

# PRESERVING THE PRESENT:

Current Best Digital Practices

Melitte Buchman  
Digital Content Manager  
NYU Libraries

Powered by NYU **DLTS**



## Scholarship remains, the discovery tools change



Painting: *School of Athens*, Raphael, 1510-11, Plato and Aristotle central figures.

It is clear that scholarship methodology has changed over the ages. It should come as little surprise that in this age of technological advances that the new generation is looking toward on-line digital tools to support and expand their scholarship. At NYU it has become apparent that there is an increasing role for digital migration of analog material (books, archival material, and multimedia) to accessible digital formats. As education becomes less about location digital libraries become a way of sharing a knowledge base with the next generation of scholars.

## Introduction

- Scholarship in the digital age
- “Digital preservation” can be defined as ongoing access, however there is a significant difference between access projects and access/preservation projects
- The burden of legacy digitization

Veracity, context, and comprehensiveness matter as they always have in scholarship. The truth, that is the look and feel of the original, needs to be guarded due to the flexibility of digital files. Also of concern is the longevity of digital files. New digital surrogates must be made to withstand the test of time so that access to the material can be ongoing. Microfilm lasts 500 years if correctly made. A digital file needs to be created to best technical benchmarks and to internationally understood standards so that it too can last. Stand alone websites are unlikely to last long and will provide only temporary access. Create ongoing access a digital project needs to be carefully considered.

Institutions have to do the best that they can about saving digital files that were created before digital standards were firmly in place. In some cases these legacy files will need to be normalized to current standards so that they can be preserved.

## Preservation Projects

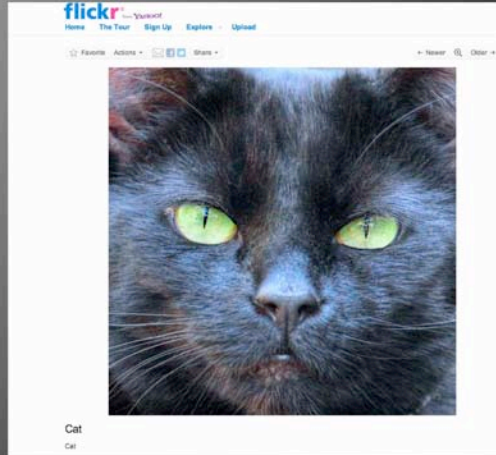
- ▣ Digital preservation of analog material includes important elements
  - Conformance with in house preservation policy
  - Most faithful representation of look and feel of source material
  - Clear discovery mechanism and permanent citation for scholarship
  - Creating files to acceptable long-term safekeeping characteristics
- ▣ Considerable common sense as well as specific expertise is need

Just as you can drive a car without understanding the mechanics of an internal combustion engine, you can also participate in guiding digital projects particularly in terms of curatorship, without knowing how to create each part of a digital library. Specialists will be needed to complete any digital project.

Selecting the content to save is entirely within the realm of the subject specialist. How to save that material needs to be guided by digital library technical services and conform to standards. To create online collections a clear digital collection policy must exist. To continue to access online collections a number of different kinds of metadata will be needed including clear discover mechanisms, descriptive metadata and technically well formed files.

## What is not a preservation project

- ▣ DAMS
- ▣ Viewers
- ▣ Flickr
- ▣ Websites
- ▣ Facebook



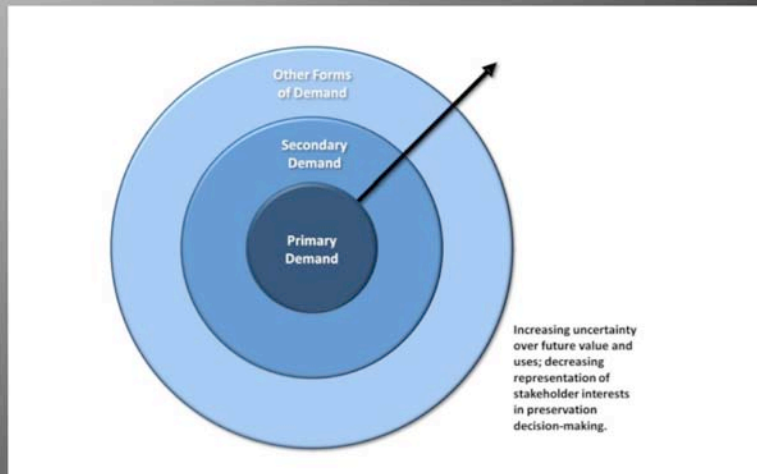
Access projects, while easily launched are unlikely to persist. Whatever access is created, either via Flickr, a finding aid, or attaching to an online catalog, there must be an underlying preservation file that is held in a repository with sufficient metadata for discovery over the long term.

## Proper Capture

- Preservation dictates that only one digitization should be done
- Good technical systems are as productive as bad technical systems
- True preservation allows for sustainable access

The following find a discussion about basic characteristics of image files that are created for preservation. Conservation, the first concern for rare physical material is paramount. Digitization should not damage the original artifact. Only one digitization of rare materials should be planned as a way of conserving the original for the future. Good technical capture takes no more time that poor low resolution capture. True preservation imaging will require good equipment and significant planning but can be extremely efficient once set up given the new generation of equipment.

## Proper Capture → Future Value



Accessed 8/3/2012, <http://brtf.sdsc.edu/>, from *Sustainable Economics for a Digital Planet: Ensuring Long-Term Access to Digital Information*, pg 19, final report of the Blue Ribbon Task Force on Sustainable Digital Preservation and Access.

It is important to remember that digital files will live beyond their original digitization need (for example for coursework) and continue to be used by scholars in different ways over time. Since those uses cannot be predicted digitization needs to be done to include future needs. High resolution images with correct color information about the original will be key to this ongoing use. Valid files that can continue to be used over time are required as well.

## In the beginning there is policy

The digital migration of analog materials should preserve, as closely as possible, the characteristics of the original and, as appropriate, the intentions of the artist/author/creator.

It is especially important to the scholar that the digital capture be as truthful to the original as possible. There should be no digital fixing of rips, clean up of stains, enhancing text or color. Since scholars will be using this material to study many different aspects of the material, for example fading of a manuscript, it is important not to use digital tools to manipulate the digital image.

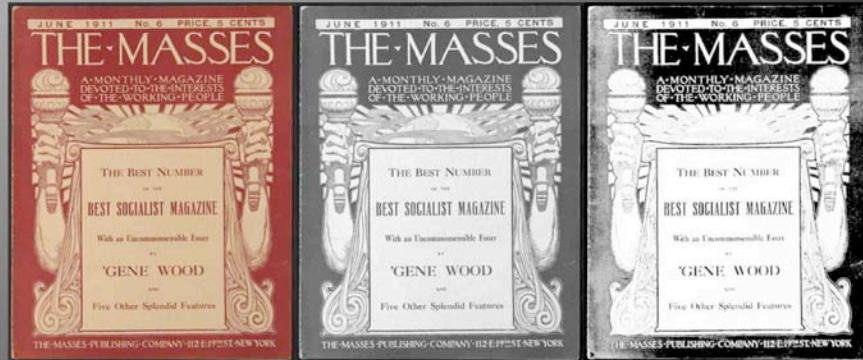


## In the beginning there is policy

The digital migration of analog materials to preserve, as closely as possible, the characteristics of the original and, as appropriate, the intentions of the artist/author/creator.

At NYU Libraries a number of our materials are artist made. Often the creator has manipulated color or other aspects of the original. The migration specialist will respect the intentions of the artist/creator.

## What is a preservation project?



Full color

Gray scale

Bi-tonal

Full color is the most desirable technical specification for researchers. Full color files maintain the look and feel of the original content and substrate (in this case paper) and will allow researchers to use these images as digital surrogates for the original materials in many cases. Grayscale, though far less useful has been used by some for preserving legibility of the content. Bi-tonal capture is nearly useless and often very difficult to read.

□ Afghanistan Digital Library



<http://afghanistandl.nyu.edu/>

This site is an example of a preservation project with an access component. Please note the amount of metadata that travels with the individual books as well as the permanent citation for scholarly publication.

## □ The Real Rosie the Riveter



<http://dlib.nyu.edu/rosie/>

This site, similarly, is a preservation project with an access “front end”.

## Current Best Practice for Preservation Digitization

- Descriptive metadata
- Technical creation of digital objects from physical original
- Other metadata
- Overview of a preservation repository, Open Archival Information System (OAIS) model

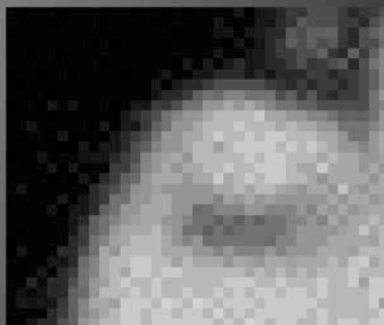
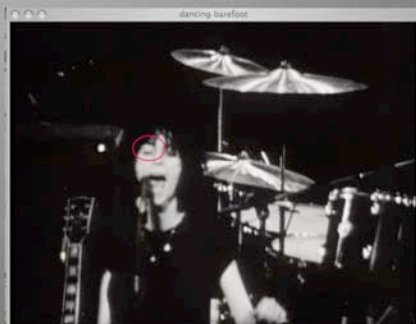
A great deal of metadata is needed to identify and understand digital projects. Descriptive metadata is the domain of the subject specialist / cataloger. In this session other classes of metadata will be discussed, though technical metadata will be explored more deeply. Experts in metadata issues are needed to build a true digital library. Some metadata will likely be needed even in access projects.

## Technical Creation of Digital Images

- ▣ Guidance on what to ask for from your in-house or vendor content provider
- ▣ Recommendation that you don't buy a scanner and start at it
- ▣ Definition of basic terms and standards

Following find a brief outline of basic information needed to have to discuss digital imaging with your digital library or vendor. As with all tools good imaging can only be done with good equipment.

## What is a pixel?



A digital picture is not a photograph. It is binary information that is interpreted by a graphical user interface (GUI) that is available on your computer. Without the GUI digital photographs would be unreadable. The pixel is the basic unit of digital imaging and it is represented by one color only.

## What is a digital image



010101010100110  
100110011010100  
101001101011010  
111011110101001  
100010110010010  
001001000010001  
101010010001000

All digital images are, at their core, binary information.



## What is format?

- Image format: a standardized method of organizing and storing image data. Can be thought of as a container.

Acceptable format for long term preservation  
TIFF6 uncompressed

Other common formats such as jpegs, pngs, jpeg2000 are not acceptable for preservation work as they are lossy. Lossy files contain compression that changes the pixels with each save. Following see an example of jpeg compression.

## What is compression?

- ▣ Compression: the objective is to reduce redundancy of the image data in order to be able to store or transmit data in an efficient form. Image compression can be lossy or lossless

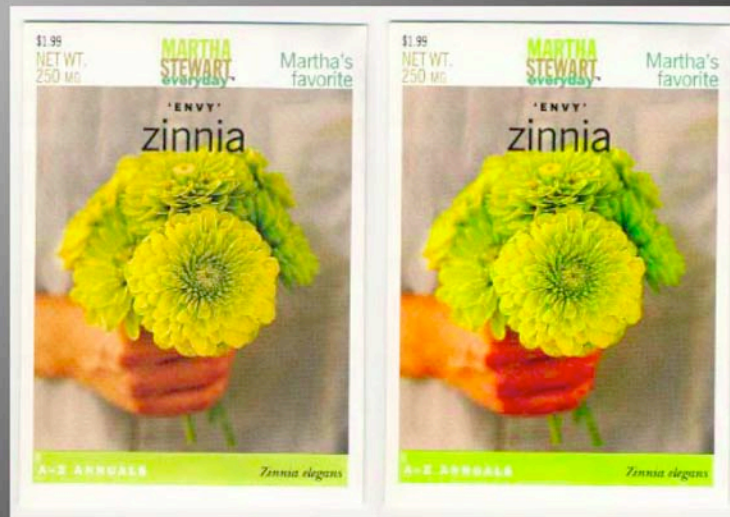
Acceptable format for long term preservation  
TIFF6, uncompressed

Lossy compression damages a file irreparably. Saving a file over and over will exacerbate this.

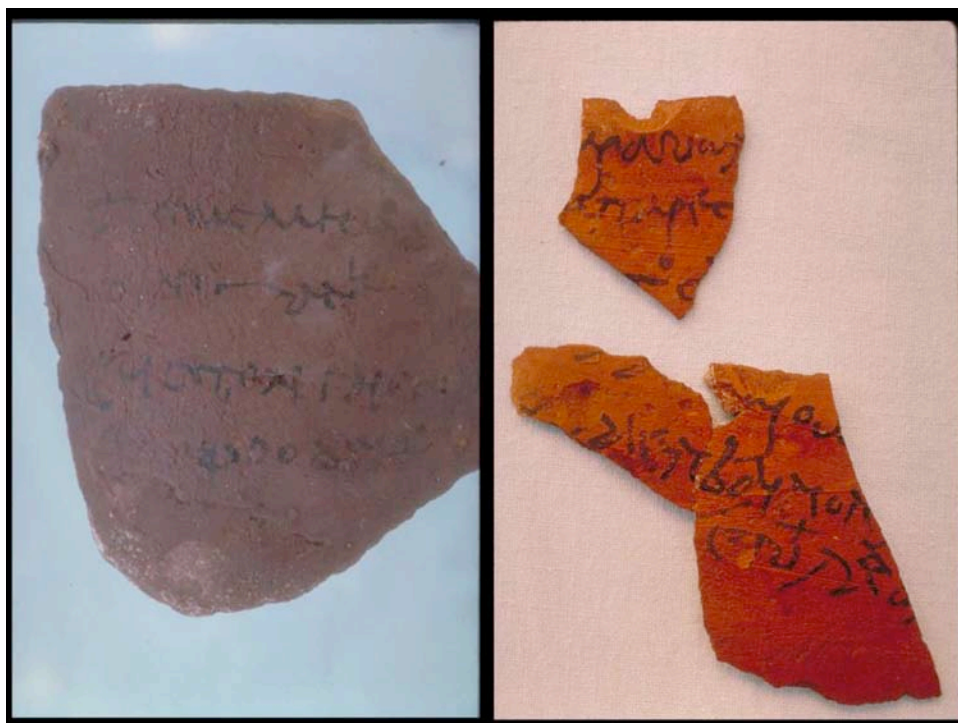


Example of jpeg compression errors.

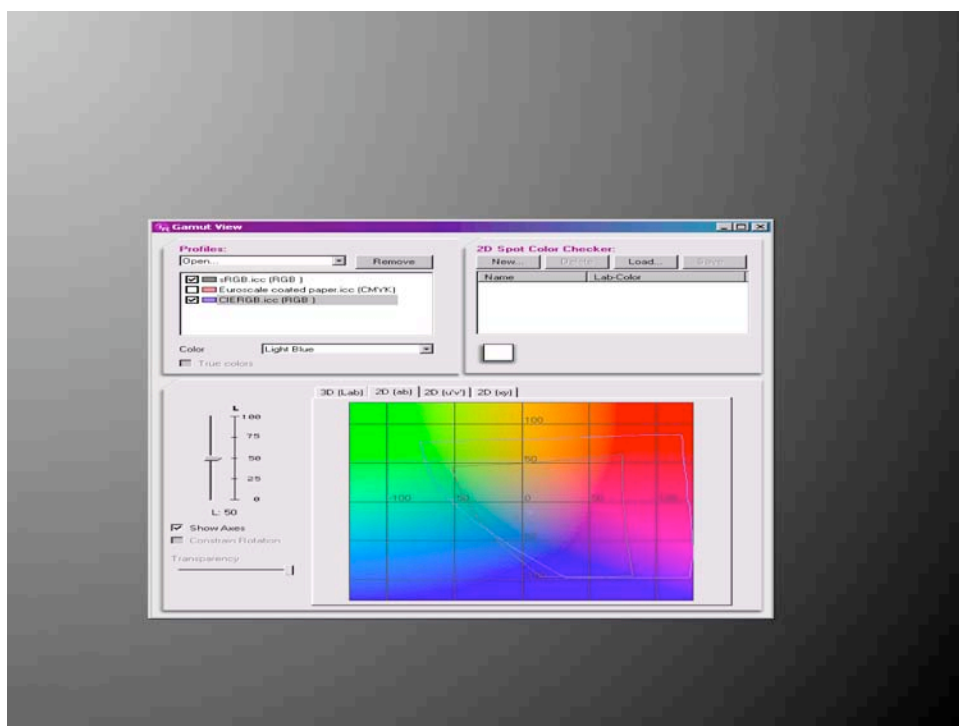
## What is color management?



A particularly tricky part of digital images is their relationship to the color of the source object. A color specialist should be consulted to make sure that your equipment is characterized and calibrated and that an appropriate color space is embedded directly into the digital image. At NYU Libraries we embed all images with the international color space CIE RGB which is a standard held by the Commission Internationale de l'Éclairage. Adobe RGB is another often used color space for archival work. Above is an example of one generation of loss from an unmanaged color file.



Bad color management will obscure the color of the original making it much less useful for researchers.



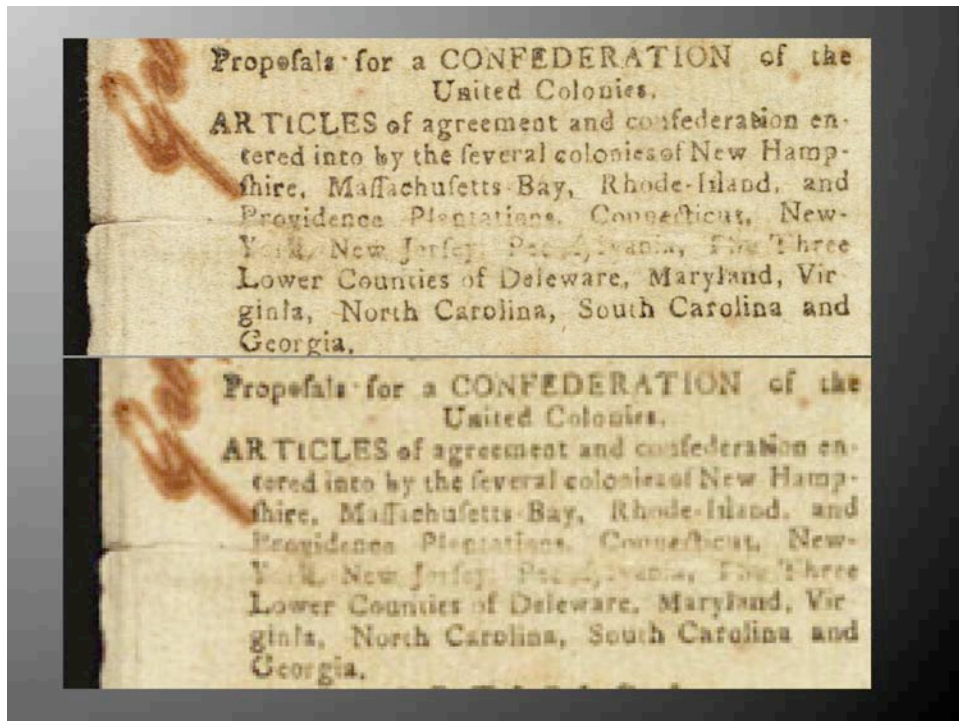
Color management is a science. There is important mathematical mapping that happens in digital images that must be maintained.

## Resolution

- Higher resolution generally better up to a point. Generally 400 ppi/600 ppi is acceptable for reflective material.
- Ultimately depends on nature of the original material (lantern slides, slides, glass negatives)

Most reflective material can be imaged at 400 or 600 pixels per inch. More detailed originals such as slides, glass negatives, transparencies and others will need a much higher resolution (often up to 4000 pixels per inch) to capture all of the detail from an original analog source.





600 ppi versus 72 ppi



## Other technical considerations

- ▣ Bit depth
- ▣ Embedded color space
- ▣ Digital damage done in Photoshop or other image tools

There are some other important technical characteristics to consider. Imaging specialists will be able to discuss these characteristics with you. Please know that cleaning pictures up in imaging software such as Adobe Photoshop after the fact results in badly damaged files. Correct files need to be made at the time of capture.

## **Preservation Metadata needs**

# Descriptive Metadata



MARC, [Archivist Toolkit](#), Dublin Core, EAD and VRA  
Core, DAC, among others

This information is created by the people who understand the material the best, the subject specialist or cataloger.

# Descriptive needs

- ▣ Catalog (MARC)
- ▣ Finding Aid (Archivist Toolkit)

COLLECTION TITLE: Nao Bustamante video collection TITLE: <b>Sans gravity</b>	RUN-TIME (HH:MM:SS): 00:01:16
DATE: 2000 WORK TYPE: Performance., Performance art., One-person shows (Performing arts).	LANGUAGE: English
CREDITS: Nao Bustamante, producer, creator, CAST/PERFORMERS: Nao Bustamante.	SYNOPSIS: Nao Bustamante ( <a href="http://www.naobustamante.com">www.naobustamante.com</a> ) is an internationally known performance artist born in California. Her work encompasses performance art, installation, video, pop music and experimental rips in time. She has performed in galleries, museums, universities and underground sites throughout Asia, Africa, Europe, New Zealand, Australia, Canada, Mexico and the United States. She is currently teaching New Media and Live Art at Rensselaer Polytechnic Institute, New York. 'Sans Gravity' is a performance/action in which Nao introduces her head into a plastic bag filled with water and then tapes it tightly onto her neck. She then sits down upright with the bag on her head. The action creates a real and urgent situation that the artist has to respond to. It ends when Nao, after a minute and twenty seconds, gets free from the water-bag finally breathing like a fish out of water.
KEYWORDS: Ritual in art, Water in art, Sensory deprivation, Hispanic American artists, Performance art, Ritual performance, Vulnerability in performance, Latino artists	COPYRIGHT INFORMATION: "Materials of the Hemispheric Institute Digital Video Library are protected by copyright. They may not be copied, downloaded, or reproduced. The owner of this work has granted NYU Libraries non-exclusive rights to include this material in the Hemispheric Institute Digital Video Library and to make it accessible to the public for educational and research purposes. Requests to purchase or for permission to use the work should be directed to the owner." COPYRIGHT HOLDER: Nao Bustamante ADDRESS: EMAIL: <a href="mailto:nao@naobustamante.com">nao@naobustamante.com</a> WEBSITE: <a href="http://www.naobustamante.com">http://www.naobustamante.com</a>
PERMANENT URL: <a href="http://hdl.handle.net/2333.1/orfjdc9">http://hdl.handle.net/2333.1/orfjdc9</a>	

## Technical Metadata



MIX, pbcore, textMD,  
NISO standards,  
exif, UPDIG

This type of metadata is generally created by digital libraries or vendors who are migrating analog material (books, documents, multimedia) to digital files.

## Structural Metadata



Digital objects do not have the intuitive structuring of the source material. Multi-volume sets, fold out pages, inserts, materials related to multimedia, etc. must all be structured so that the digitized object can be understood clearly by researchers.

## **Administrative, Provenance and Rights Metadata**

These types of specialized metadata are best handled by experts. Rights metadata help digital libraries control items that have not cleared copyright.

## Repository Overview

- ▣ Validation
- ▣ Fixity
- ▣ Preservation Goals

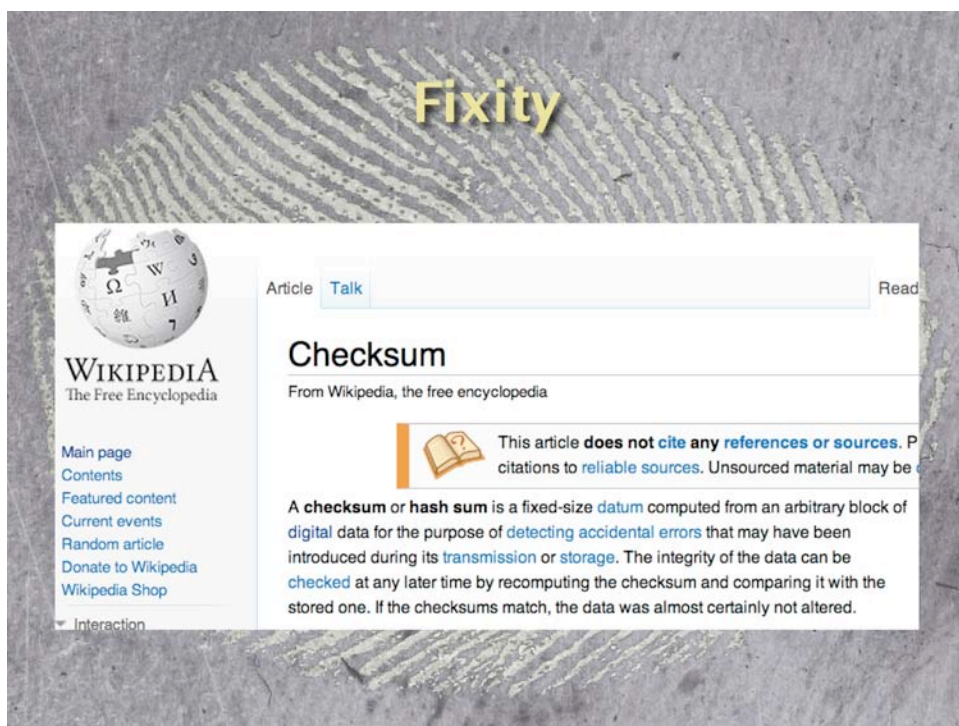
To preserve digital files they need to be held in an IT (Information Technology) structure that preserves the files and the information about the files. This is a very complex field. Following find an overview of information about repository needs.



# Validation

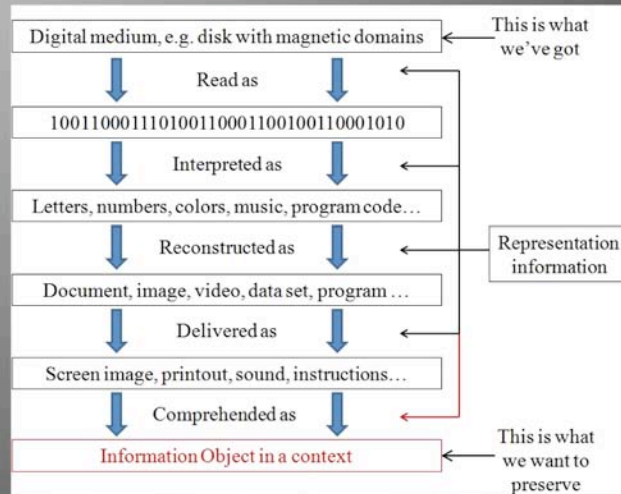


Validation is the process of evaluating a digital object to determine whether it meets the standard that it purports to meet. For example a scanner can make an image file such as “object.tif”. This file may open in Adobe Photoshop which is built to open anything it can, valid or not. The file may however be fundamentally flawed or out of standard. Over time such invalid files will become illegible. jHove, <http://hul.harvard.edu/jhove/>, created by Harvard University Libraries, will validate an image file and inform you that your file is “well-formed and valid”. This is a critical step in creating preservation images.



NYU Libraries currently uses SHA-256 checksums. Running checksums in accordance with good IT principles will guarantee that your files have not changed over time.

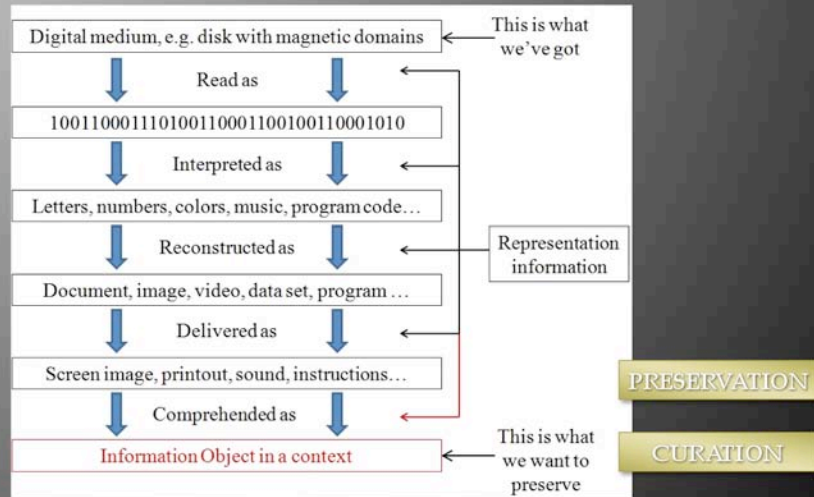
# Preservation Goals



Special thanks to Robert DeCandido, Database Coordinator, The Morgan Library and Museum

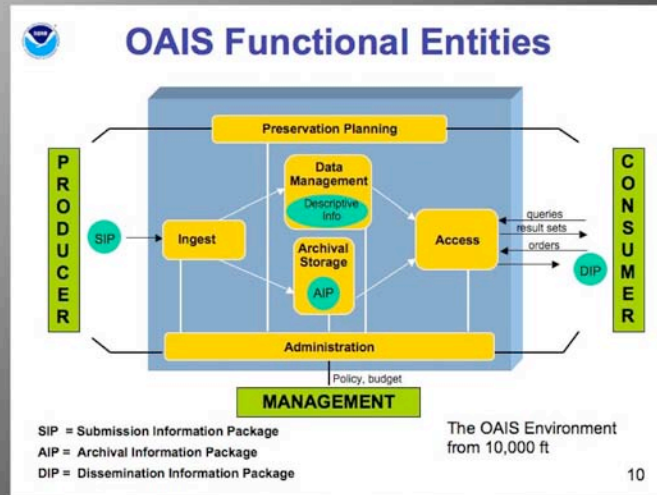
From the long view we need to remember that preservation is about saving not only the information object, but also all that supports it. This includes a complex hierarchy of information that has to be preserved to allow the digital file to be perceivable to the researcher. To see the image of a digitized page the screen image has to be comprehended. Behind that lies layers of graphical interfaces, metadata instructions, binary information etc. All have to be preserved and managed so that the end result is human readable.

# Preservation Goals



Special thanks to Robert DeCandido, Database Coordinator, The Morgan Library and Museum

# OAIS model



<http://www.oclc.org/research/publications/archive/2000/lavoie/>

A preservation repository is much more than a backed up computer.

□



Thank you.