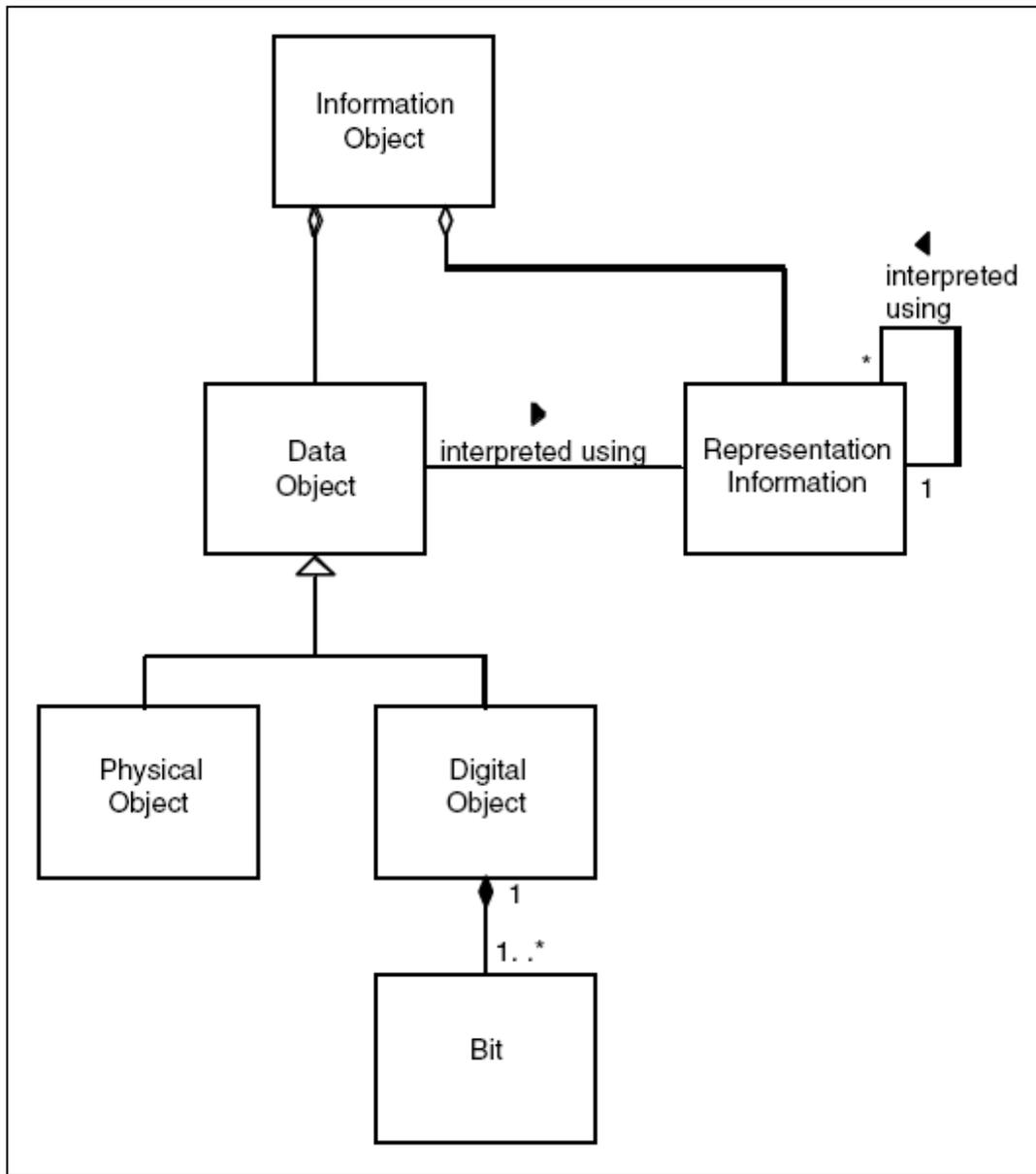
**Figure 4: A bit stream can represent anything at all**

# Representation Information

- “Information that maps a Data Object into more meaningful concepts” (OAIS) - makes humanly-perceptible properties happen
- Examples: file format, encoding scheme, data type



Reference Model for an Open Archival Information System (OAIS). Consultative Committee for Space Data Systems, 2002.: Figure 2-2



OAIS: Figure 4-10

# A Simple Example of Making Sense of a Bit Stream

011110011101001101001  
111010011011001100101011  
1110010011011101011110101110  
11011101101010101

# Hint: These are Bytes

0111100

1110100

1101001

1110100

1101100

1100101

0111110

010 0110

111 0101

111 0101

110 1101

110 1100

1010101

Add Byte-Level Encoding  
Knowledge:

This is ASCII

<title>&Uuml;

Add Data Structure Knowledge:

This is HTML

So text is intended to be a  
document title

Add further character encoding  
knowledge:

HTML supports ISO 8859-1  
character entities

&Uuml; = “latin capital letter U  
with diaeresis”

Add character rendering (font)  
information:

Using Arial font, “latin capital letter U  
with diaeresis” can be rendered as the  
following glyph:

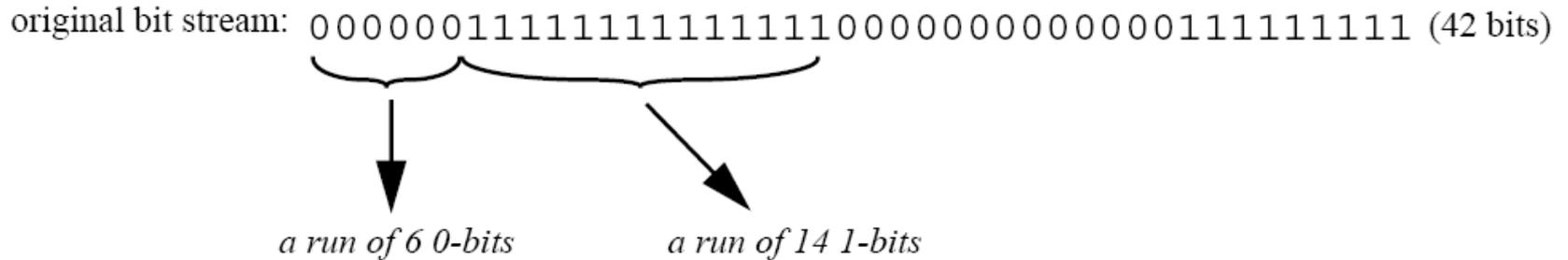
Ü

When you're staring at a bunch of bits, things will often be much messier than the previous example. Some complication factors...

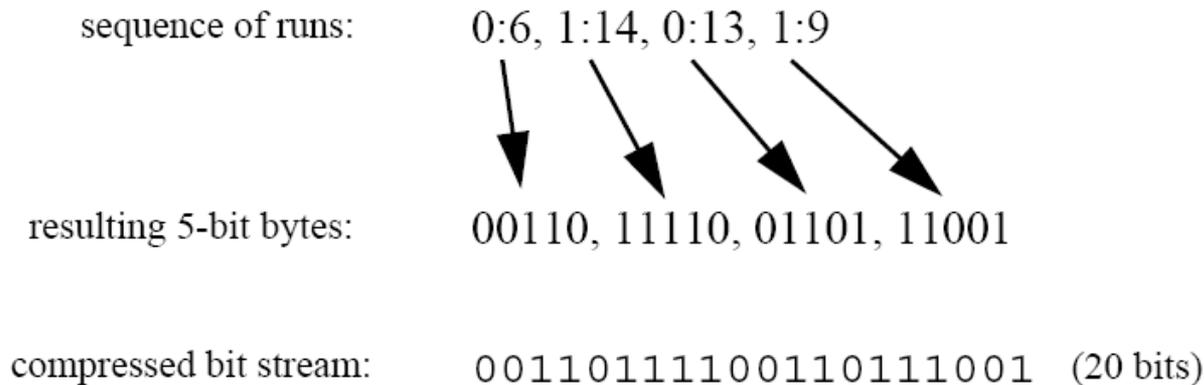
Formats that are proprietary or otherwise not sufficiently documented.

# Compression

As a simple example of compressing a bit stream without loss, “run-length encoding” replaces each sequence of 0s (000...0) by a count, indicating how many 0 bits were present in the given “run” (similarly for 1s). This can reduce the size of a bit stream without losing any information. For example, each run in the original bit stream shown can be represented by a 5-bit byte whose first bit specifies whether the run is of 0s or 1s and whose remaining 4 bits specify the length of a run (of up to 15 bits). This scheme is most appropriate for data that contains long sequences of 0s or 1s, such as digital imagery.



Representing each run in the original bit stream as a pair **b:n** (where b is 0 or 1 to indicate the contents of the run, and n is the length of the run) produces:



**Figure 7: Compressing a bit stream**

# Three Levels of Compression\*

- Format of file implements compression internally - e.g. body of JPEG file is compressed but not file header
- Application creates completely new, compressed copy of file(s) – e.g. WinZip, gzip
- File system compresses data units – e.g. not writing data to series of sectors that are all filled with zeros

\*Carrier, Brian. *File System Forensic Analysis*. Boston, MA: Addison-Wesley, 2005.

# Encryption

- Special data (“keys”) and algorithms used to transform data into a form that is purposely less easy to read
- Used for:
  - Confidentiality
  - Integrity
  - Non-repudiation
  - Authentication

# Encryption at Various Levels\*

- Application that creates the file
- Application that reads an unencrypted file and creates an encrypted file
- Operating System – “Before a file is written to disk, the OS encrypts the file and saves the cipher text to the data units. The non-content data, such as the file name and last access time, are typically not encrypted. The application that wrote the data does not know the file is encrypted on the disk.”
- Encrypt an entire volume – implemented in storage system below file system level

\*Carrier, Brian. *File System Forensic Analysis*. Boston, MA: Addison-Wesley, 2005.

# Robustness of File Formats

- Image compression is specifically designed to remove a certain form of redundancy (that which is presumed to be imperceptible or irrelevant to a user of the image)
- Header information can help to make sense of corrupted portions of a file
- Can be very helpful to also store header information outside of the file – corruption of header can otherwise seriously inhibit use of file
- More serialized formats – e.g. XML rather than a binary format – usually easier to recover (in whole or in part) after data errors

# Obsolescence

"Those who forget the past are  
condemned to reload it."

- Nick Montfort, July 2000

- All layers undergo change over time, at varying rates.

# New Conception of “Long-Term”

“A period of time long enough for there to be concern about the impacts of **changing technologies**, including support for new media and data formats, and of a **changing user community**, on the information being held in a repository.” (OAIS, emphasis added)

# Risks Associated with Obsolescence

- Vendor Lock-In
- Legacy Data
- Need for “Digital Archeology” (more about this tomorrow afternoon)

# Approaches to Preserving Layers of Meaning

- Make information useful
- Policies and procedures (periodically revisited and audited)
- Creators' and users' awareness of issues
- System development
- System administration
- Ongoing maintenance - copying, converting, reformatting, emulating, normalizing, migrating

# Significant Properties

- “Whoever takes the decision that a particular digital object should be preserved will have to decide what properties are to be regarded as significant. The submission agreement could usefully specify a list of significant properties.”<sup>1</sup>
- “properties of digital objects that affect their quality, usability, rendering, and behaviour”<sup>2</sup>
- Essence = “characteristics that must be preserved for the record to maintain its meaning over time”<sup>3</sup>

1. Holdsworth, David, and Derek M. Sergeant. "A Blueprint for Representation Information in the OAIS Model." Paper presented at the IEEE Symposium on Mass Storage Systems, College Park, Maryland, USA, March 27-30, 2000.

2. Hedstrom, Margaret, and Christopher A. Lee. "Significant Properties of Digital Objects: Definitions, Applications, Implications." In *Proceedings of the DLM-Forum 2002, Barcelona, 6-8 May 2002: @ccess and Preservation of Electronic Information: Best Practices and Solutions*, 218-27. Luxembourg: Office for Official Publications of the European Communities, 2002.

3. Heslop, Helen, Simon Davis, and Andrew Wilson. "An Approach to the Preservation of Digital Records." National Archives of Australia, 2002.

# Defining Significant Properties can Serve a Variety of Purposes

- Writing **specific** provisions into **submission agreements**
- Developing **criteria** and **empirical tools** for evaluating preservation approaches
- **Documentation of preservation decisions** in terms of specific properties
  - allowing professionals to revisit previous decisions
  - indicating to researchers what properties have not been retained

# Traditional Dichotomy: Emulation vs. Transformation/Migration

- **Emulation** – Use of software to imitate obsolete computer equipment on new computer equipment, i.e. trick files and applications into thinking they're still running in their original environment
- **Transformation/Migration** – Digital object that depends on obsolete computer equipment is changed in order to run directly on new equipment
- Advocates of emulation contend that it better supports notion of preserving an “original,” along with its “look and feel,” and it can be more cost-effective than repeated transformations of digital objects

# Emulation - *Oxford English Dictionary*, Second Edition

“To reproduce the action of or behave like (a different type of computer) with the aid of hardware or software designed to effect this; to run (a program, etc., written for another type of computer) by this means.”

# Migration

- Periodic transformation of the bits/bytes to run directly on newer platforms.
- Used widely as an approach to actively managing legacy systems.
- Work can be expensive and introduce errors of translation.
- Since the resulting objects can run directly on newer platforms, layers of technology can be minimized.

# Not Just “Emulation vs. Migration”

- All strategies use standards in some way
- General consensus to keep original bits
- Transformation can be minor or extensive
- Transformation/Emulation can take place in Producer environment, upon Ingest, as part of preservation activities within a repository, or at time of access

# Cost-Benefit Analysis of a Preservation Approach

- **Cost** = sum of all resources one must commit in order to carry it out
- **Benefits** = value one can derive from the digital objects that have been preserved based on that approach
- **Opportunity costs** = failure to derive benefits that one could have had by choosing a different approach

# Using Properties in Making Guesses about Benefits

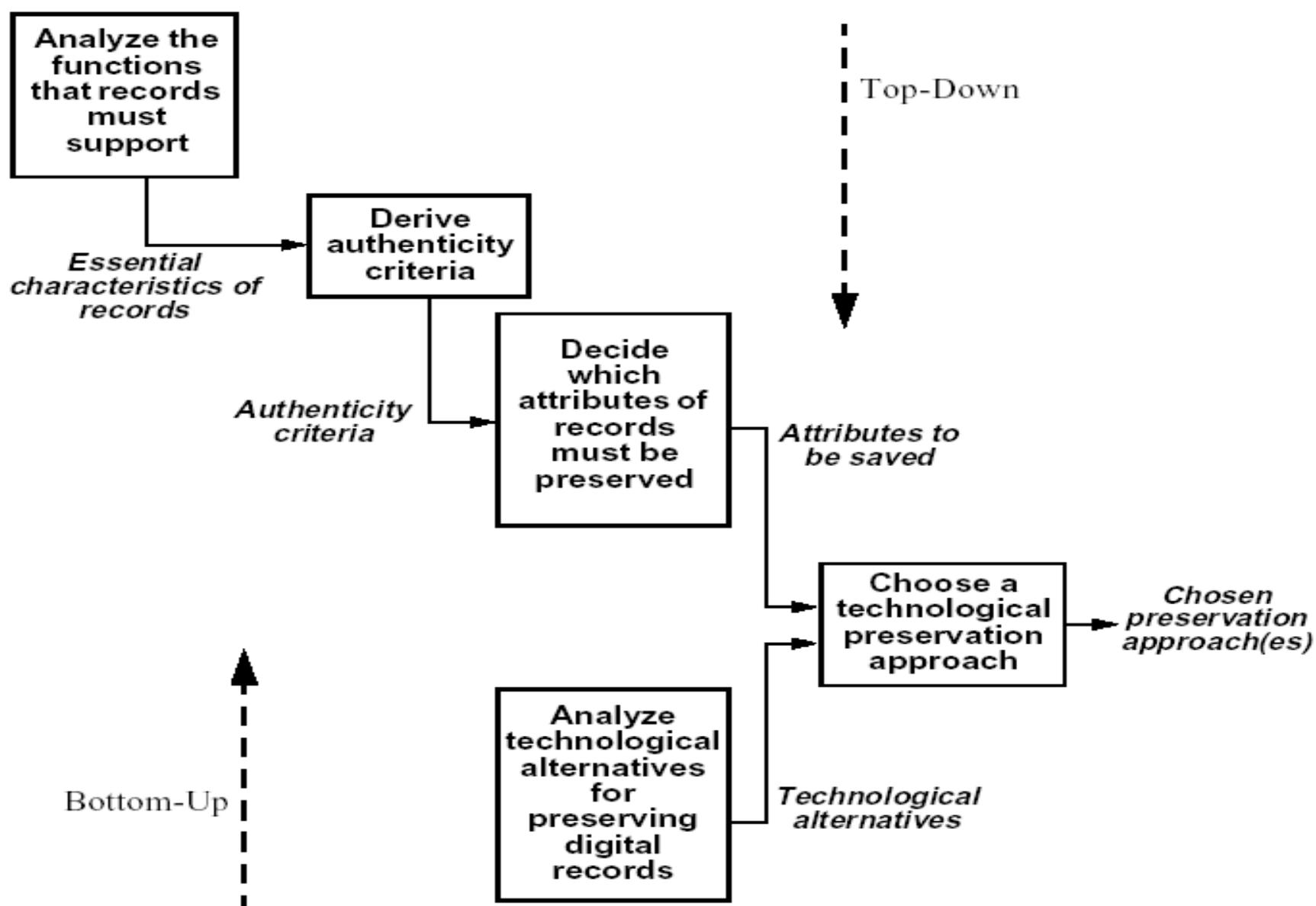
- Impossible to directly measure now the value of future use, so we must guess as to their expected value.
- Users derive value from digital objects by performing various high-level functions.
- Properties that facilitate those functions should have instrumental value.
- One would hope to preserve properties that serve the widest possible range of uses, though one may weigh some types of use more heavily than others.

# Punch Line 1: No Such Thing as Benign Neglect

- Ongoing preservation effort is assumed, regardless of the strategy adopted.
- Goal is to minimize (rather than eliminate) work and maximize the benefits.

# Punch Line 2: Identify What's **Desirable** & What's **Possible**

- Best, most informed guess about **how** objects will be **used**
- **Characteristics** that support such use
- Currently available **technical approaches**
- Whether using any given approach can **cost-effectively** preserve those characteristics
- All decisions should be **well-documented** and **revisited** periodically



**Figure 3: The preservation strategy**

*Source: Jeff Rothenberg and Tora K. Bikson, "Carrying Authentic, Understandable and Usable Digital Records through Time", 1999*

Digital preservation is not an all-or-nothing proposition.

# Categories of Service

- Categories of contributors or content types that carry similar set of promises for:
  - Amount of work before and during accession process
  - Validation
  - Creation of metadata and documentation
  - Intellectual property protection
  - Security
  - Access Controls
  - Long-term preservation of content, contextual information, structure, behavior

# Factors Defining Categories

- **Institutional arrangements and agreements**
- Scope and policies for appraisal and collection development
- Complexity of digital objects and relationships
- Significant properties of objects and relationships
- File formats: availability of documentation of formats, openness and industry support
- Availability and sustainability of:
  - Technical resources (hardware, software, systems)
  - Human resources (attention and expertise)

# Examples and Sources Related to Levels of Service

- “Developing Permanence Levels and the Archives for NLM’s Permanent Web Documents.” U.S. National Library of Medicine. November 2007.  
<http://www.nlm.nih.gov/psd/pcm/devpermanence.html>
- “Format Support.” DSpace@MIT. Massachusetts Institute for Technology.  
<http://libraries.mit.edu/dspace-mit/build/policies/format.html>
- Lavoie, Brian F. "The Incentives to Preserve Digital Materials: Roles, Scenarios, and Economic Decision-Making." Dublin, OH: OCLC Research, 2003. [See “High-end” – perpetual access, preserving “look and feel” vs. “Low-end” – short-term, “intellectual content only”]
- LeFurgy, William G. "Levels of Service for Digital Repositories." *D-Lib Magazine* 8, no. 5 (2002).
- “NDSA Levels of Preservation.” National Digital Stewardship Alliance, 2013. <http://www.digitalpreservation.gov/ndsa/activities/levels.html>

# To care responsibly for digital collections over time:

- Intervene to prevent bits from becoming unreadable or corrupted
- Know much more about the underlying technology than a typical end user does

# Concluding Remarks

- Don't wait for a single, ultimate solution to emerge.
- The pieces of the puzzle are in place to build preservation environments.
- Go forth and preserve!