Down the Rabbit Hole of Image Storage and Retrieval in CONTENTdm

By: Gillian (Jill) D. Ellern

Abstract

Do you know the difference between a display image and a thumbnail image in CONTENTdm? How and why do you pick between a jpg, jpg 2000, pdf, or tif image formats as you scan an image for loading into the Project Client? Are you confused about the difference between a dpi, a ppi or megapixels? Why does one image that you scan fill the screen while another displays so small you can barely see it? This presentation explores these questions and more. Using this map of rabbit hole (a workflow diagram of CONTENTdm) will help you to better understand the software as you process and upload images.

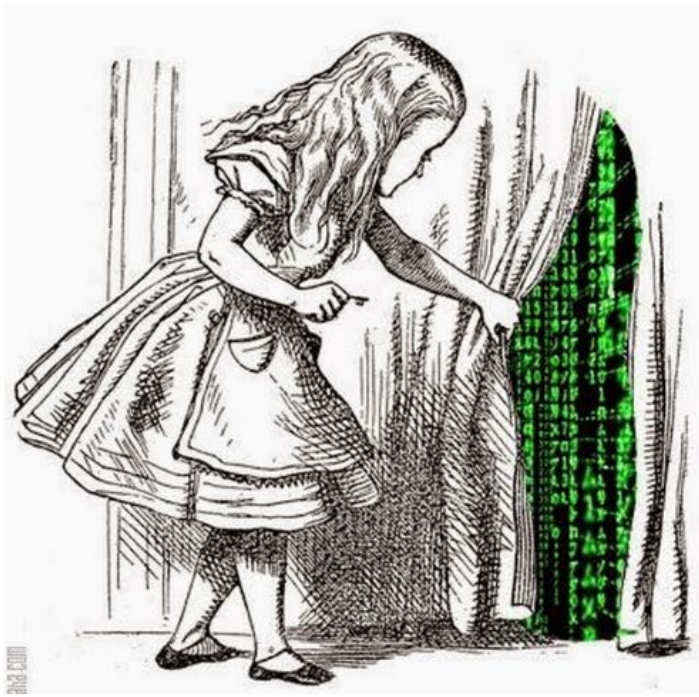
Ellern, G.D. (2018, August 2). *Down the Rabbit Hole of Image Storage and Retrieval in CONTENTdm* [Presentation]. CONTENTdm User Group Meeting, Columbus, Ohio.

Archived version from NC DOCKS available at: http://libres.uncg.edu/ir/wcu/listing.aspx?styp=ti&id=37395.

DOWN THE RABBIT HOLE WITH IMAGE STORAGE AND RETRIEVAL IN CONTENTdm

CONTENTdm Community Insights
Webinar May 2021

Gillian (Jill) D. Ellern
Western Carolina University



Presentation Outline

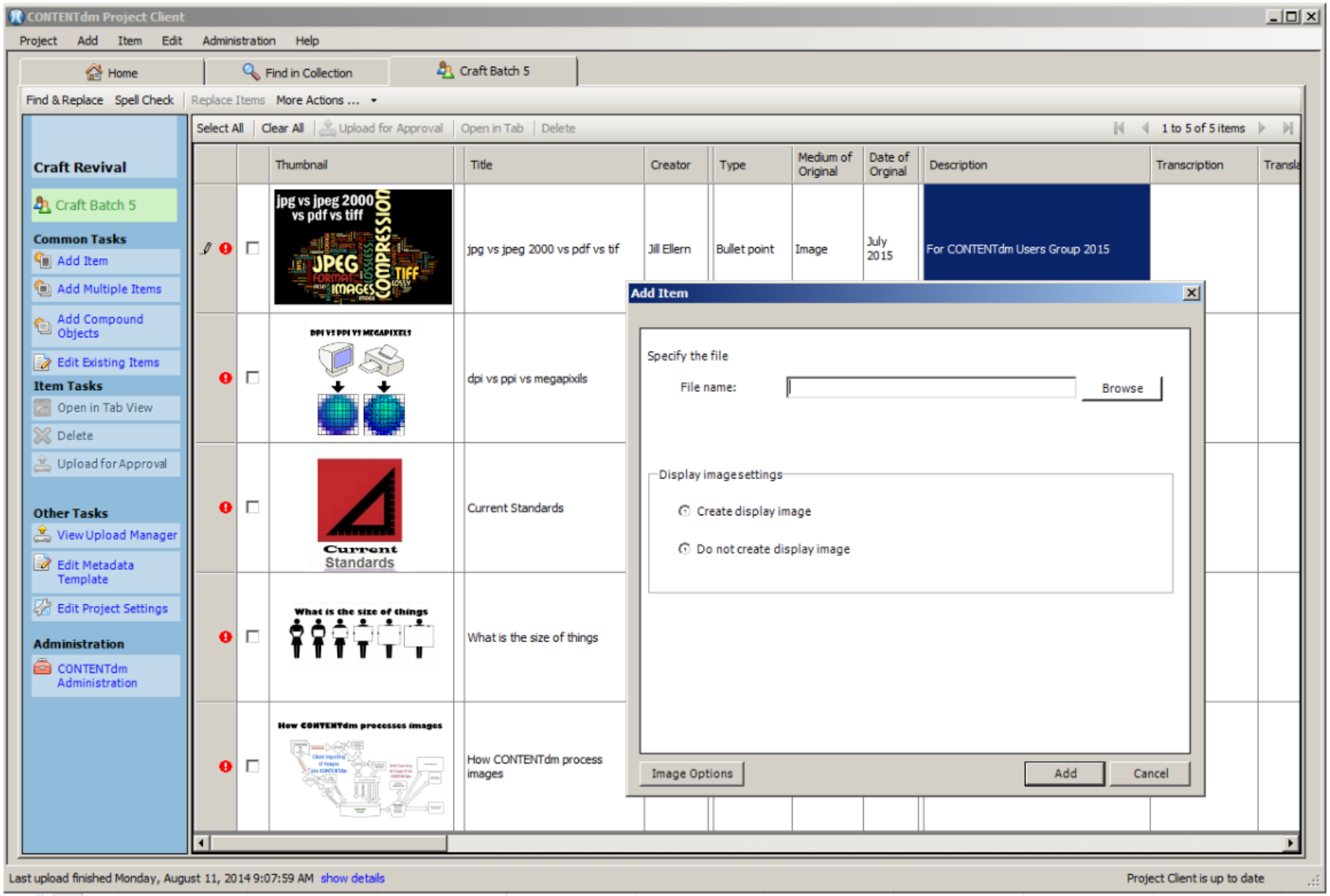
The question that started my research

About me

Down the rabbit hole with tons of questions

Four areas of answers I found

Process of Images in and out of CONTENTdm


CONTENTdm Project Client
Project Add Item Edit Administration Help
Home
Find in Collection
Craft Batch 5
Find & Replace Spell Check Replace Items More Actions ...
Craft Revival
Craft Batch 5
Common Tasks
Add Item
Add Multiple Items
Add Compound Objects
Edit Existing Items
Item Tasks
Open in Tab View
Delete
Upload for Approval
Other Tasks
View Upload Manager
Edit Metadata Template
Edit Project Settings
Administration
CONTENTdm Administration
Select All Clear All Upload for Approval Open in Tab Delete
1 to 5 of 5 items
Thumbnail
Title
Creator
Type
Medium of Original
Date of Orginal
Description
Transcription
jpg vs jpeg 2000 vs pdf vs tif
Jill Ellern
Bullet point
Image
July 2015
For CONTENTdm Users Group 2015
dpi vs ppi vs megapixils
Current Standards
What is the size of things
How CONTENTdm processes images
How CONTENTdm process images
Add Item
Specify the file
File name:
Browse
Display image settings
Create display image
Do not create display image
Image Options
Add
Cancel
Last upload finished Monday, August 11, 2014 9:07:59 AM show details
Project Client is up to date

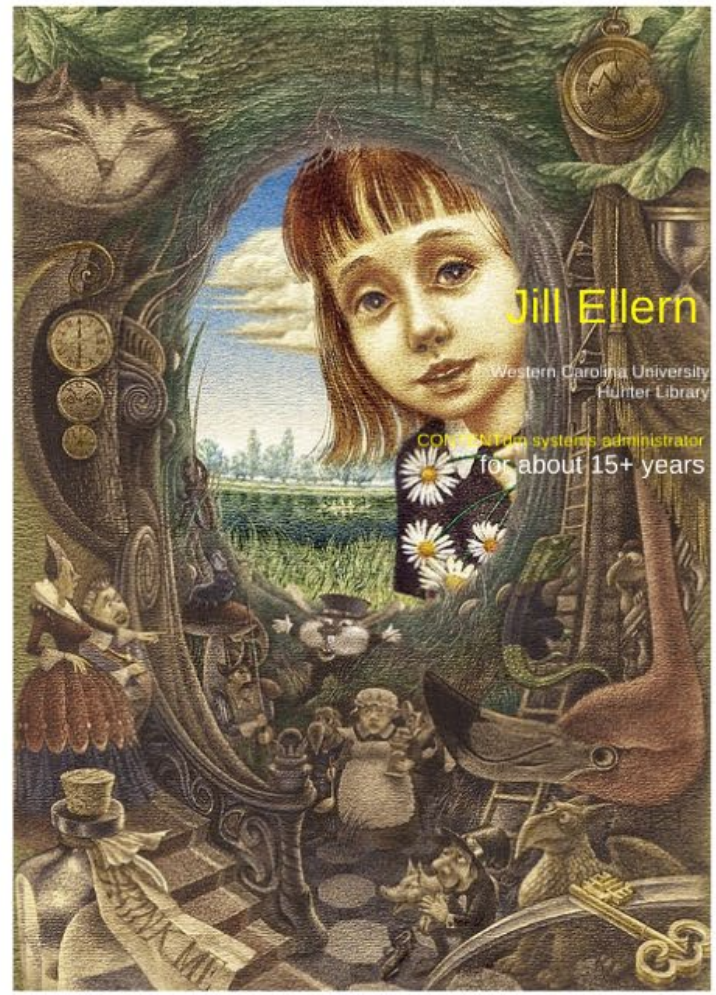
Should we be checking this "Create a display image" button?

Much more to that answer than I ever expected...

It led to TONS of questions...

Led to more questions relating to answers

Which led to more.....

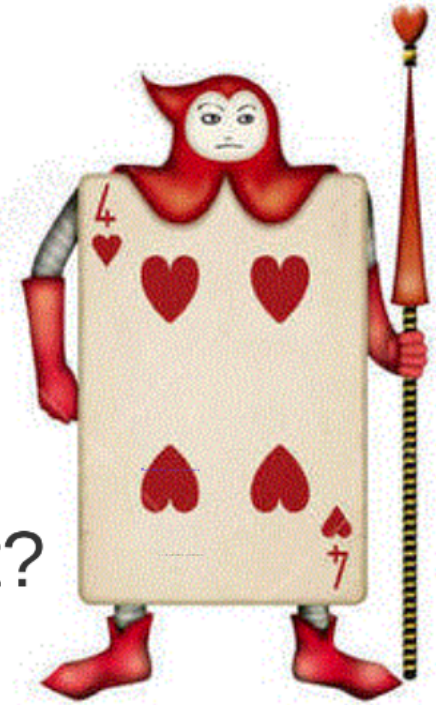



KEEP
CALM
AND FOLLOW THE
WHITE
RABBIT

CONTENTdm Community Insights Webinar: Down the Rabbit Hole with IMAGE STORAGE AND RETRIEVAL IN CONTENTdm


Should we be checking this "Create a display image" button?
Drink me
eat me

Display Images?



What exactly is the display image?

Are there benefits if we let the Cdm client create it?

Are there ramifications of deciding NOT to create one?

How does it create it and what are the settings?

More display image questions

How is the display image different from the image we add?

What happens to my original image if I create one?

How is a display image different from the "thumbnail"?

What about our standard jpeg image?

The standard "DPI" settings we've been using for years....

If we create display image, do we need to modify those setting?

What are the current recommended image standards?

Where on the web are those standards located?


Technology has advanced since we set these standards
Should we consider...
Increasing the "dpi" of an image?
Changing compression rates?
Other image formats (Jpeg 2000)?
Would patrons even notice?
Perhaps some our decisions were influenced by disk space and storage fees...
What about the CONTENTdm's archiving function?

What about Jpeg 2000?

OCLC changed some of our old files to this format

How is it different from jpeg as a file format?

Should we change it to be our primary file format?

What are the ramifications of changing to this format?

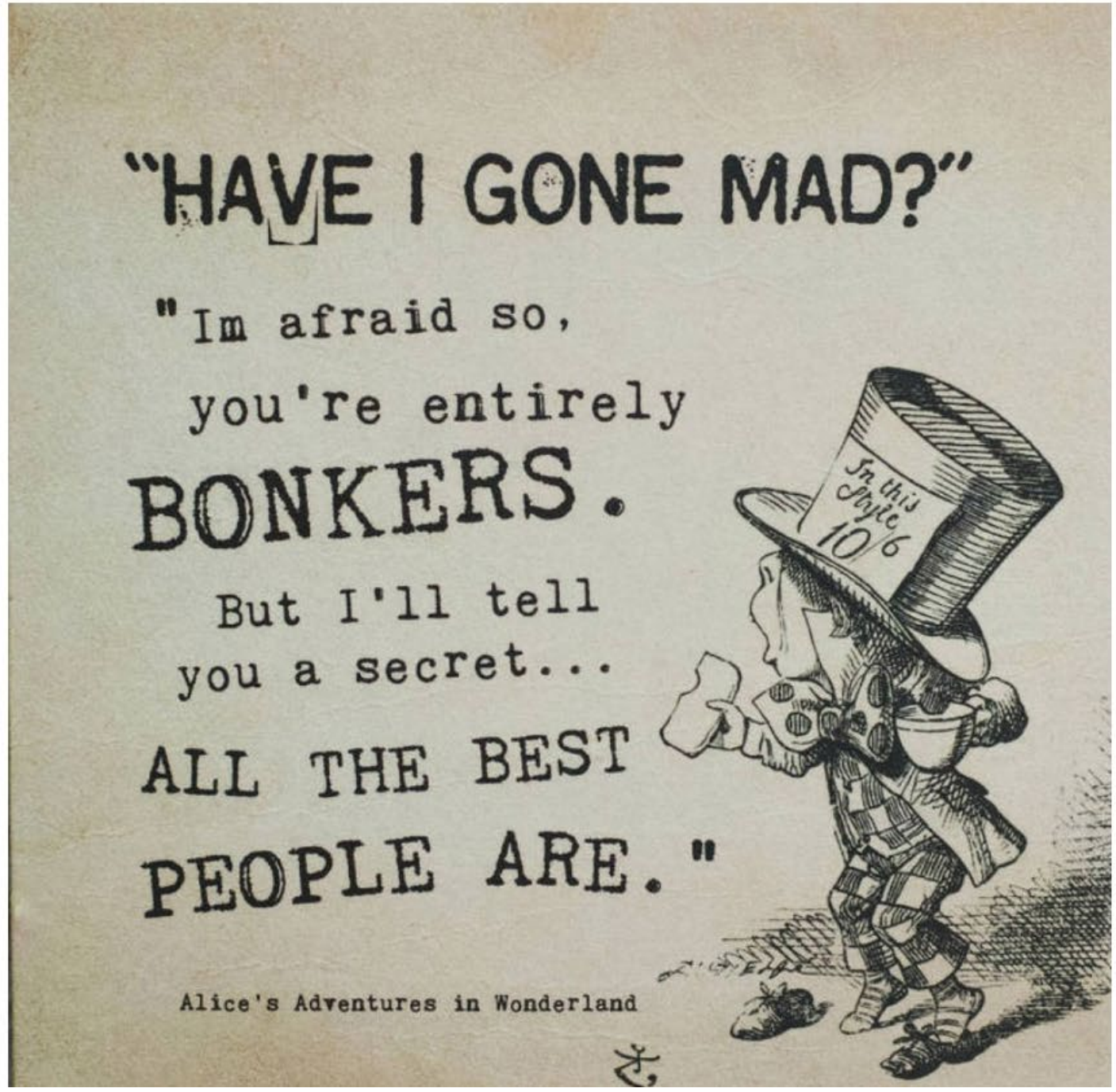
Purpose of CONTENTdm

Is it a display device? Publicity vehicle? Research tool? Storage device?

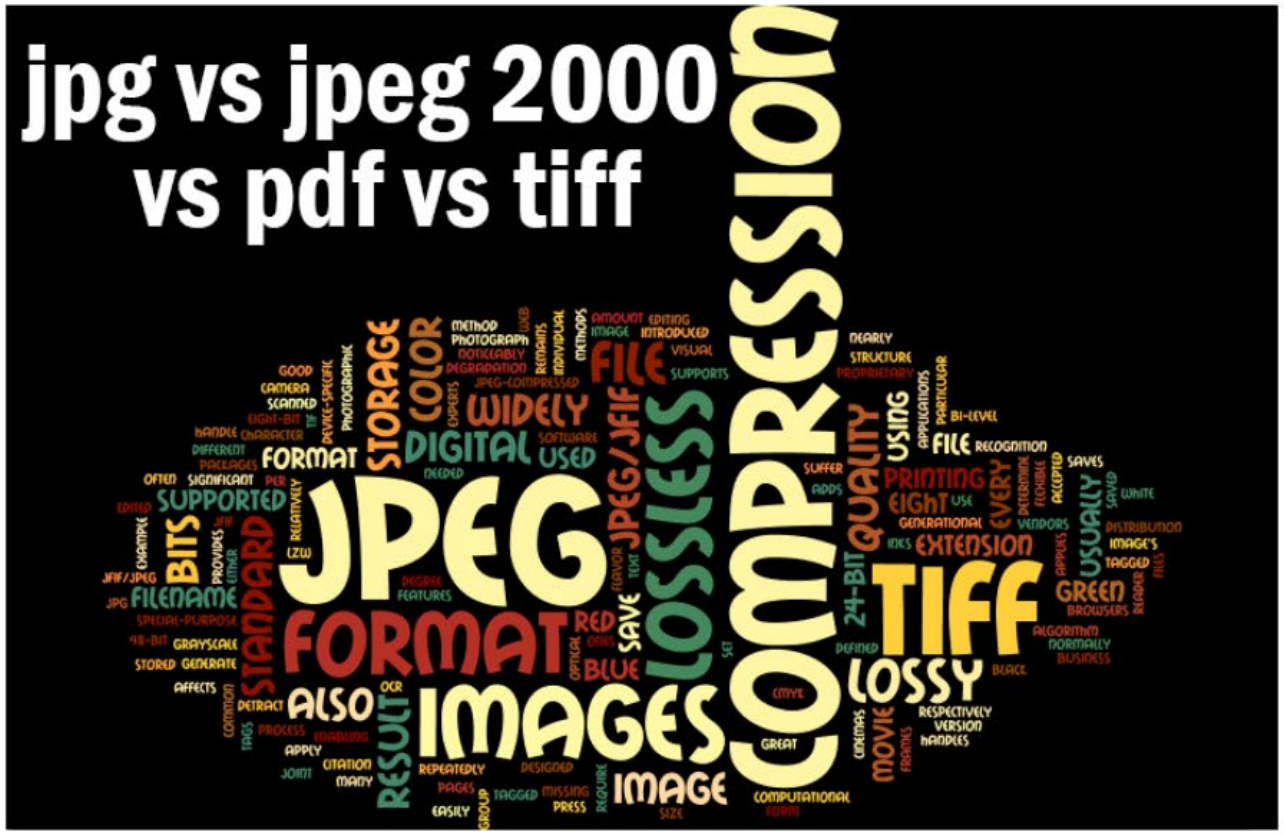
All of these?

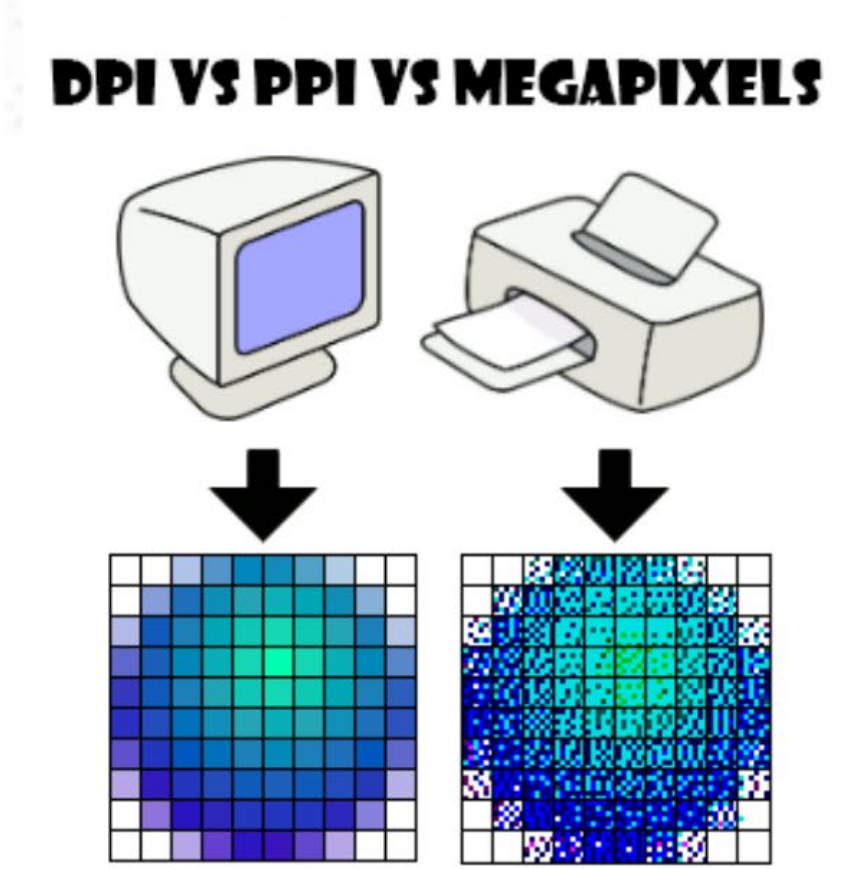
Does it do all these functions well?

Are we using all of them well?

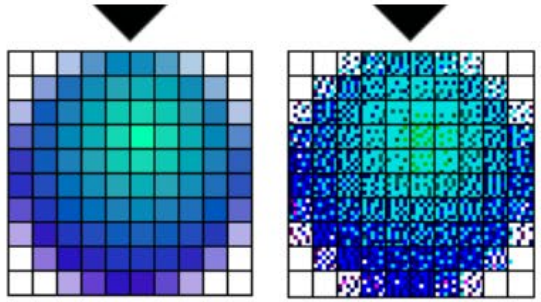


some other fundamental qu

		Thumbnail	Title
			jpg vs jpeg 200

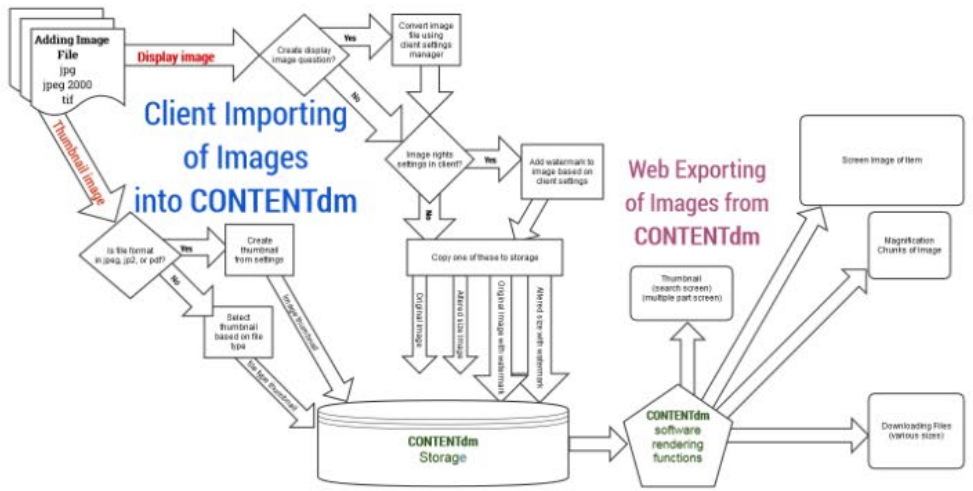


dpi vs ppi vs m



Current Standards

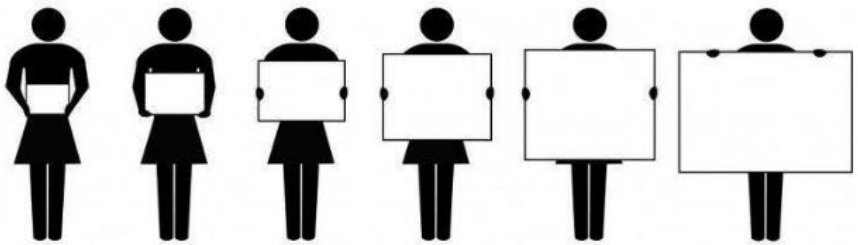
How CONTENTdm processes images



How CONTENTdm pro
images

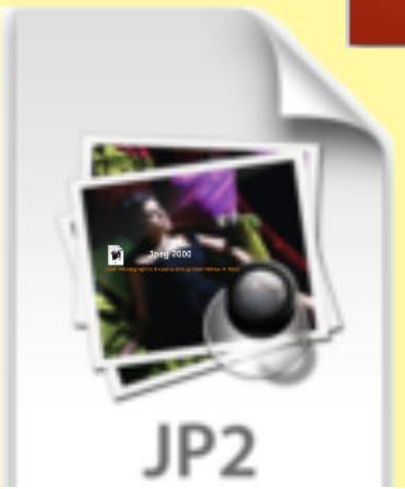
Current
Standards

Current Standards



What is the size of thin

jpg vs jpeg 2000 vs pdf vs tiff



jpg or jpeg

Joint Photographic Expert Group

- Has been the de facto standard for web images
- Uses redundancies in file to compress data
- Can be compressed significantly
- "Lossy" - you lose data when compressing
- Not ideal for text, cartoons, line drawing, crisp lines or sharp contrasting edges

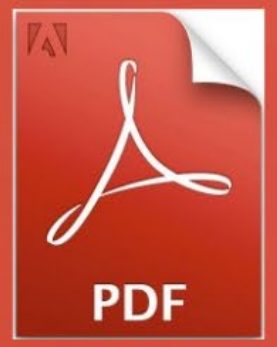
Original image, with good color grade

Loss of edge clarity and tone "fuzziness" in heavy JPEG compression

Jpeg 2000

Joint Photographic Experts Group Committee in 2000

- Shares the same name but very different underneath
- Uses a different scheme to compress the file
  - Discrete wavelet transformation vs. cosine transformation
- Better image quality at smaller sizes (even at 80:1)
- Can use both lossy and lossless options
- Web browsers often require a plugin to use this format
- Didn't take off as an industry standard



pdf

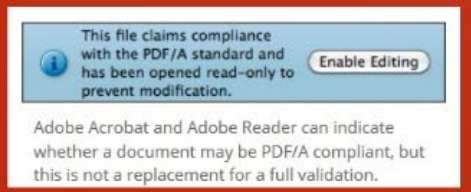
Portable Document Format

- Designed specifically for text (page oriented documents)
- Not so good for images (pixelates them)
- Allows for scaling, zooming and printing using a reader
- Can have limitations on document size (10GB+)
  - + 20Mb not displayed in Cdm inline - use compound objects
- Accessible for disabled persons
- Can be encrypted (password protected)

pdf/a

Portable Document Format - Archiving

- Specialized version of pdf for digital preservation of electronic documents
- Documents are self-contained so contents are rendered exactly the same way each time
- Multiple levels of pdf/a compliance (A-1a, A-1b, A-2 and A-3)
- Embeds all the fonts so often larger files
- No standard validation tool



tiff or tif

Tagged Image File Format

- Standard universal format for storing high quality images
- Used as a major format for most digital reproductions
- Has a variety of subtypes - Latest version 6.0 (1992)
- Lossless format
- Minimal compression options
- Very large sized files
- Not a native format for most devices
  - RAW is more common and more useful

Why change to jpeg 2000 (jp2)?

Increased compression of file size

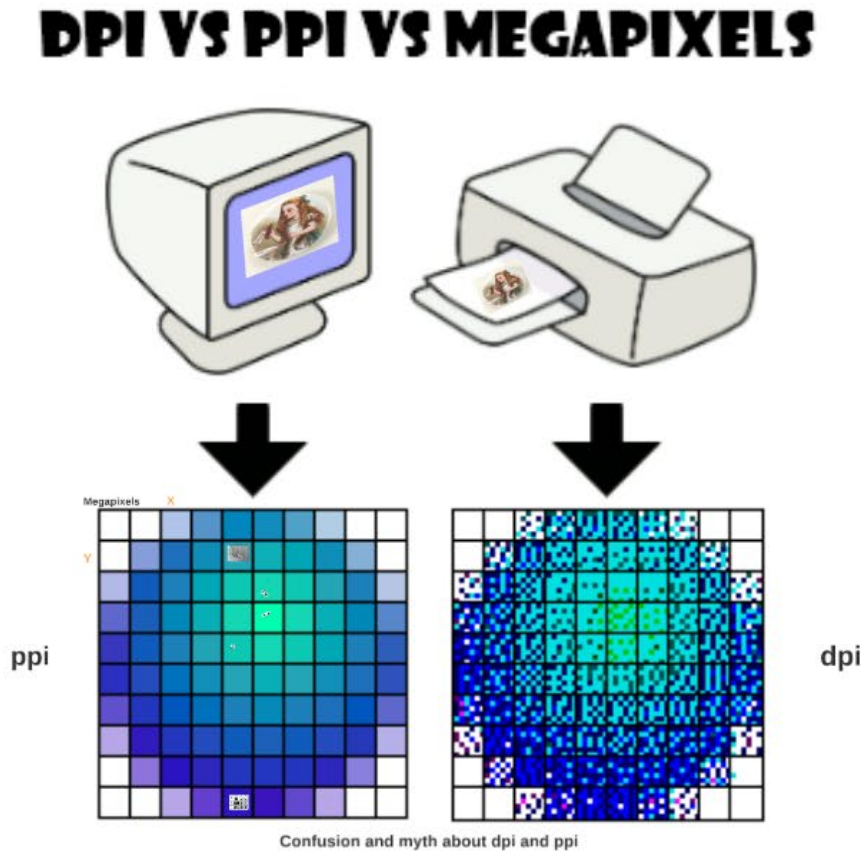
- 20% smaller then same jpeg
- Converting from a tiff to jp2
  50% file reduction, lossless

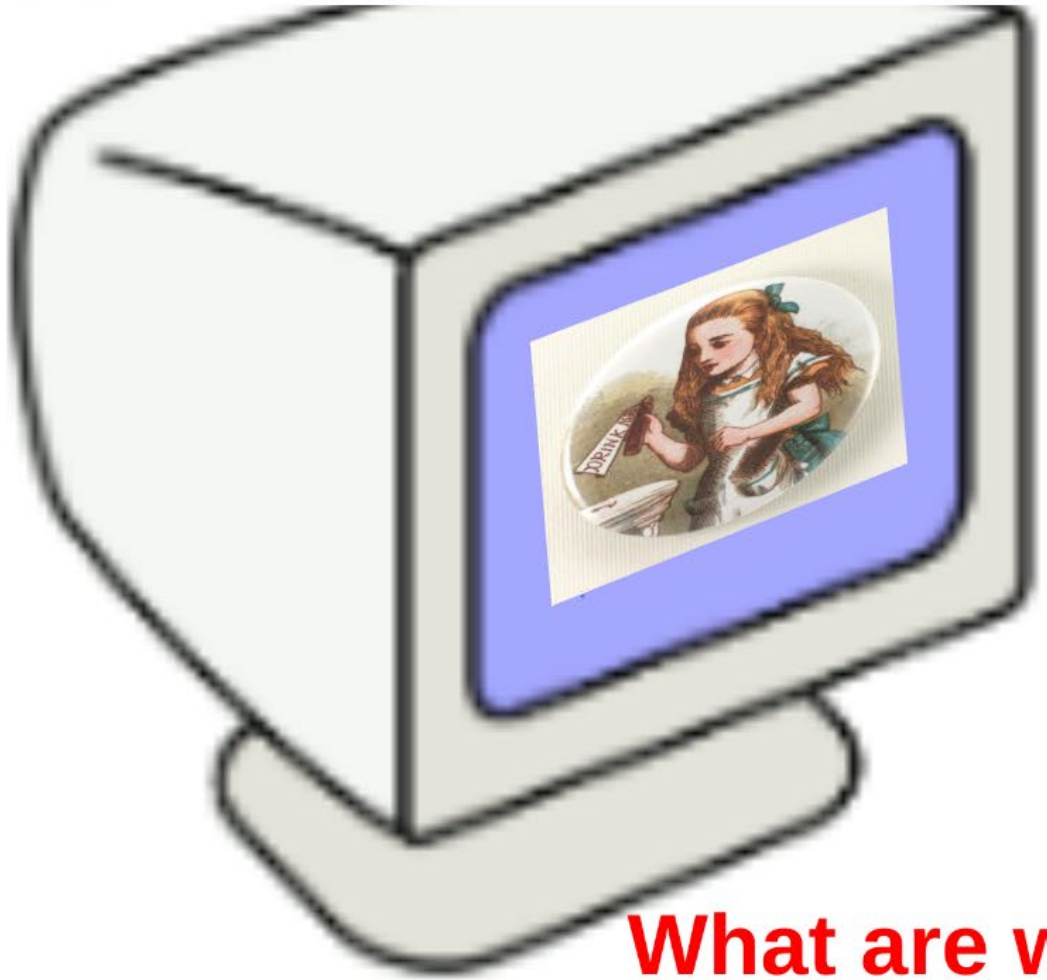
Compression of both lossless and lossy in one

Works well with a variety of images

Uses a progressive display

Increasingly becoming a viable format for archival images


DPI VS PPI VS MEGAPIXELS
Megapixels
X
Y
ppi
dpi
Confusion and myth about dpi and ppi



What are we trying to achieve?

High Digital Image Quality

"Full resolution image"

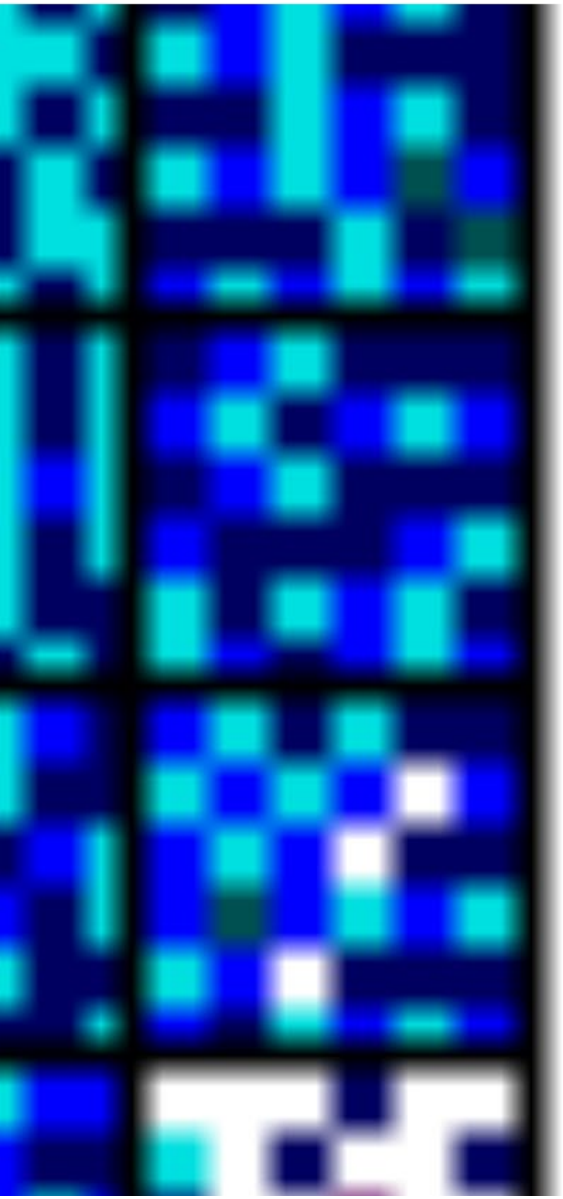
How do we measure quality in a digital image?

We do it with image resolution!

Higher resolution means more image detail.

dpi

ppi



dpi

dots per inch

Used for printing an image on paper

How many tiny, tiny droplets of ink are laid out on a page of print

ppi

pixels per inch

Used for displaying images on a monitor

How many pixels by length (x) and width (y) are created by a device to describe an image

Confusion and myth about dpi and ppi

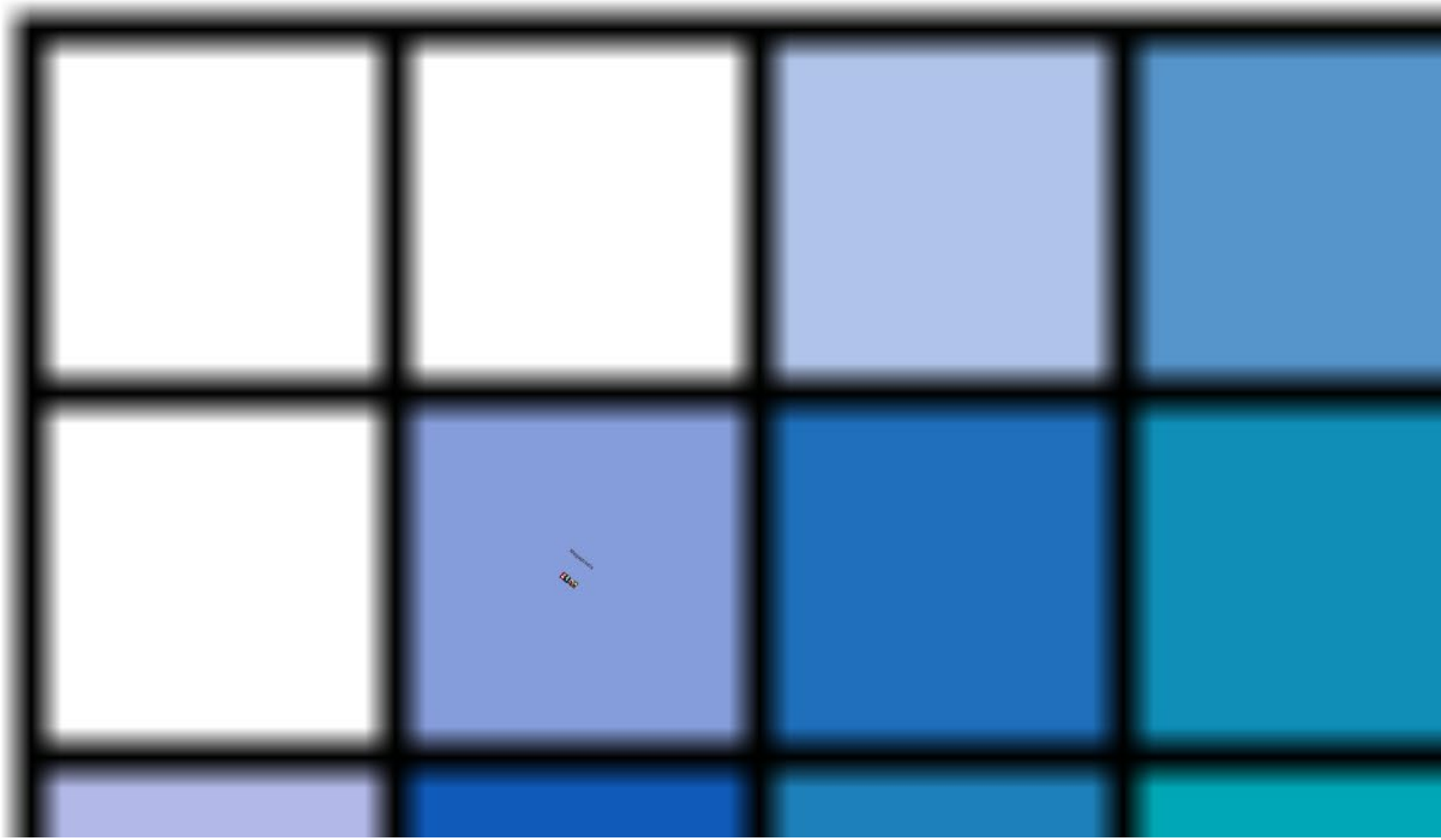
- Used interchangeably but they are not the same
- 1 pixel does not equal one printed dot Hasn't for decades
- Technically even scanners scan "pixels" and not dots
- The designations simply refer to a conversion calculation

Not even about digital resolution, it's pixel dimensions

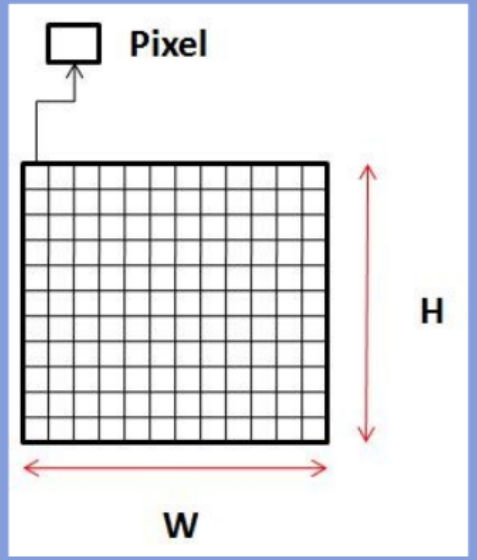
Megapixels

X

Y



Megapixels



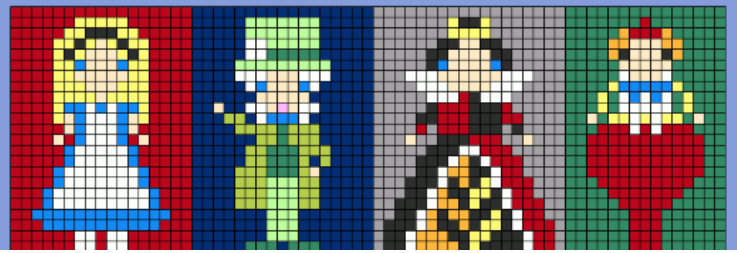
one million pixels

X * Y = #Mpx

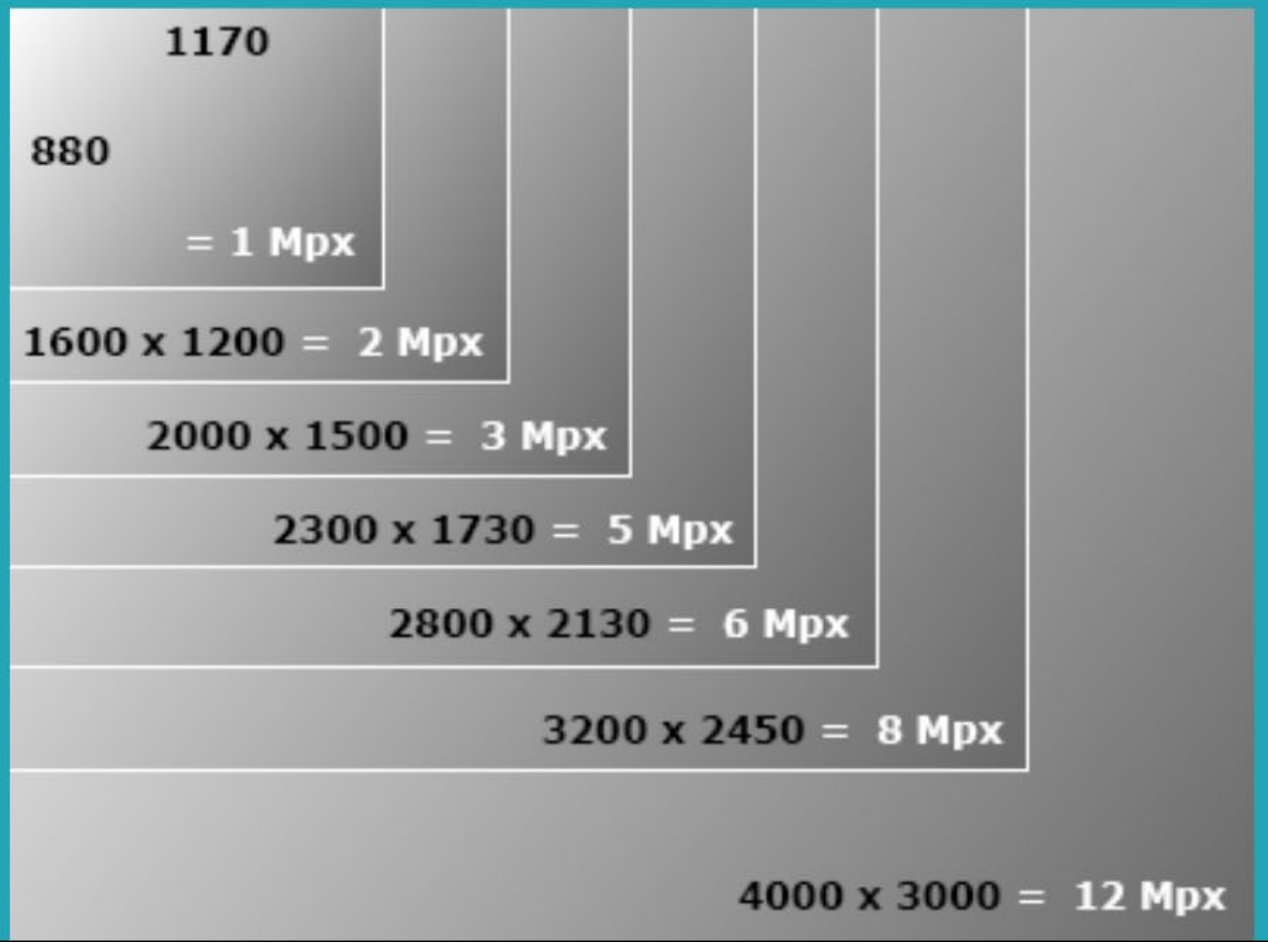
Rounded into single number for X and Y

Used for the number of pixels in an image

Express the number of image sensor elements in a camera



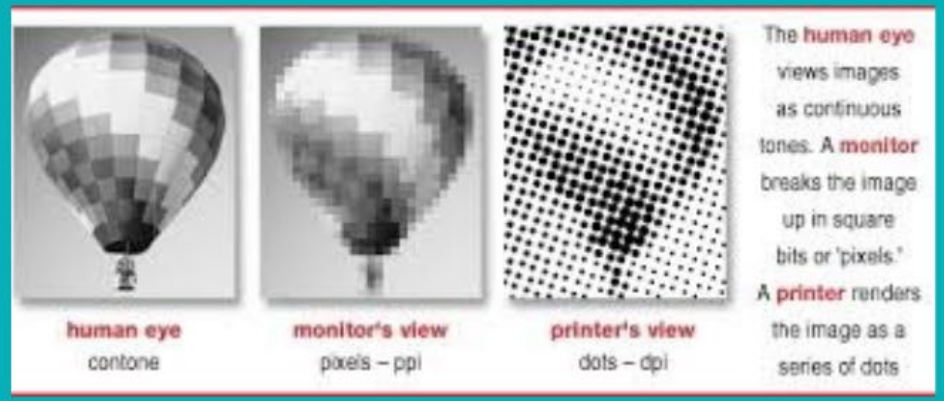
Pixel dimensions in Mpx



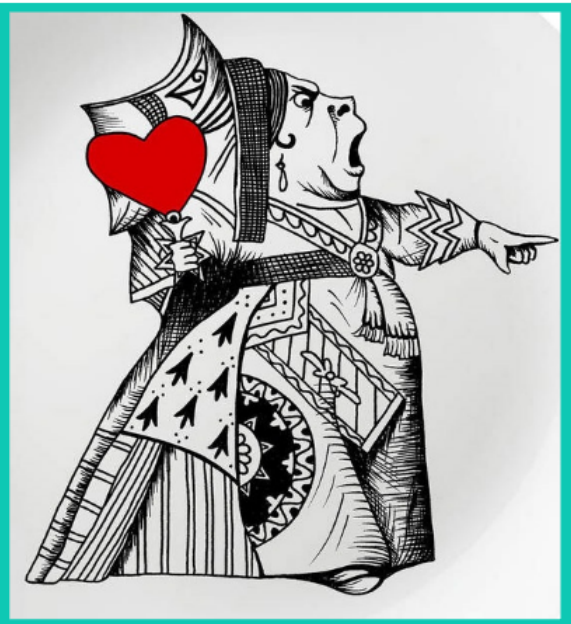
Our current standards

Using the same resolution needed to print them back again!

WEB ≠ PRINT



I don't think we are always asking ourselves what we are going to do with the image.

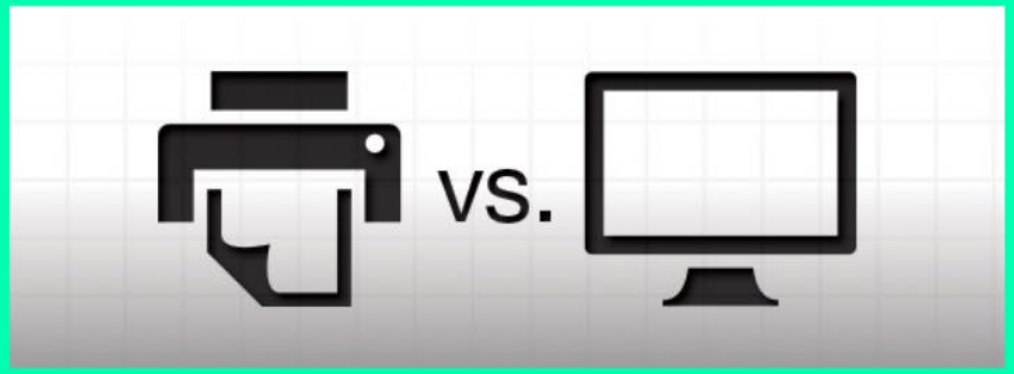


Digital Image Quality

Using just dpi or ppi will not give you a quality digital image!

Elements of Quality

- Size (in pixels) of the digital image
- Quality of the recording device (optics and sensors)
- Digital format it is stored in (lossless or lossy compression)
- Technical proficiency and eye of the photographer



Scanning a photo at 300 - 600 ppi

- Standard is for printing a "high resolution photo"

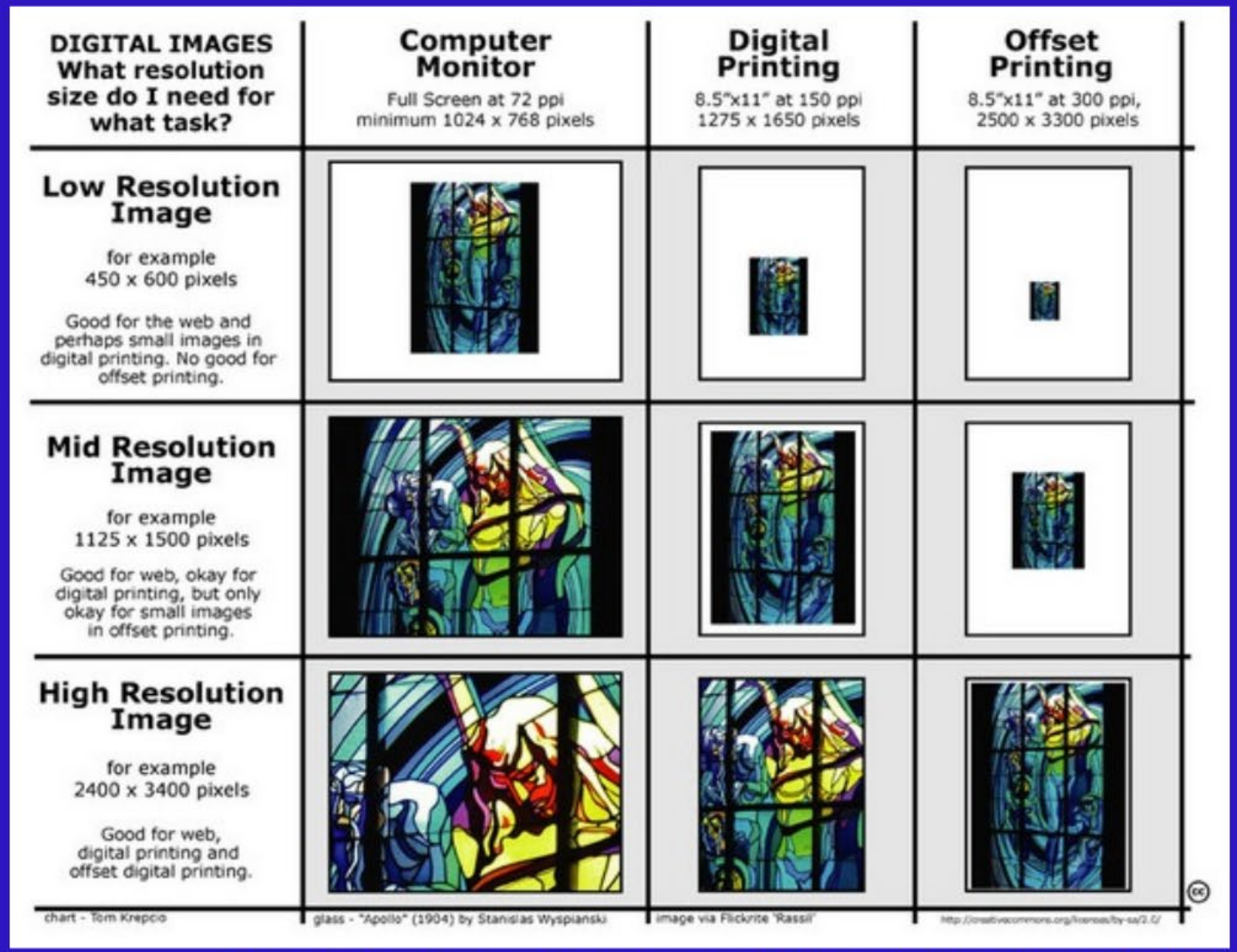
Viewing image on the web at 72 - 100 ppi

- Scanning far bigger then is needed for regular web use

Our technology today an scan much higher (12,800 x 12,800 ppi)

Compression rates (i.e. jp2), storage capacity, the cloud applications and storage, hand held display devices for imaging and display... are all changing the equation quickly

A need to continually evaluate standards



from http://www.krepcio.com/vitreosity/archives/2010_04.html

Image processing standards

Evaluating preservation vs researching needs

How much zooming and panning is really used or needed?

- Do tend to like a standard resolution for all image
- Don't tend to vary ppi per image
- Use same setting for images and text

How much zooming is going to happen with a text document in pdf?

What are your standards?

Current Standards

Federal Agencies Digitization Guidelines Initiative – 2016

http://www.digitizationguidelines.gov/guidelines/digitize-technical.html 

Technical Guidelines for Digitizing Cultural Heritage Materials

...from page 33

Prints and Photographs

Performance Level:

	1 Star	2 Star	3 Star	4 Star
Master File Format	TIFF	TIFF	TIFF	TIFF
Access File Formats	All	All	All	All
Resolution	100 ppi	200 ppi	400 ppi	600 ppi[1]
Bit Depth	8	8	8 or 16	16
Color Space	Grey Gamma 2.2 SRGB Adobe 1998 ProPhoto ECIRGBv2	Grey Gamma 2.2 SRGB Adobe 1998 ProPhoto ECIRGBv2	Adobe 1998 ProPhoto, ECIRGBv2	Adobe 1998 ProPhoto, ECIRGBv2
Color	Grayscale or Color	Grayscale or Color	Color	Color

FADGI guidelines uses a star system

"Our mission is to define what is practical and achievable today... FADGI does not recommend digitizing to less then three-star."

- One star imaging should only be considered informational, in that images are not of a sufficient quality to be useful for optical character recognition or other information processing techniques. One star imaging is appropriate for applications where the intent is to provide a reference to locate the original, or the intent is textual only with no repurposing of the content.
- Two star imaging is appropriate where there is no reasonable expectation of having the capability of achieving three or four star performance. These images will have informational value only, and may or may not be suitable for OCR.
- Three star imaging defines a very good professional image capable of serving for almost all uses.
- Four star defines the best imaging practical today. Images created to a four star level represent the state of the art in image capture and are suitable for almost any use.

Library of Congress Format Spec 2017-18

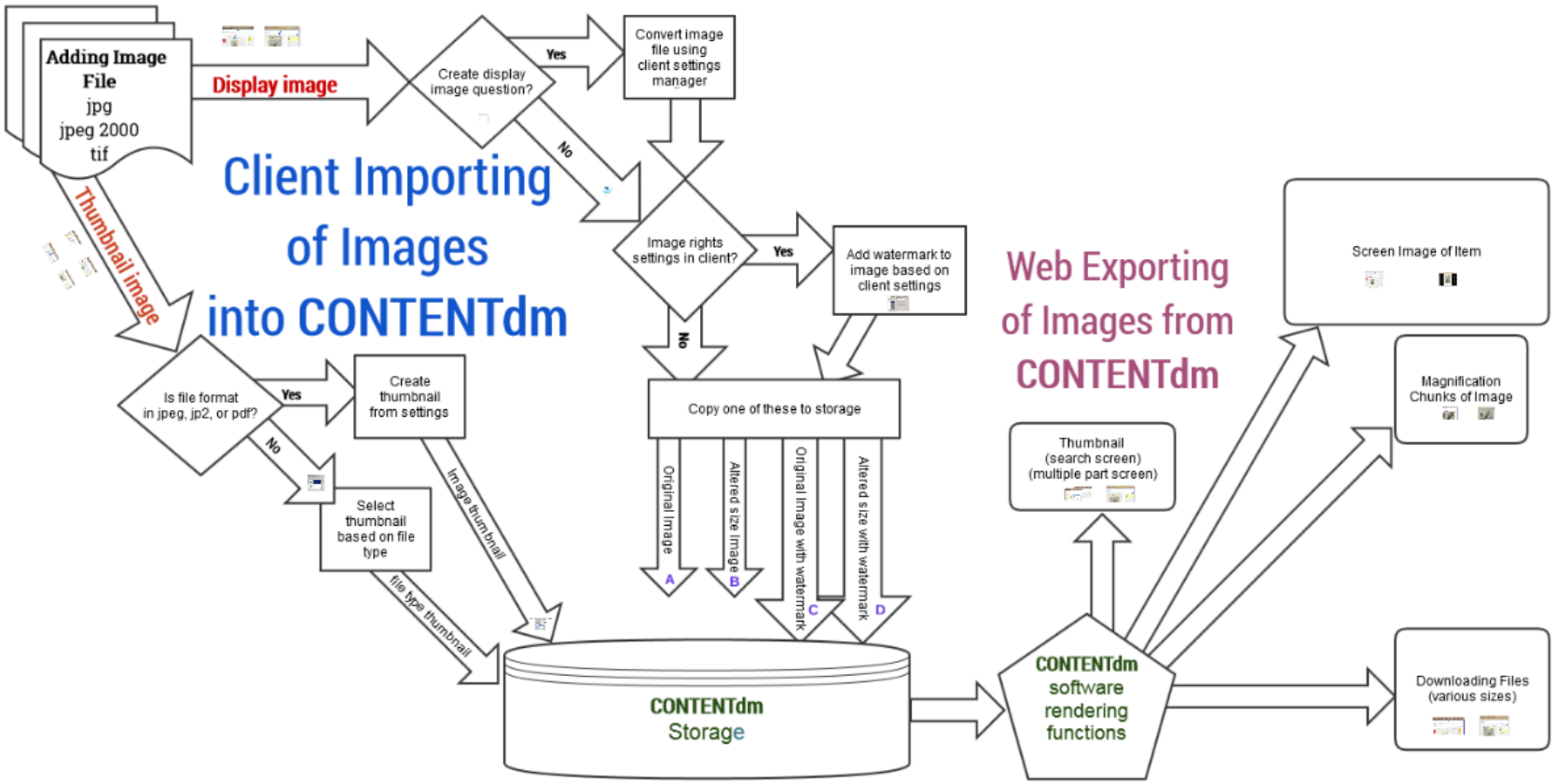
http://www.loc.gov/preservation/resources/rfs/stillimg.html

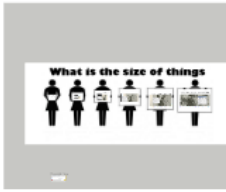
Digital photographs

In order of preference

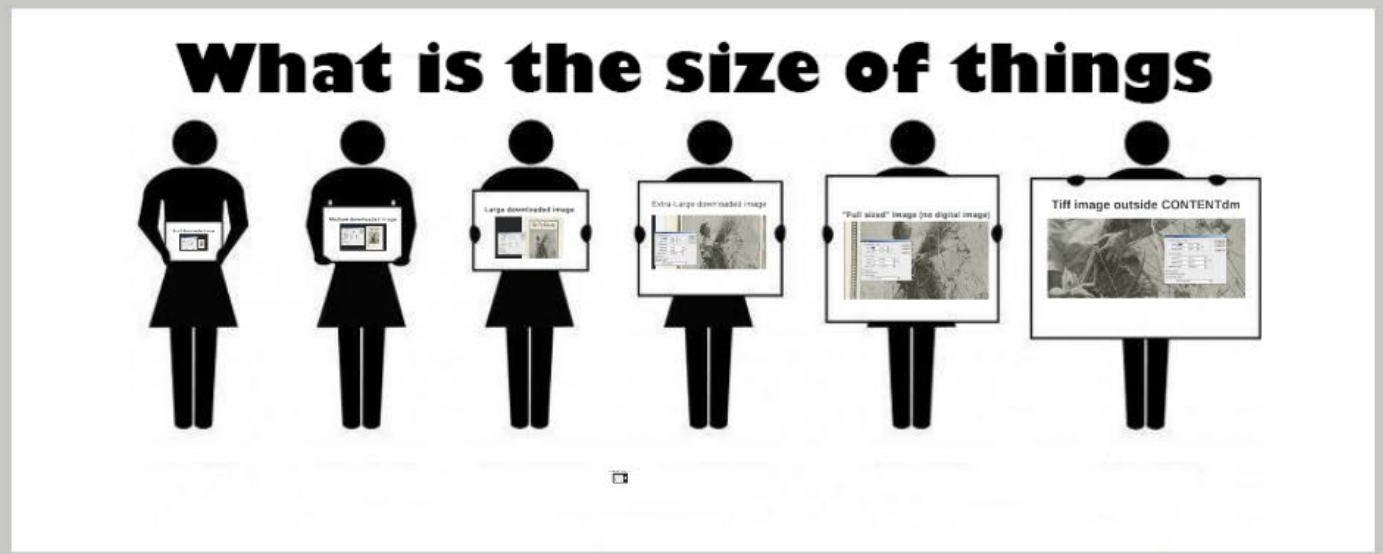
- TIFF (uncompressed)
- JPEG2000 (lossless (*.jp2)
- PNG (*.png)
- JPEG/JFIF (*.jpg)
- Digital Negative DNG (*.dng)
- BMP (*.bmp)
- GIF (*.gif)

How CONTENTdm processes images





CONTENTdm Community Insights Webinar: Down the Rabbit Hole with IMAGE STORAGE AND RETRIEVAL IN CONTENTdm



Download menu

Purpose of CONTENTdm

Is it a display device?

Publishing vehicle?

Research tool?

Storage device?

Display Image Management System