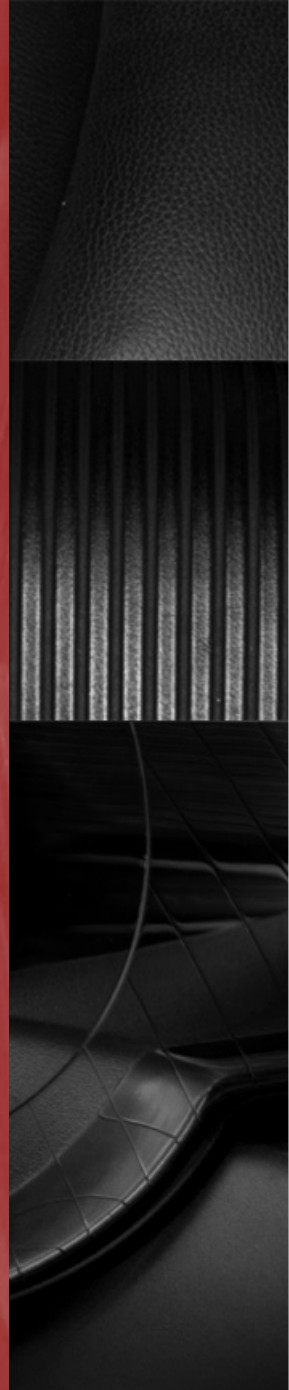


Synthesizing Bits into Objects: Data Management for Information Preservation

Eliot Scott

The University of Texas at Austin



What constitutes Data Management?

- Prior Forms of Data Storage
 - Stone
 - Clay
 - Papyrus
 - Paper
- Object Preservation Challenges
 - Fire
 - Water
 - Molds
 - Pollutants
- Catalog Objects to Facilitate Access



<http://www.telegraph.co.uk/culture/books/7772052/The-Vatican-Archive-the-Popes-private-library.html>

Digital Data Management

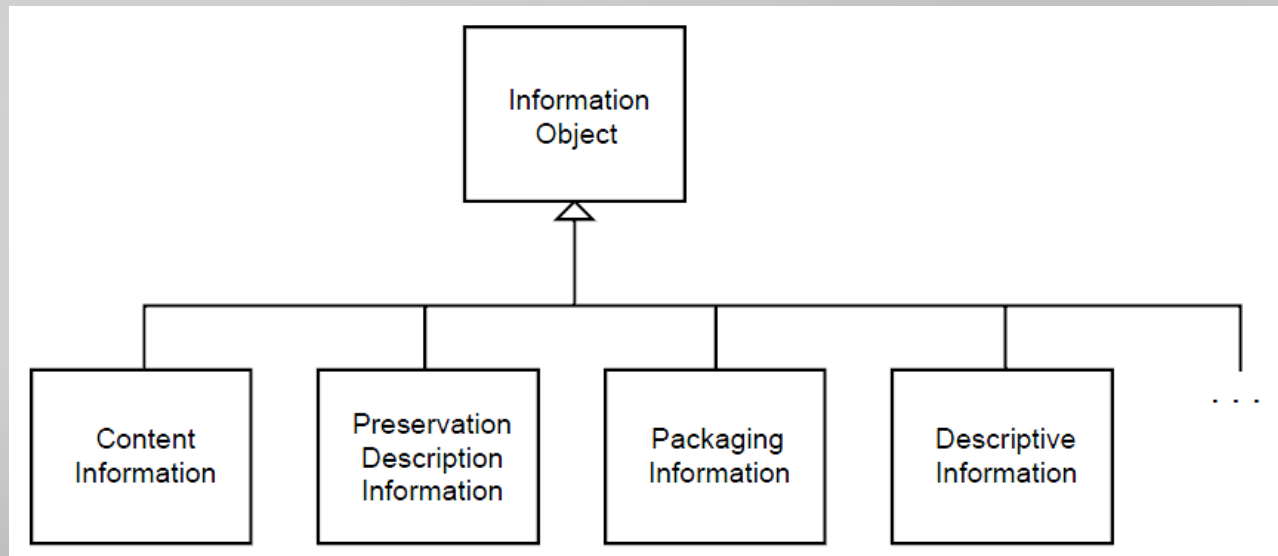


<http://1userverrack.net/2011/05/03/server-room-4/>

- Digital Data Storage More Complex
- Preservation of Object
 - Requires moving the bits to other physical mediums
 - Requires Environment to read
 - Hardware and Software Dependencies
- Cataloging Objects for Access
 - Inherently contain some cataloging data
 - Clones in numerous locations

How Bits become Objects

- Digital Objects
 - Physical – Bits written on media (HDD, CD, DVD, etc.)
 - Logical – Recognized by application software (MIME types, encodings, compressions)
 - Conceptual - Unit of human recognizable information (book, map, photo, etc.)



Consultative Committee for Space Data Systems (CCSDS) Reference Model for an Open Archival Information System (OAIS) (2002).

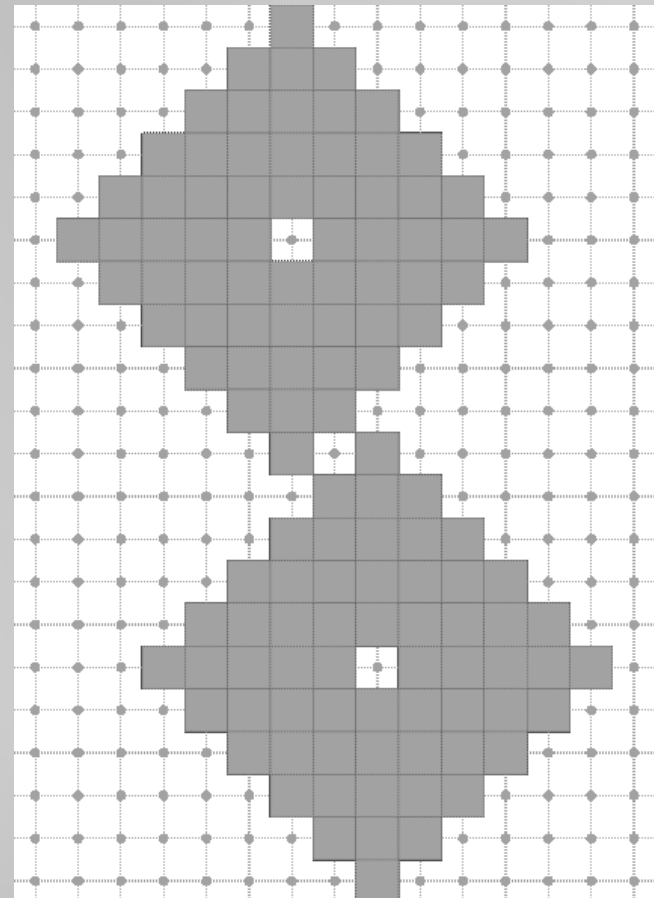
What Data formats are we Managing?

- Digital Objects

- Data Sets (db's, xls, etc)
- Documents (rtf, pdf, docx, etc)
- Images (jpg, tif, png, etc)
- Videos (mpg, avi, etc)
- Web Pages (HTML, CSS, Images)
- eBooks (epub, azw, lit, etc)

- Digital Technologies

- Source Code
- Applications
- Operating Systems and Environments



http://liris.cnrs.fr/dgtal/doc/nightly/dgtal_dgtalboard.html

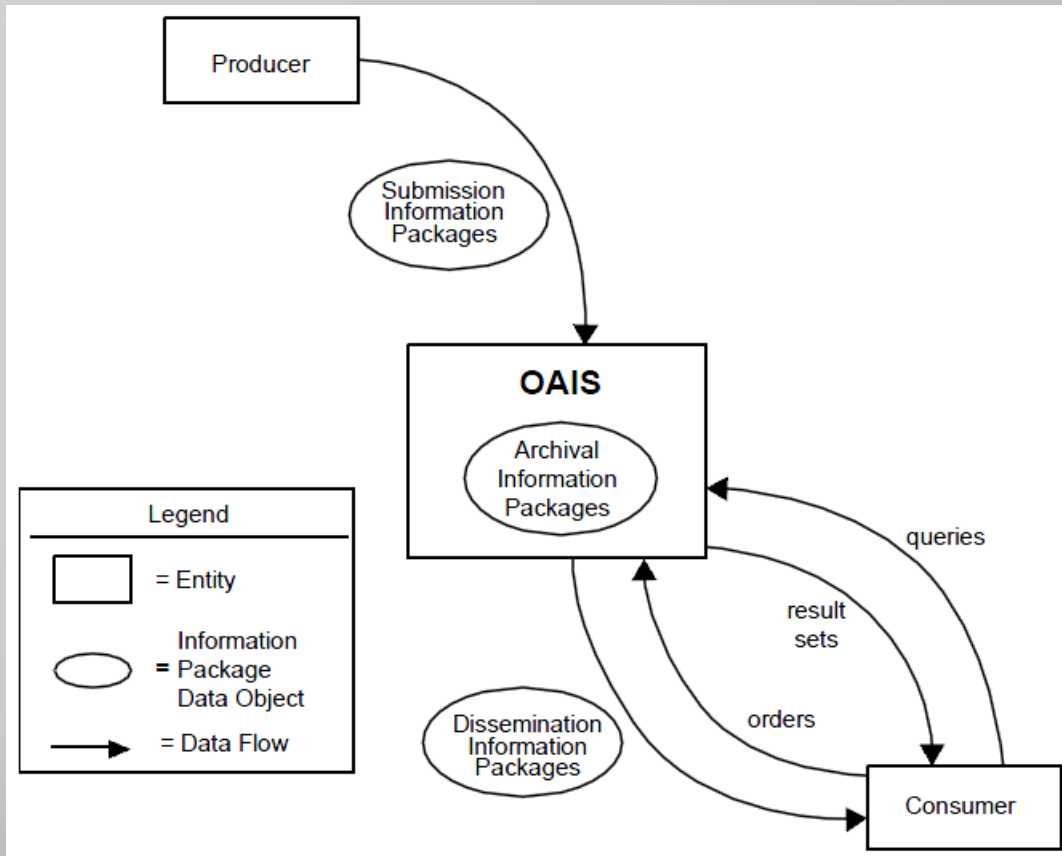
Where does Data Management Begin?

- Producers ▪ Creators ▪ Owners
- Authenticity vs. Reliability



<http://www.gizmag.com/identifying-authors-of-anonymous-emails/18091/>

Digital Data Management Solutions

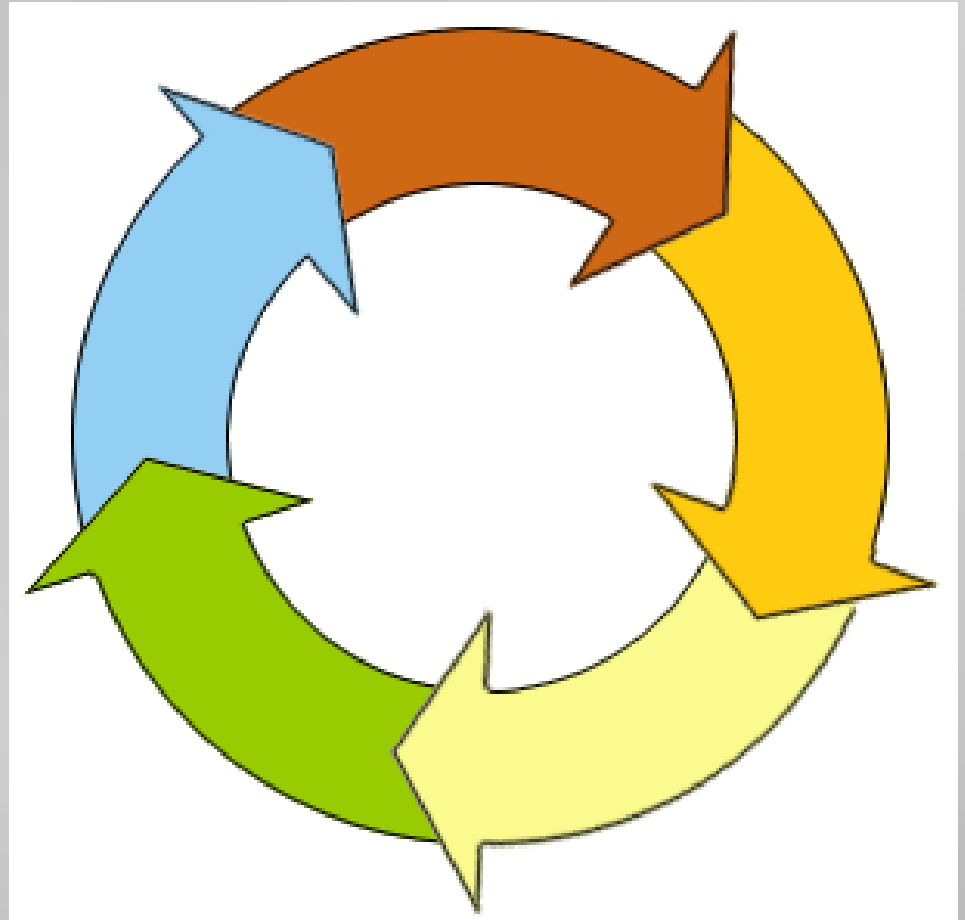


- Open Source Libraries & Content Management Systems (Access focused)
 - XTF, DLXS, Greenstone, Drupal
- Open Source Repositories (Preservation focused)
 - D-Space, Fedora, LOCKSS, Archivematica, Islandora
- Commercial
 - ContentDM, Veridian

Consultative Committee for Space Data Systems (CCSDS) (2002)
Reference Model for an Open Archival Information System (OAIS)

Information Lifecycle

- Creation
- Versions
- Use
- Appraisal
- Transfer
- Authenticate
- Describe
- Rights
- Preservation
- Use



Creation and Use



<http://www.dashpunk.com/blog/curation-v-creation.html>

- Malleability of digital
- Not fixed on media
- Constantly changing
- Easy to overwrite
- Context can be lost
- Even opening older files can destroy provenance

Appraisal

- Keep?
 - Associated Costs
- Keep but do nothing
 - Let it Die
 - Need to save environment
- Repurpose
 - Separate Form and Content
- Destroy?
 - Smelt HDD?



<http://blumsteinatcorcoran.wordpress.com/2010/05/14/do-not-stand-idle-against-an-unfair-appraisal/>

Accession or Transfer

- Receive objects from prior context
- Metadata to reconstruct context
- Integrate into new function as fixed object
- Original order vs. Order as Received
- How to Transfer files without losing valuable data and authenticity?
- Problems with Copy
 - Compression and file segmentation changes, often along with modification dates
- Clone
 - Object as a string of sequenced bits ▪ Calculate with checksums ▪ Use Write blockers to prevent Disaster



<http://thinkis.co.uk/16/Content-Management>

Authenticate and Validate Object

- Ingestion gives object fixity
- Can guarantee authenticity if done properly
- Guarantee of prior art , patents, etc
- Assist in Intellectual property
- Protects integrity of object
- Can restrict access (ie limited use)
- Can monitor constantly for use or theft



<http://www.mitfile.com/pdf/digital-identity-authentication-in-e-commerce.html>

Cataloging and Description

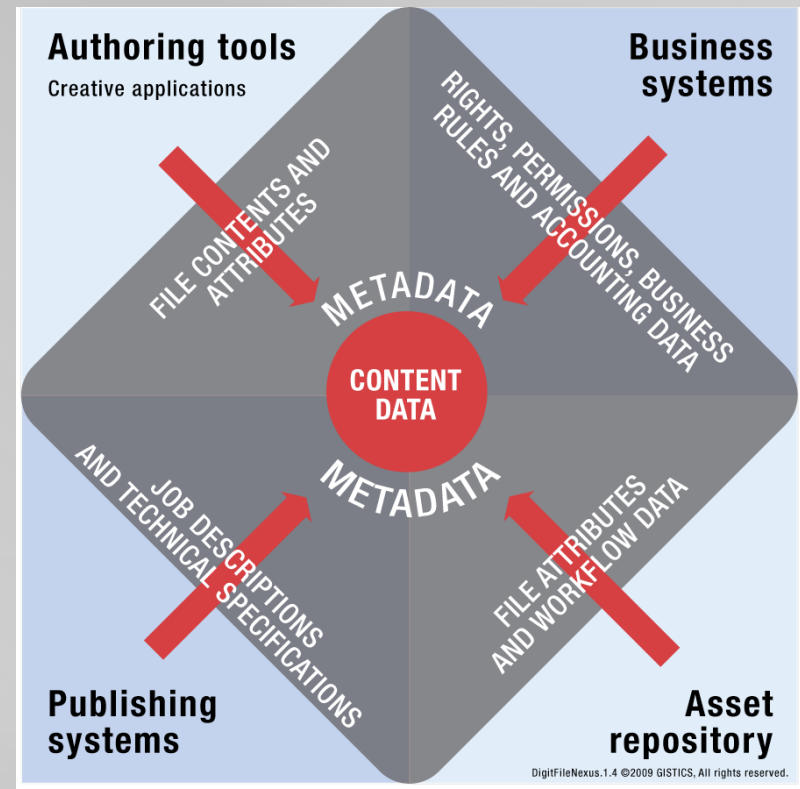
- Catalog upon accession vs.
- Continual cataloging from point of creation
- Rights and Access restrictions
- Finding Aids
- Catalog hyperlinks?
- Or archive associated content?



<http://www.granitemedia.org/2011/09/cataloging-book-ii-the-fellowship-of-the-titles-or-one-title-to-rule-them-all/>

Description of Digital Objects

- Metadata
 - Descriptive
MODS, DC, EAD
 - Administrative
Management, Access, Rights
 - Technical
MIX Images, VideoMD, AudioMD, textMD
 - Structural
METS, File Data, MIME Types
 - Preservation
PREMIS
 - Usage -generated by repository



<http://damformarketing.com/making-the-case-for-digital-asset-management-in-marketing/case-for-dam-making-digital-assets/>

XǒKǒòúûn?DU

E' u Cj9N'e{#kCvÄWai_ü _ðrai+cæÆN_UZBÜ_f...%mDž[S' + Üë?#áÆ@ú[Ü.z09Ž>§N{9,æéí+P9Yv c†vuÇG D³æÄwEoR€" wà¹æ¶
 ìlèŠ~%%~8;IWòZè""V0)¿"ùBA„,">"pÁÆýuQwHøoi-áñ ýÿ PK?? ? ? !?·ó- N?
 [?]_rels/.rels C[]([?
 Ç'ÚJA
 †[æAè]7Ü"ÒÙpHjw"è¶,"[w][w]xU¼%£ 9PÜ^æöçÈOÖ}f;Ô;<[ayÖ Ø`Gßkxm< PYÈ[§,g

igi@µC×6_Ö]7~fiØË%œÈ aoÜ.b*llErðö,el0i%ñ-6aiçÖðDÿ_çŽ..., i %!ó|uceZ^ZtÜcyç~;?lY,?){ûCf~/h>? ýÿ? PK???? ? ? !Öd³Qú 1? ? ? ?word/_rels/document.xml.rels ç? ? (?
-‘ÍjÄ0Ö,ï,...%fØ{-;ýi„È”@*~û Š%þi²?>¶~üCR†?÷âç`Fhæ””>iwoÄ?êœU%?)¥«:Ü(x/vwİ ^µ-‘q??
H°!oo6~h4ÇCÔvžDL±n eök)©!±x”8ñÎ.ôš£
ø°üð

U8q|hæœ°pázÊ8 Öf°äUÿ Srð#æ¶§_Êj+6' %ø·y·ðæ+€8sè6¶§ÇA T
Zq|Ä>e"ð?ð°Wðð-|V/c·7J5äÜð§äó|'päp/ ýÿð PKððððð !-µ-â-ð Pð ð word/theme/theme1.xmlIYo0Ü6ð¿ðøw toc'vð uSð°-M-ðÄn+ð¶%--
ðPc+ð|ð)ðüäðÄðä°+ðððm+â[ðø¿ü4Ü:|ðð°GR'ÄX'â6ð§>§ðüäüÿðð©«x|ç
ð¶|)OÜ°ÿr|ç§ñy+°°ÿðð-ÿH°æð°ñ, '½|'pµß§>§XUdb, 'j°xqÜ©°Jx---ðÄX'æ|j'n|ñECE%§§p)ðððèEli|V[]§1M-c°àèPðððOðP°ð6äð=ð°%zÄgð |ðg...Äð uSÜEððððöÖEOÄðä%ððÄRÄDÜ«°™ÿ·qu
-g:'CZ°¶°°oÿUðAP|XðpT0-ð+|ð¶|ðð|ðæxððø°%ZÄ|°fif|Vp-2|fð-ðÉi-|ðøqzv-0-5|°%ðEæel-N§Öle°Xcðððððøµu|çsÜÄðÄ7çðð°»èà

dXÑ?©iJÆØ‡(îâx\$(Ö

~«JÊtøe\O*ôthG%ëHYMæ-}KNBÁP+*Y%É|±αŠÛTÑ¼#/#øA7ÂqZ...ðÐ\$*c?çCçiqUßân+èwðNñøúñŽ»O~·iè'4
 ð=3Ú—PâQöäíÊ1FÞ
 \\9†øêÇ'œŕĈÂMØ“a2auDù,,;Yt»ðó!~[x'üòóúē]Wyr%j|Él'ig-'³Uew÷

[illegible][illegible]

Default Paragraph Font Riöý³R

00
Table Normal öô 4Ö 4Ö äö
(k ôÿÁ (

No List

PK [Content_Types].xml-‘ĕjÃÖ÷...pfđŧøº(¥ðİciw),Öä±-jß4 Éßwi,Pº-t#bí™{U“ótéU^h...d)ã¨«òúî»×*1Pƒf~ô
“^xxWá0)”™!Téz9<“Ü Šyi)A~@tæ¶(ïÆÄHœñǺĐuù*

DxzfÊ/OŠÇ° ðûü \$€
X«Č3azCòÀà,°D0 j~è3B¶ĩbăĩ¬>đđí3ê\`ð?ě/c[đđ-¶Géá□ǼÍ-Y-ÚRK.“spÓ»—·’aaē-? ýy PK !¥ÖŞçÀ _rels/.rels,Ǻ0 ěī...½fN)QÒǺ%v/¥Ø/E} á[āh”ŮĖĪŬOÇ»„ɦ÷©=p•‘ùâ”č Ÿš²ǺĀc?ĔháV=¿,É...ş§%xp+£{Ůµ_ %PN£<1¶YH¶0•ŲŮÔOR*BdÑÉŇŬE4bşŞq_xq~žğ6LÓōŦRx7*ĚÝĂ0lżOÁŢ”,ăEñn7”Liabj~/ăS¼eªŌĐμ,űőý ýy PK !ky-ťř Š theme/theme/themeManager.xml İMǺ @á}ıwđ7c»(Eb²Ĕ»ö CœAC ÖŶŮ×ääřf7Í8ĐÖ.K

Y,œ

Šeí' á:đ ħm'í'ÍHĤA lácĥæývÉ'ŋRlÈsQ)šǫĥ-íY Ůll+Ǫlí YÁ'si=GWèÓ÷láEë+8r8yǎ wǎw PKĥĥĥĥĥĥ | -u-â-ĥ Pĥĥ theme/theme/theme1 xmlîYQoŋî6qāēōw tōc'yĥ ušǫt)-MĥÄntþ%-

Accessing the Digital Object

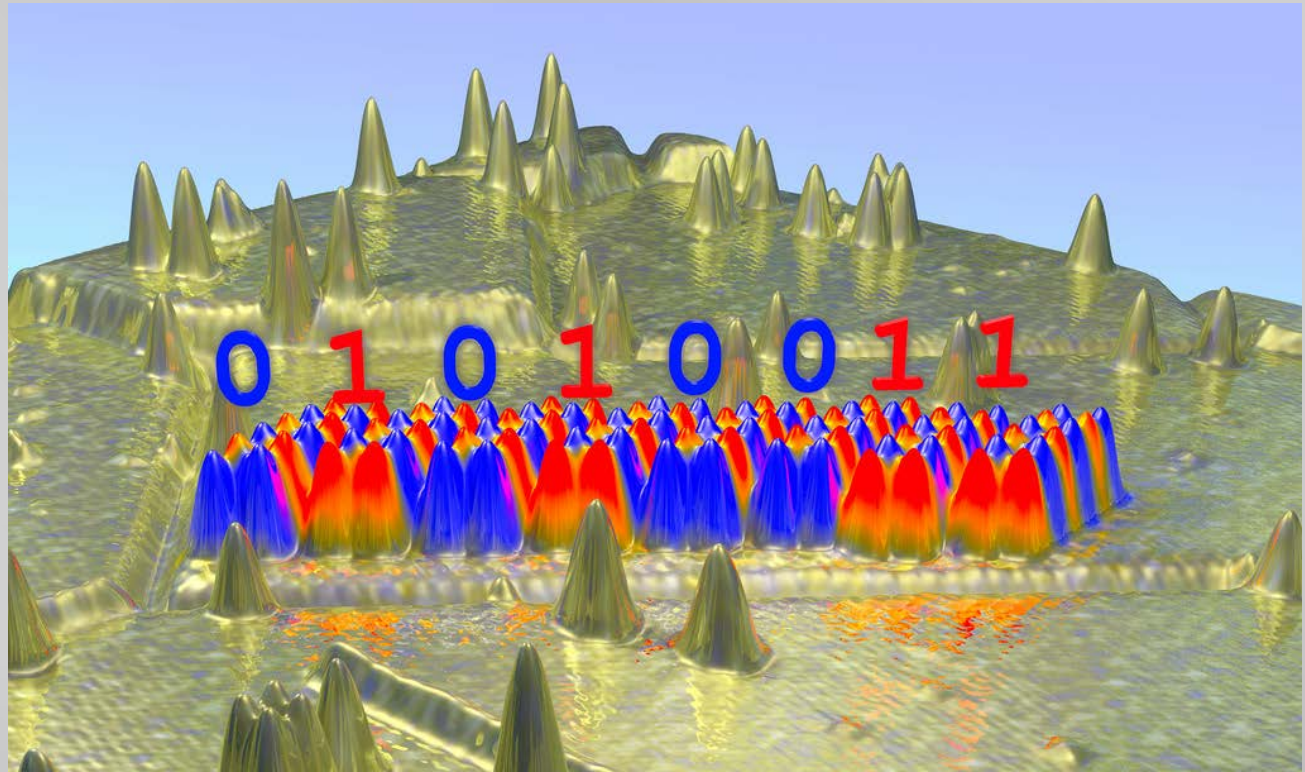
- Maintain Original Bitstream for Preservation
- Reconstitute Bitstream as logical object for application software so that conceptual object can be read by human beings
- Preserve the ability to reproduce the digital object in order to access it



http://ugis.ls.berkeley.edu/isf/resources/image_LS/5994aab.jpg

Storage of Digital Objects

- Media Type?
- Repository Type?
- Encrypt?
- Access Levels?
- Media
- Databases
- Bitstreams
- Files



<http://inhabitat.com/ibm-creates-the-worlds-smallest-storage-device-and-its-12-atoms-in-size/>

Preservation Considerations

- Feasibility
 - Hardware and software
- Sustainability
 - Will it be feasible in the future?
- Practicality
 - Within technical and economic limits for institution
- Appropriateness
 - Is method reasonable for object in question



<http://easydigitalpreservation.wordpress.com/2010/06/09/video/>

Preservation of Digital Objects



- Preservation Methods
 - Emulation
 - Migration
 - Refreshing

<http://blogs.loc.gov/digitalpreservation/2011/11/the-artifactual-elements-of-born-digital-records-part-1/>

Emulation



- Began when writing Mainframe code for not yet released hardware.
- Software emulation based on wiring diagrams
- Preserve Hardware Technology
- Preserve Software Environment on new hardware

http://news.cnet.com/8301-13860_3-20020759-56.html

Migration

- Transforming the object into a new state
- Porting to a new application
(ie Word Perfect to Word)
- Information dropped during conversion
- Keep original as new techniques evolve
- Sequential migration
 - Make access copies from original as new applications come out
- Loss of features from original
- Content vs. Design Preservation



<http://www.culanth.org/?q=node/355>

Refresh

- Clone to put on new media
 - Floppy Disk -> Hard Disk
 - Hard Disk <-> Optical Disk
- Necessary for Preservation as Physical Media Decays
- Used in Conjunction with Emulation of Migration



Future Models?

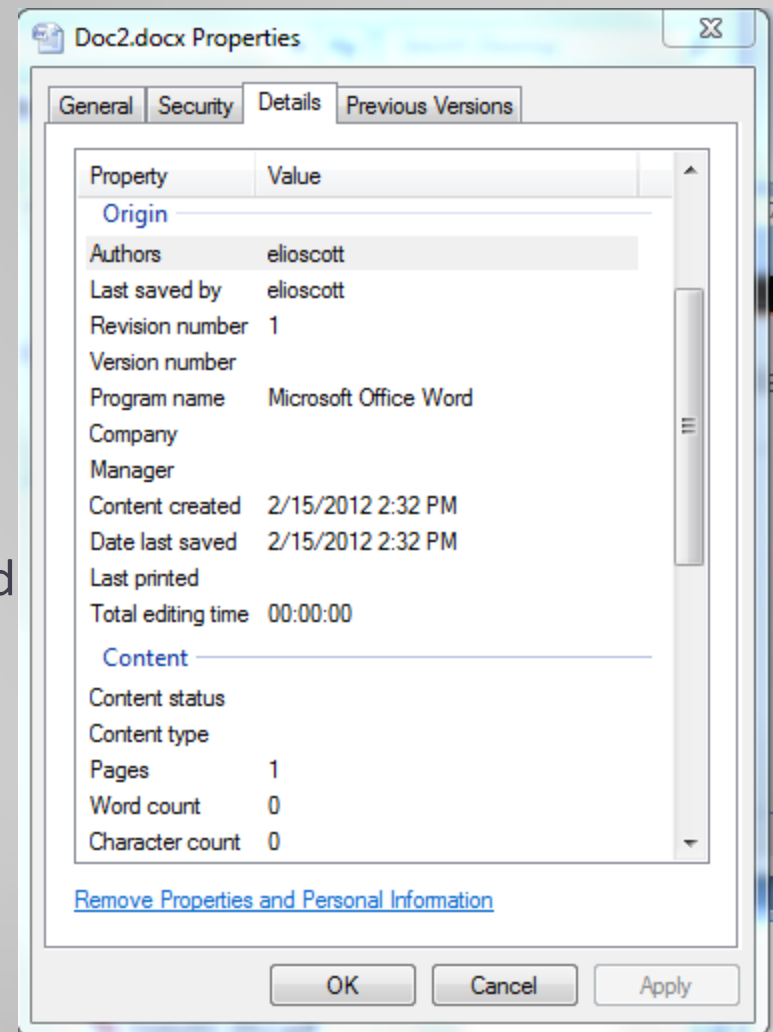
- Expand on OAIS
- Hardware upkeep expensive
- Mix Emulation and Migration Strategies
- Use by use case
- Preserve Legacy Systems and Software in Bitstream Format
- Have a Library of Virtualized Systems
- Hope virtualization software remains open source and that can emulate environments on newer chipsets (ie ARM)



http://www.ciudadesdigitales2011.com/eng/que_es.html

What can you do to ensure your data as information?

- Add author metadata to documents
 - MS Office
- Practice good versioning
- Speak with your librarian, archivist or data manager if you need to archive valuable information as it is being created
- Keep documentation of activities and processes surrounding creation
 - Chain of custody
 - Circumstances of Creation





Questions?

■ References

- Abrams, Stephen ,Sheila Morrissey and Tom Cramer (2009) What? So What?" The Next-Generation JHOVE2 Architecture for Format-Aware Characterization. The International Journal of Digital Curation Issue 3, Volume 4 .
- Caplan, Priscilla (2009) Understanding PREMIS . Library of Congress Network Development and MARC Standards Office.
- Consultative Committee for Space Data Systems (CCSDS) (2002) Reference Model for an Open Archival Information System (OAIS) (2002).
- Duranti, Luciana (2009) From Digital Diplomats to Digital Records Forensics. Archivaria 68:39-66.
- Hedstrom, Margaret and Christopher A. Lee (2002) Significant properties of digital objects: definitions, applications, implications. Proceedings of the DLM-Forum 2002 Parallel session 3: 218-223
- Henderson, Deborah (2010) DAMA-DMBOK Guide(Data Management Body of Knowledge) Framework Paper and DAMA-DMBOK Guide : Overview January 2010
- Hockx-Yu, Helen, and Gareth Knight (2008) What to Preserve?: Significant Properties of Digital Objects. The International Journal of Digital Curation Issue 1, Volume 3: 141-154.
- Kirschenbaum, Matthew G. , Richard Ovenden, and Gabriela Redwine (2010) Digital Forensics and Born-Digital Content in Cultural Heritage Collections. Council on Library and Information Resources Washington, D.C.
- Lawrence,, Gregory W., William R. Kehoe, Oya Y. Rieger, William H. Walters, and Anne R. Kenney (2000) Risk Management of Digital Information: A File Format Investigation. Council on Library and Information Resources Washington, D.C.
- Mosely, Mark (2008) DAMA-DMBOK Functional Framework Version 3.02. DAMA International.
- The State of Digital Preservation: An International Perspective Conference Proceedings(2002) Council on Library and Information Resources Washington, D.C. July 2002