

Levels of Representation in Digital Collections: A Framework and Implications for Archival Research

Cal Lee

School of Information and Library Science
University of North Carolina at Chapel Hill

5th Annual SAA Research Forum
Chicago, IL
August 23, 2011



UNC
SCHOOL OF INFORMATION
AND LIBRARY SCIENCE



What makes computers different
from us?

“No computation without
representation.”

- Brian Cantwell Smith*

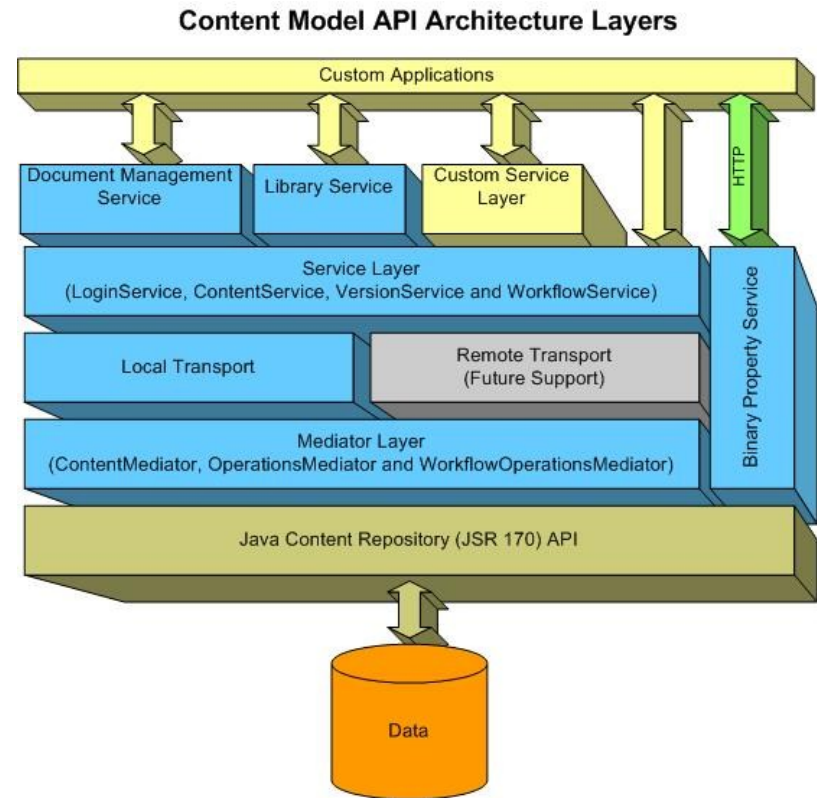
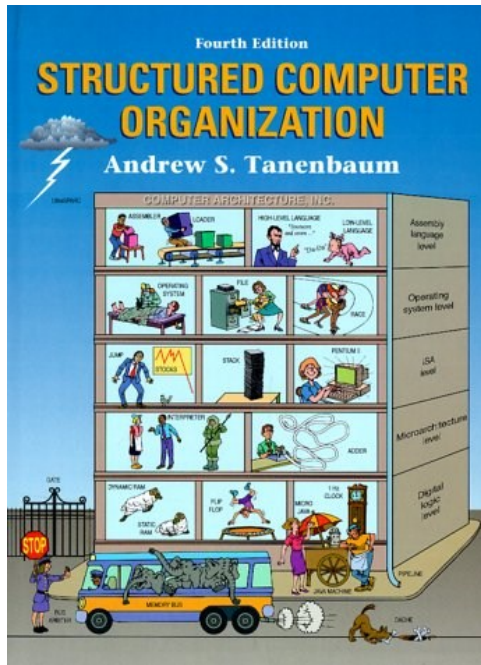
*"Limits of Correctness in Computers." In Computerization and Controversy: Value Conflicts and Social Choices, edited by Rob Kling, 810-25. San Diego, CA: Academic Press, 1996. 815.

Layers and Abstraction

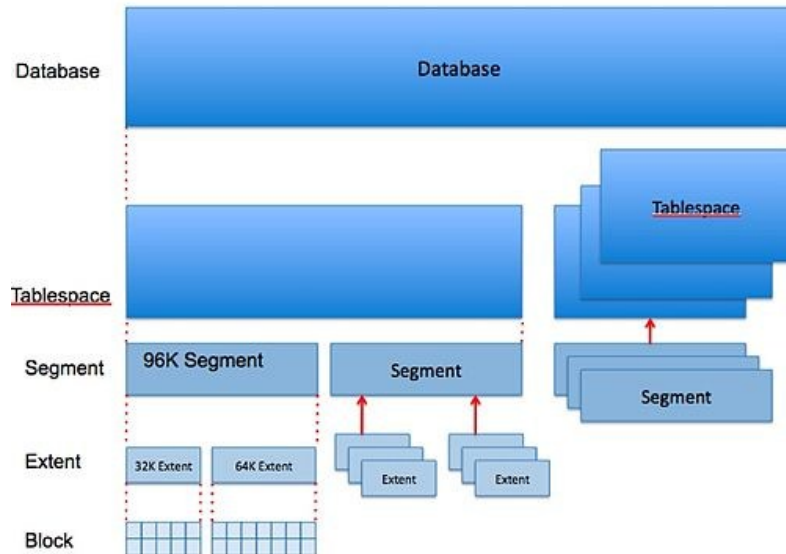
"Computer science is largely a matter of **abstraction**: identifying a wide range of applications that include some overlapping functionality, and then working to **abstract out** that shared functionality into a distinct service layer (or module, or language, or whatever."
(emphasis added)

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

Layers, Layers Everywhere



http://www.ibm.com/developerworks/websphere/techjournal/0607_kubik/0607_kubik.html



http://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>

Great News:

No single layer has to understand or retain all that is meaningful about a digital object.

In fact, the components of computer systems can be designed so that no other component has to “know” how it works internally.

More complicated news:

Because meaningful information can reside at all levels, the archivist must decide what level(s) to preserve in order to reflect the intended electronic records.

The “Lifting Problem” for Digital Archives

- Reuse of a digital object requires “lifting”¹ it out of its original context, then making sense and use of it in a new context.
- Support for lifting requires a proper balance between:
 - Providing too little contextual information, so that the user does not understand what she is interacting with, and
 - too much contextual information, so that she “will drown in unnecessary, unhelpful, or conflicting data”²

1. Guha, R., and John McCarthy. "Varieties of Contexts." In *Modeling and Using Context: 4th International and Interdisciplinary Conference Context 2003 Stanford, CA, USA, June 23-25, 2003 Proceedings*, edited by Patrick Blackburn, Chiara Ghidini, Roy M. Turner and Fausto Giunchiglia, 164-77. Berlin: Springer, 2003.
2. Ackerman, Mark S., and Christine Halverson. "Considering an Organization's Memory." In *CSCW '98: Proceedings: ACM 1998 Conference on Computer Supported Cooperative Work, Seattle, Washington, November 14-18*, 39-48. New York, NY: Association for Computing Machinery, 1998.

Fundamental Assertions

1. Preservation = ensuring the **conveyance of meaning** over time
2. Digital preservation = ensuring that important characteristics and values of digital objects can be **consistently reproduced** over time within an acceptable range of variability
3. Main criterion for success of preservation = meaningful use of the information through reproduction of **state information** that is acceptably similar to the object as it was originally captured and retained
4. Digital resources are composed of interacting components that can be considered and accessed at different **levels/layers of representation**
5. Each level conveys information and can convey **meaning** (emergent properties directly based upon, but not fully reducible to level below) **not available through any of the other levels** - moving between levels involves translation that adds and removes information

Digital Resources - Levels of Representation

Level	Label	Explanation
7	Aggregation of objects	Set of objects that form an aggregation that is meaningful encountered as an entity
6	Object or package	Object composed of multiple files, each of which could also be encountered as individual files
5	In-application rendering	As rendered and encountered within a specific application
4	File through filesystem	Files encountered as discrete set of items with associate paths and file names
3	File as “raw” bitstream	Bitstream encountered as a continuous series of binary values
2	Sub-file data structure	Discrete “chunk” of data that is part of a larger file
1	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)
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	Mounting a hard drive and then generating a sector-by-sector image of the disk using Unix dd command
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

Bitstream on physical media

Examples

ContextMiner Alpha 3.0

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

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

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

Object or package

In-application render

File through filesystem

File as “raw” bitstream

Sub-file data structure

Bitstream through I/O equipment

Bitstream on physical

ContextMiner Alpha 3.0

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

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

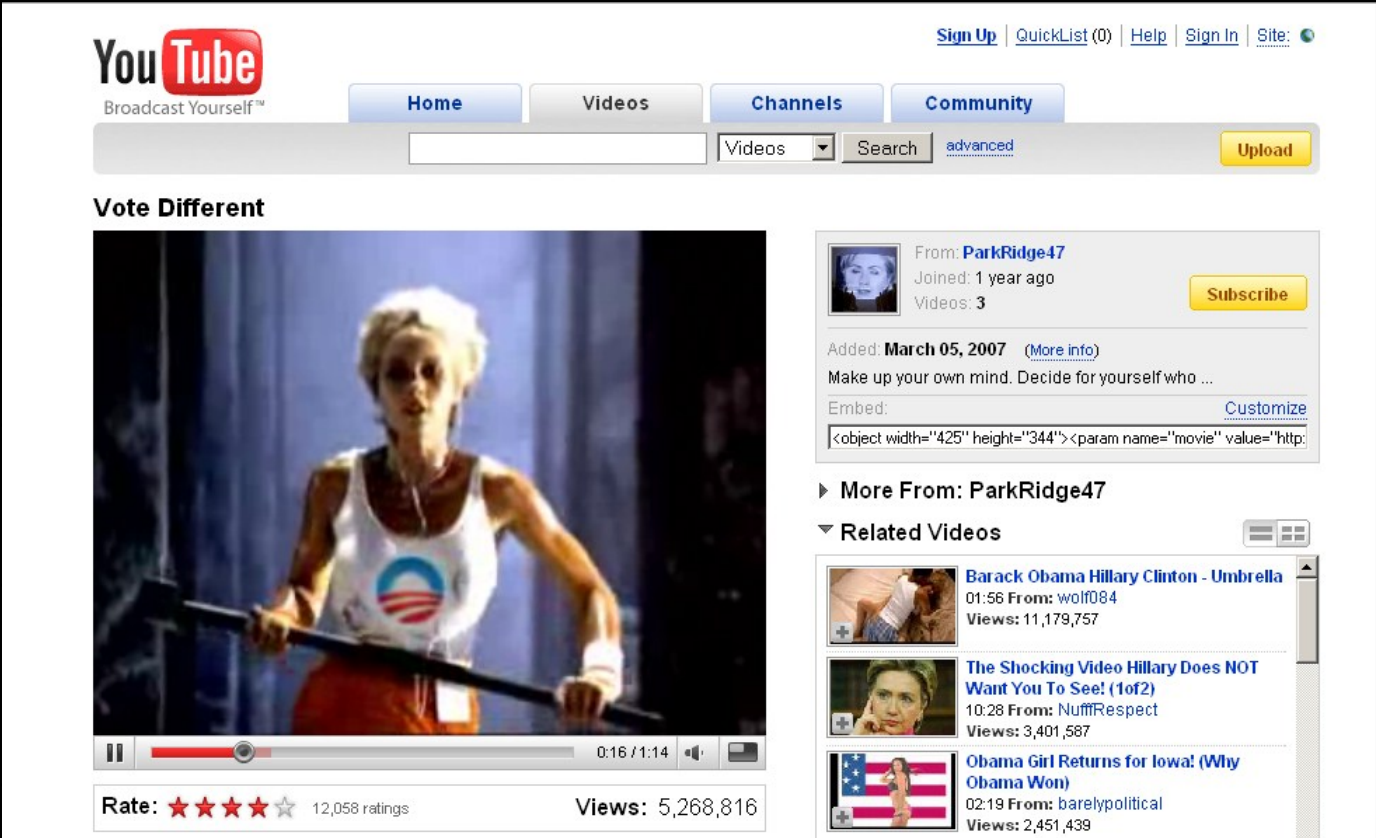
Crawling since 2007-07-19

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	32404	335	3.57	311	5	54	0	--

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
File through filesystem	
File as “raw” bitstream	
Sub-file data structure	
Bitstream through I/O equipment	
Bitstream on physical medium	

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

Bitstream on physical medium

```
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:\>
```

Viewing contents of a folder using Windows Explorer

typing
cont

Oper

Extra
value

Mou
secto

Using
pictu
of a

Name



.Spotlight-V100



.Trashes



._.Trashes



._nc-busmodels-jpw2009.pptx



EMSS Meeting.ppt



nc-busmodels-jpw2009.pptx

Interaction Examples

Level

Aggregation of objects

Object or package

In-application rendering

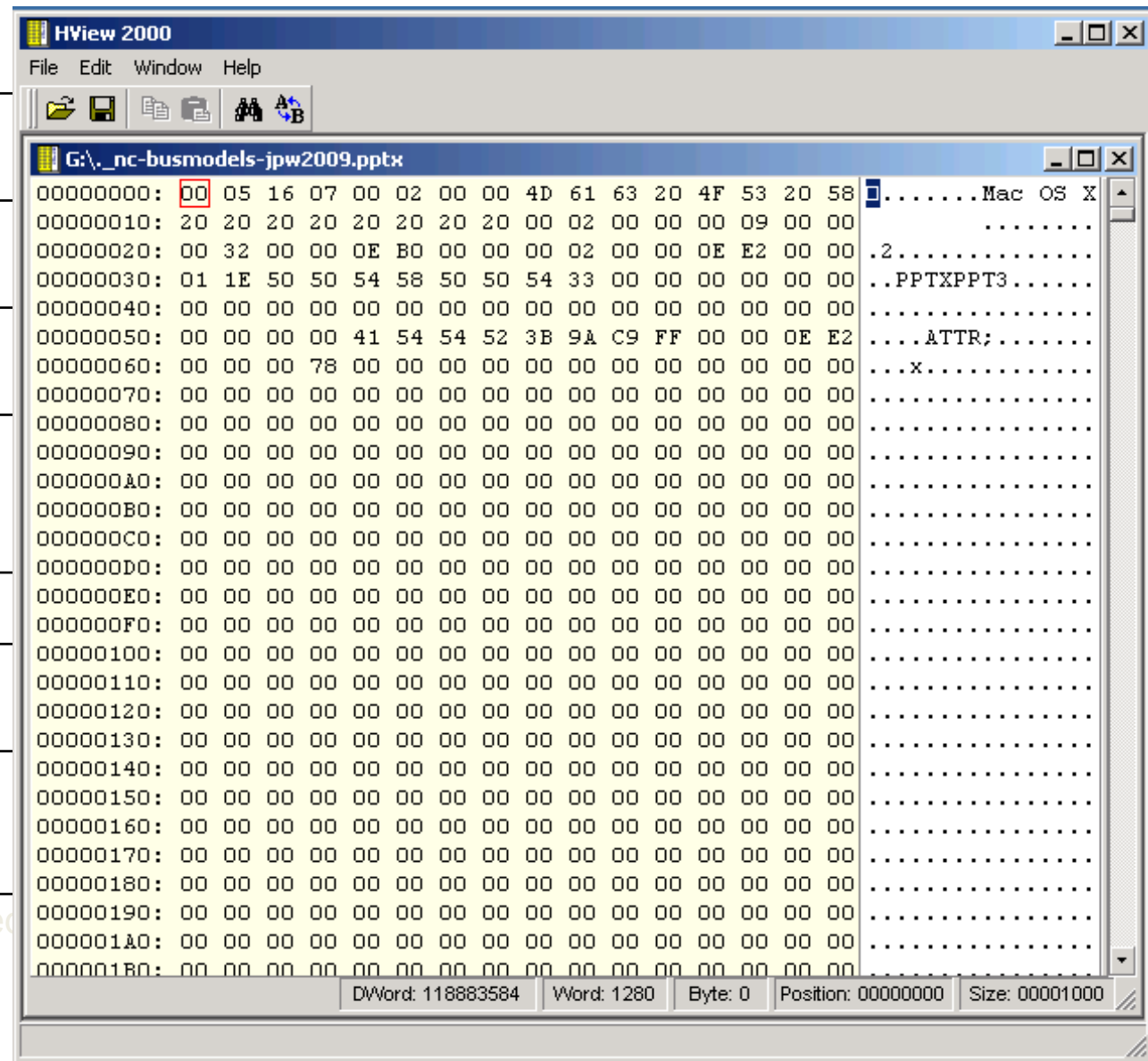
File through filesystem

File as “raw” bitstream

Sub-file data structure

Bitstream through I/O
equipment

Bitstream on physical media



Level

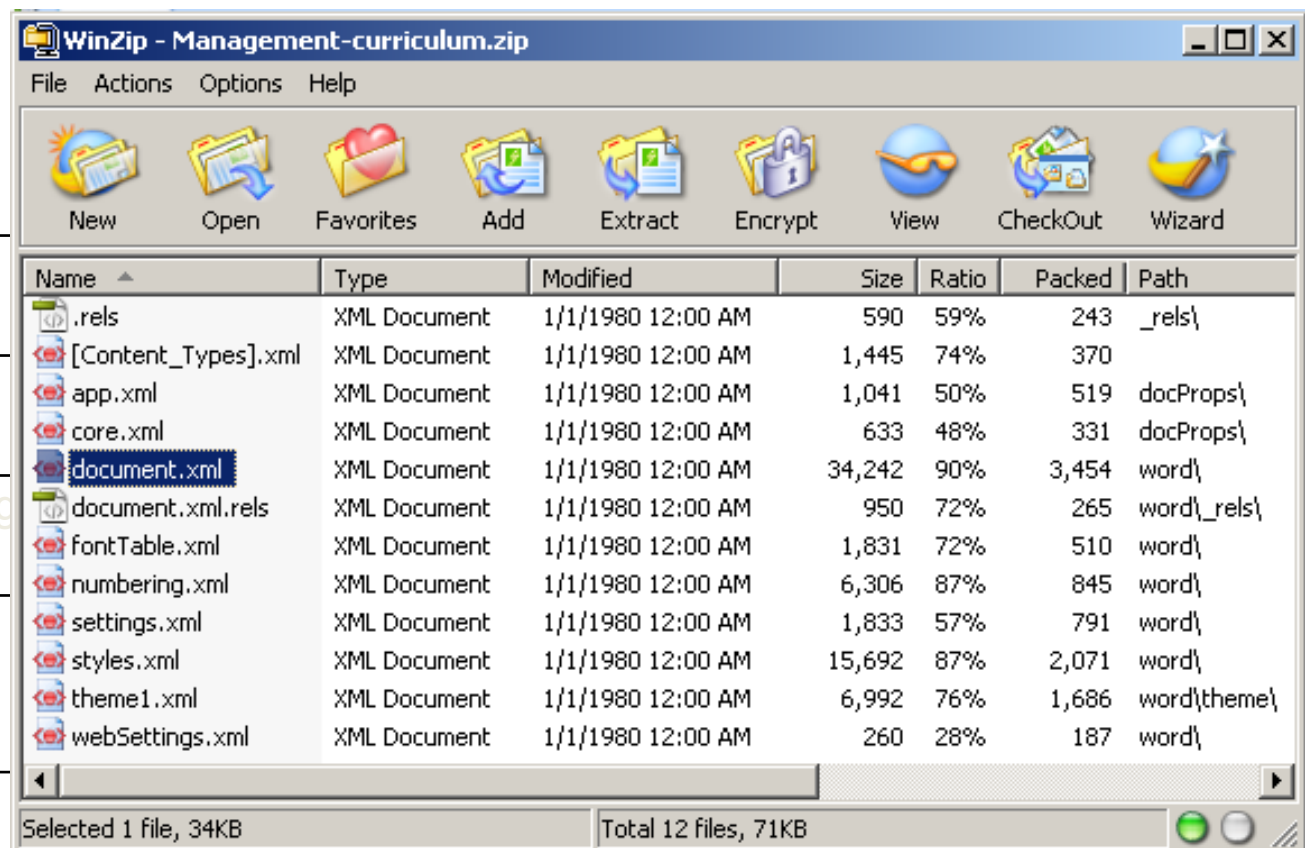
Aggregation of objects

Object or package

In-application rendering

File through filesystem

File as “raw” bitstream



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

Mounting a hard drive and then generating a sector-by-sector image of the disk using Unix dd command

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

Level	Segment =====	arg =====	length =====	data =====
Aggregation	afflib_version	0	7	"3.3.3"
	aff_file_type	0	3	AFF
	acquisition_commandline	0	36	aimage /dev/sda /mnt/charlie-002
Object or package	acquisition_device	0	8	/dev/sda
	sectorsize	1024	0	
	pagesize	16777216	0	
In-application	devicesectors	2	8	= 9999864 (64-bit value)
	acquisition_macaddr	0	18	00:0b:db:4f:6b:10.
	acquisition_dmesg	0	27298	[0.000000]
File through I/O				Initializing cgro
	image_gid	0	16	7256 F895 DE4F E304 233E 21C0 2347 CCC5
	acquisition_date	0	20	2009-11-12 19:12:18.
File as "raw"	md5	0	16	0609 2DFE AA4F B183 946F 95D8 AD84 519E
	acquisition_seconds	1570	0	= 00:26:10 (hh:mm:ss)
Sub-file data	imagesize	2	8	= 10239860736 (64-bit value)
Bitstream through I/O equipment		Mounting a hard drive and then generating a sector-by-sector image of the disk using Unix dd command		
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

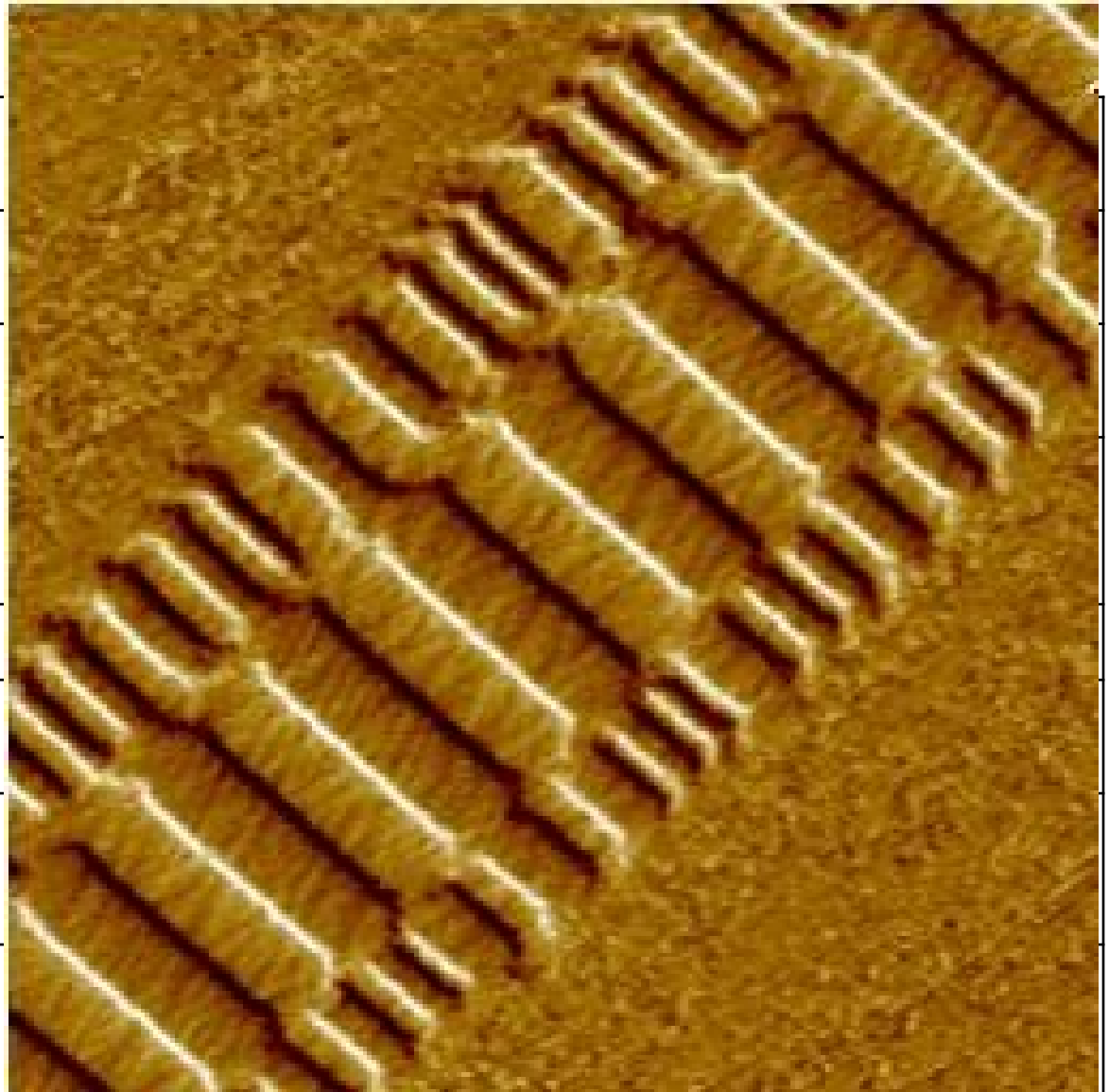
File through filesystem

File as “raw” bitstream

Sub-file data structure

Bitstream through I/O
equipment

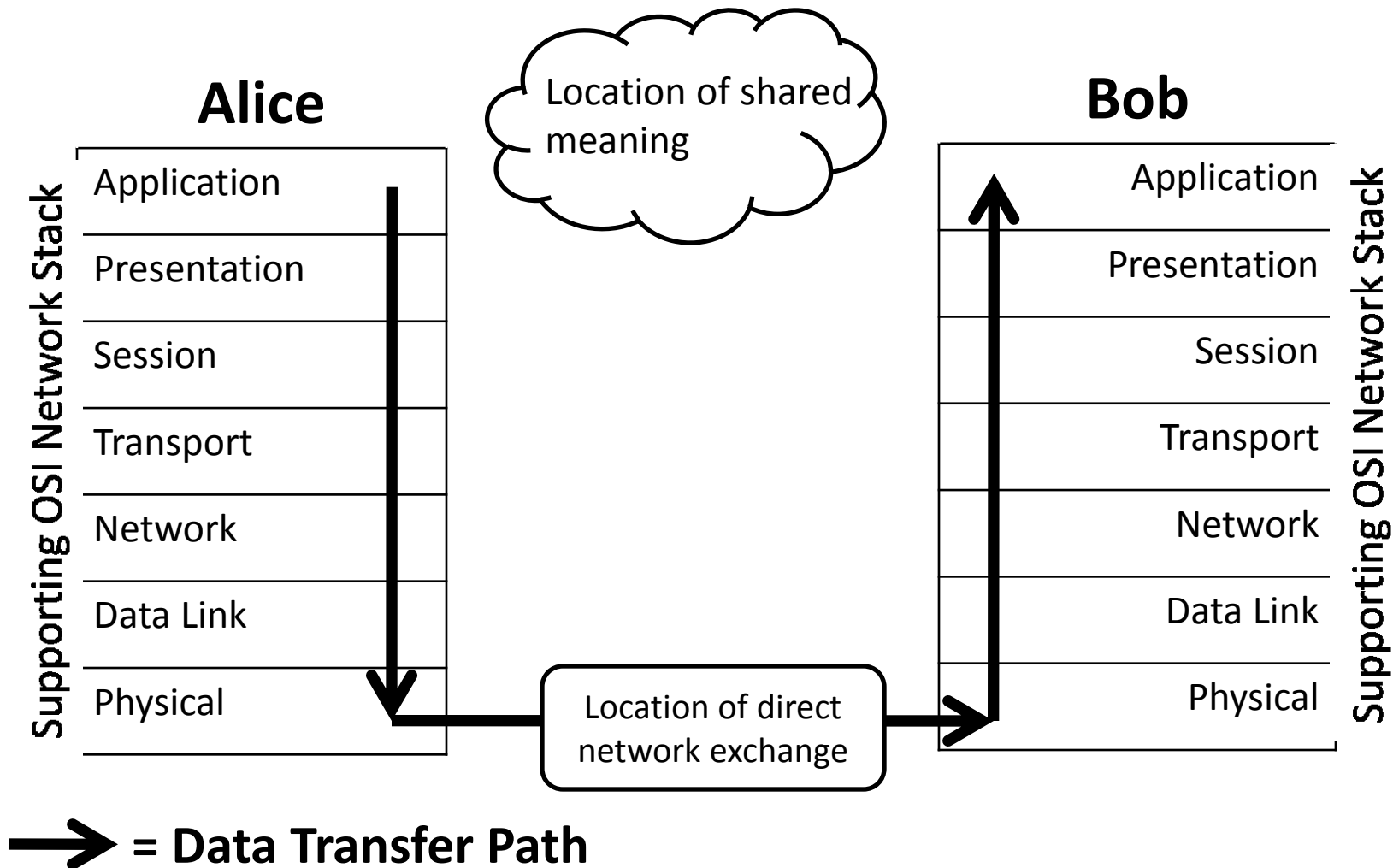
**Bitstream on physical
medium**



A layered model that you may have
encountered before:

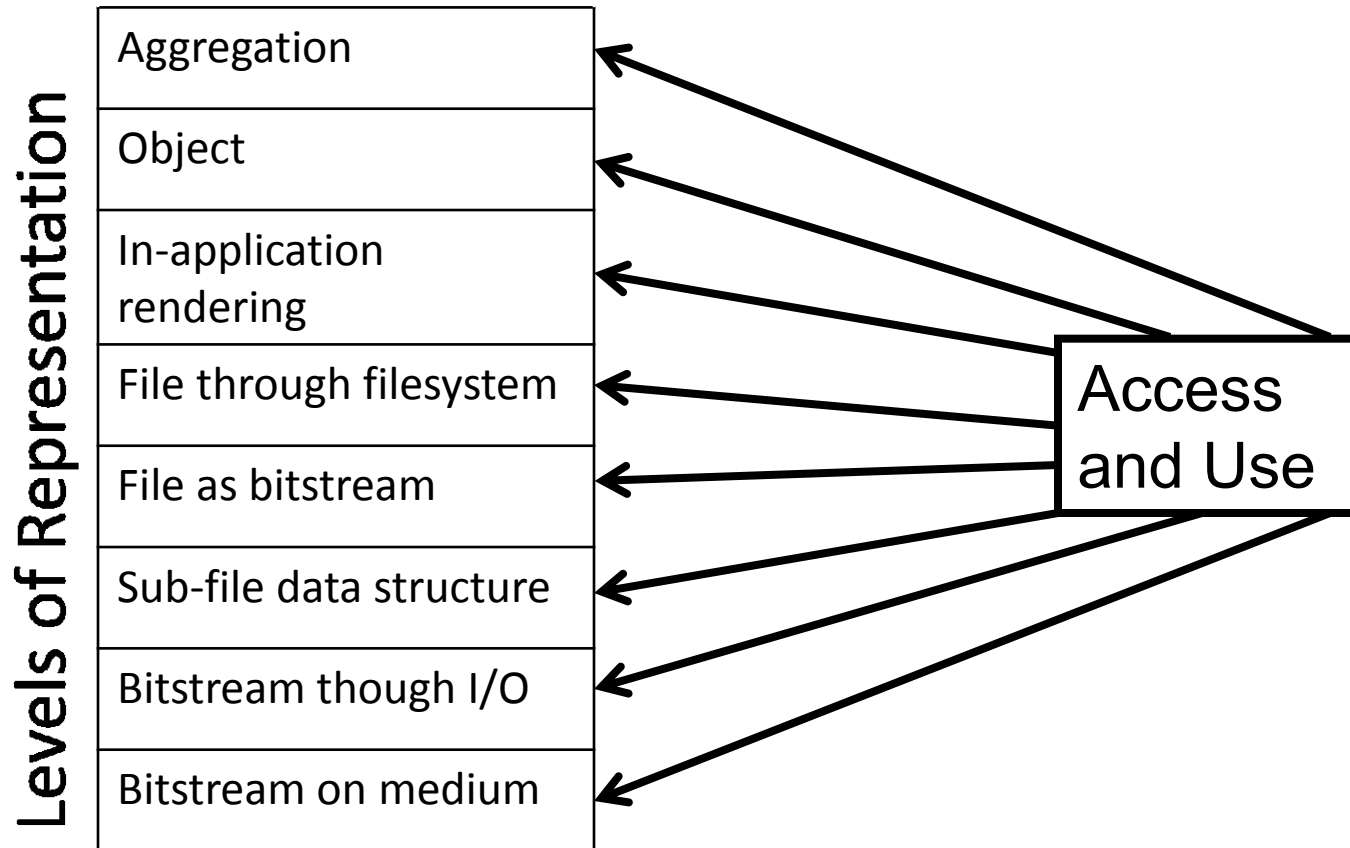
OSI stack for network
communication

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



Source for this and next slide: Lee, Christopher A. "Digital Curation as Communication Mediation." In *Handbook of Technical Communication*, edited by Alexander Mehler, Laurent Romary and Dafydd Gibbon. Berlin: de Gruyter, forthcoming.

Alice's Digital Objects



→ = Potential paths of interaction

Note contrast to previous figure: not just conveying meaning across the top layer

Implications for Archival Research and Practice

- Need to further elaborate many fundamental archival concepts – an important form of elaboration is through **implementation** in working systems
- Archival description and metadata conventions – archival description standards already support description at multiple levels but don't currently address the digital representation levels I've been discussing
- Digital preservation strategies – both what is studied and measures of success
- Professional education – value of ability to navigate across levels and recognize the relationships between them
- Oh, and it's **ethics** all the way down...

Ethical Issues – Stakeholders and Vectors of Interest

Relation of Stakeholder to Information*

Is **about** him/her

Is **of** his/her actions

Is **for** him/her

Is **to** him/her

Is **from** him/her

Is created **by** him/her

Vector of Interest*

Control

Access

Destroy

Level of Representation

Aggregation

Object

In-application rendering

File through filesystem

File as bitstream

Sub-file data structure

Bitstream through I/O

Bitstream on medium

*Not mutually exclusive