



Workshop: The fundamentals of digitisation & their implications for archiving

South African Society of Archivists Conference
July 2018



Richard Higgs
richard.higgs@uct.ac.za
Twitter: @higgsri
www.lisc.uct.ac.za



Introductions

- Me
- You:
 - What scares you most?
(about digital archives)
- What scares me the most?
(about digital archives)



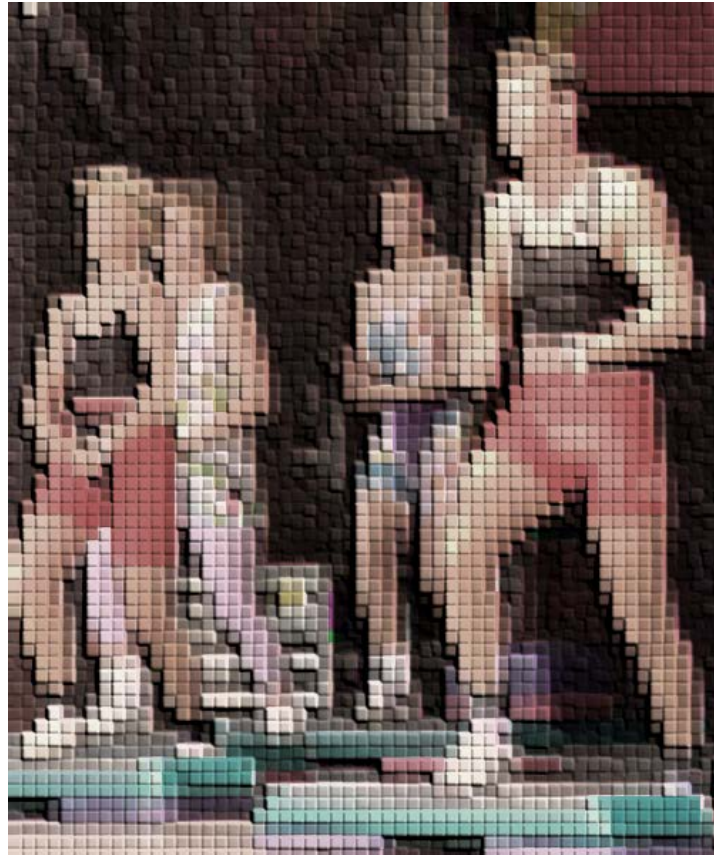
GAIL HAMPSHIRE / CC-BY-2.0



SIGNALS, ENCODING, MEDIA & FORMATS

Let's get digital

- Exercise



© User: ShinyFan / Wikimedia Commons / CC-BY-2.5

Analogue and Digital



BrokenSphere / Wikimedia Commons / CC-BY-SA



ParentingPatch / CC-BY-SA

Signals

- A change in the property of a something over time
- Do signals exist if we aren't there to see them?



Photographs by JarrahTree / Wikimedia Commons / CC-BY-2.5

Medium

- Carries a signal



Clay tablet with cuneiform inscription, Los Angeles County Museum of Art
Public Domain

Format

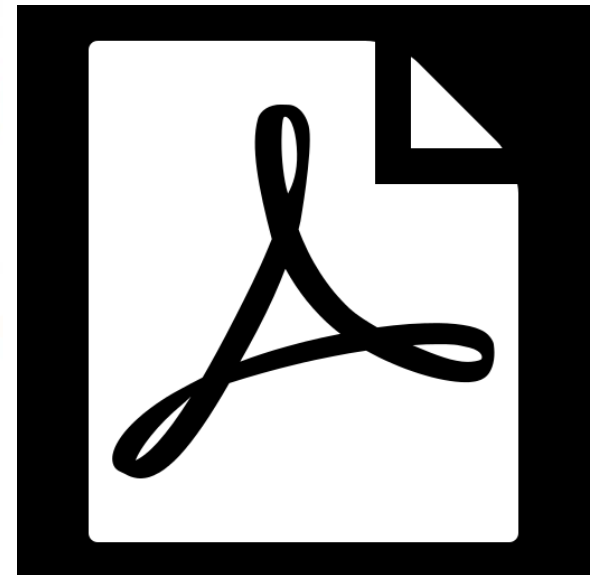
- The constraints on a medium
- The rules for getting a signal out of a medium



SOURCE: <https://obialone.wordpress.com>



SOURCE: <https://pressgazette.co.uk>



Encoding

- Exercise

5 , 24 , 5 , 18 , 3 , 9 , 19 , 5

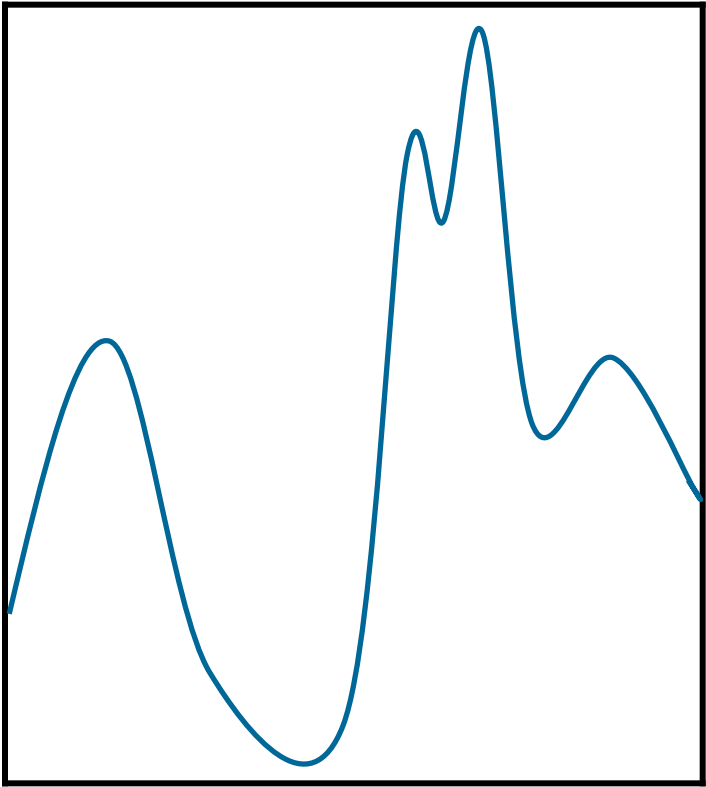


Transducers

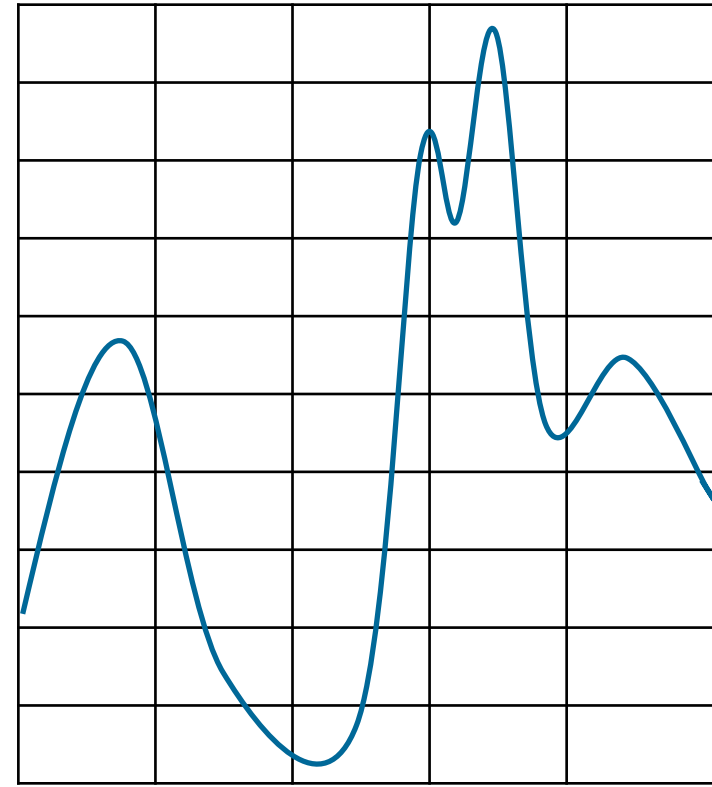
- Transfer signals from 1 medium to another
 - Why would we want to?
 - Examples?

Converting analogue to digital signals (1)

Analogue

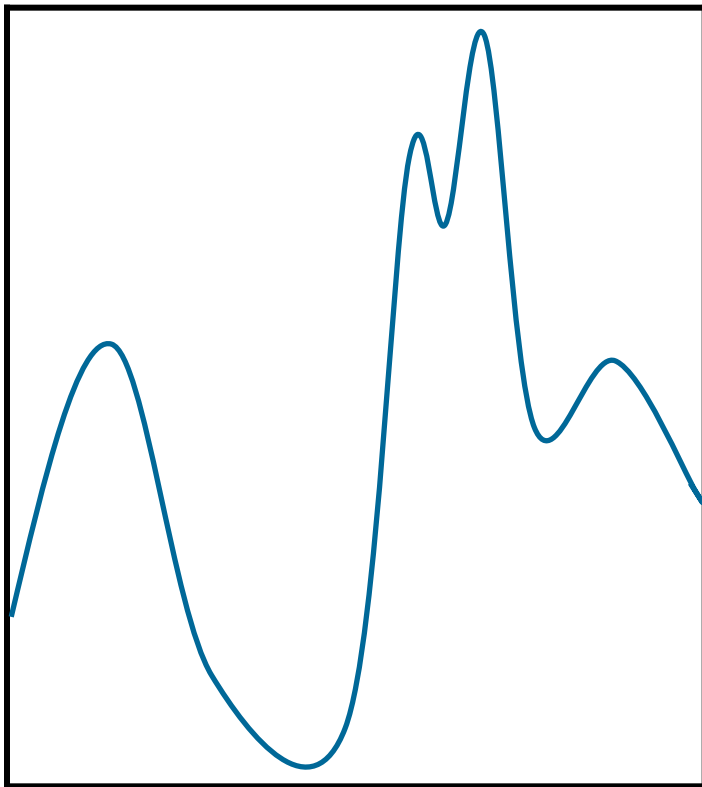


Digital

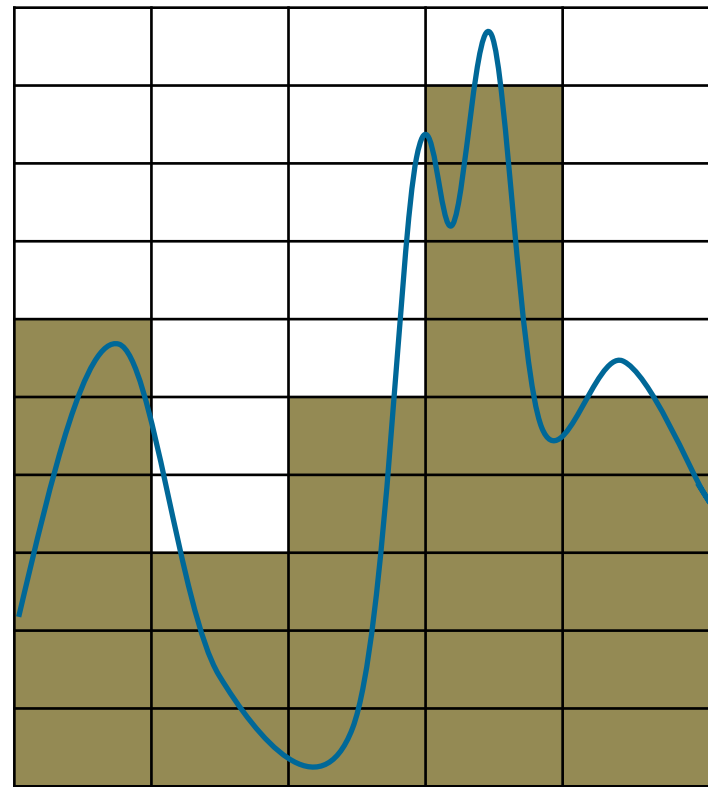


Converting analogue to digital signals (2)

Analogue

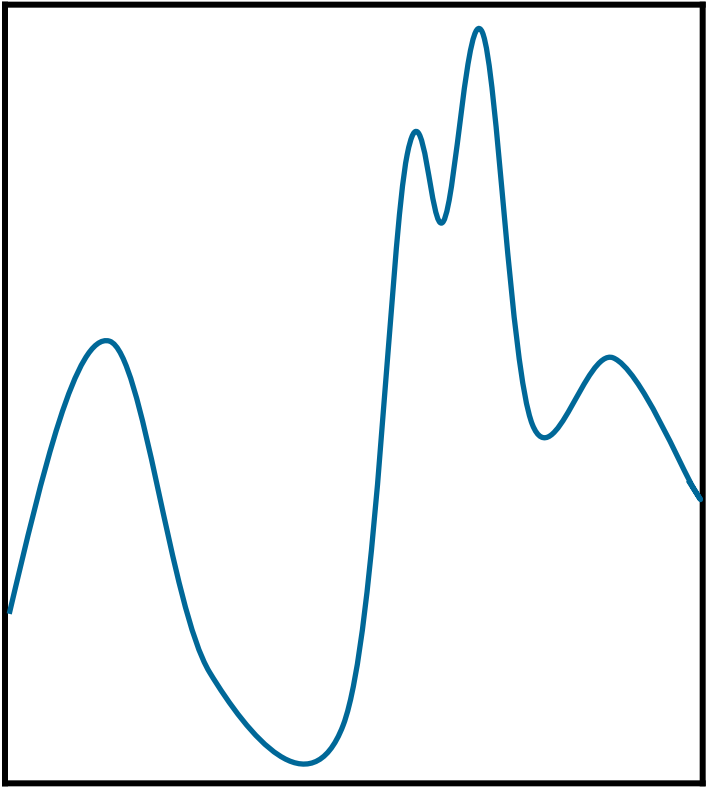


Digital

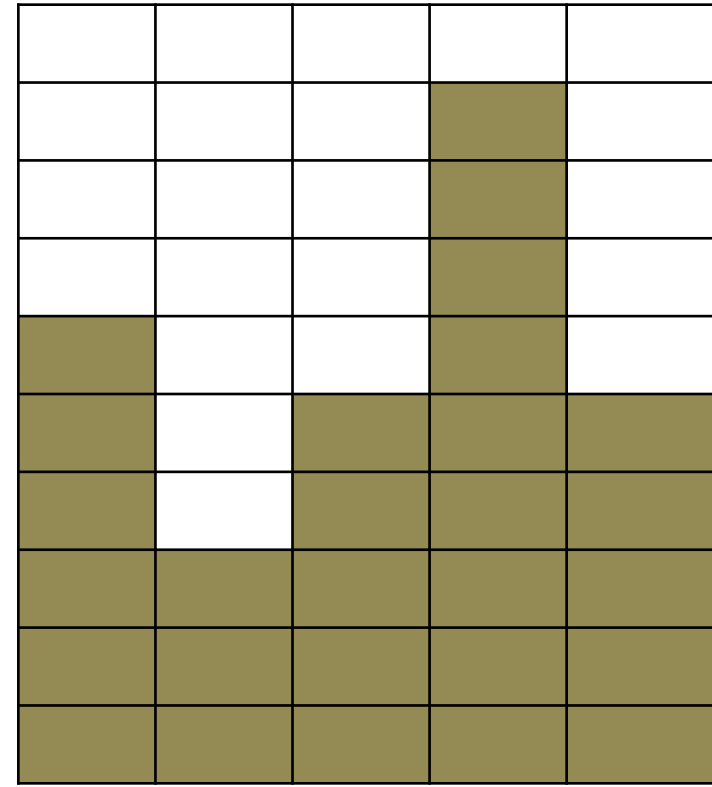


Converting analogue to digital signals (3)

Analogue

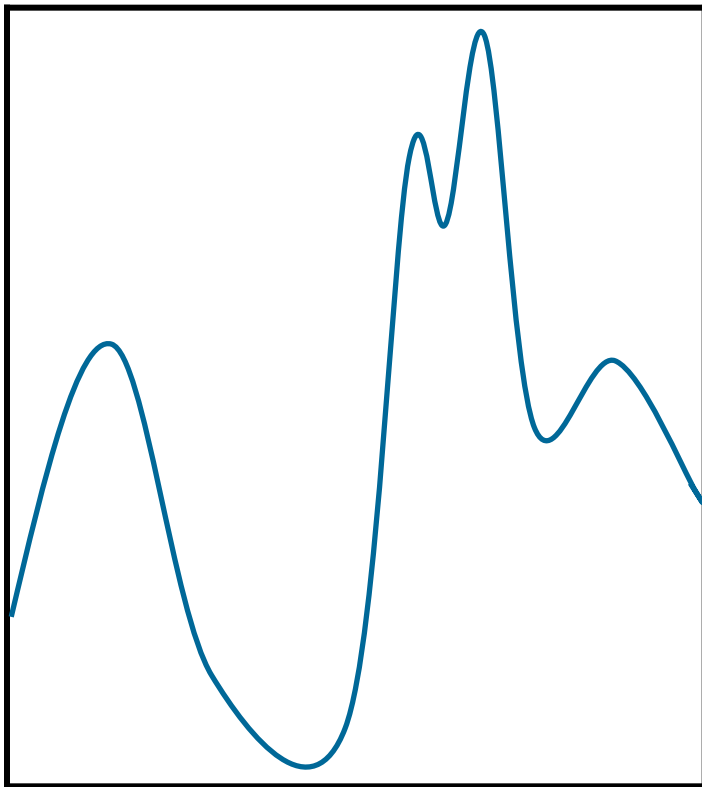


Digital

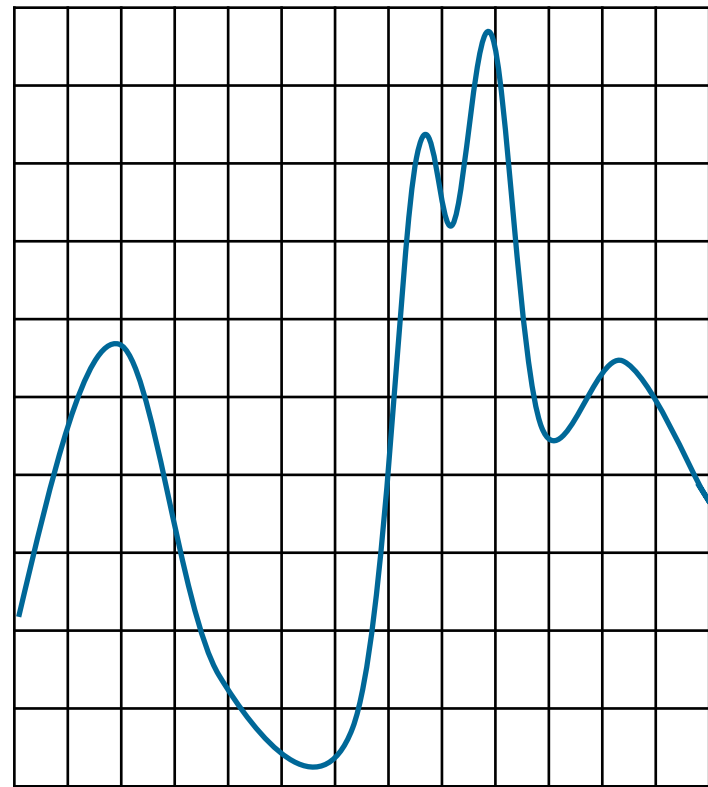


Converting analogue to digital signals (4)

Analogue

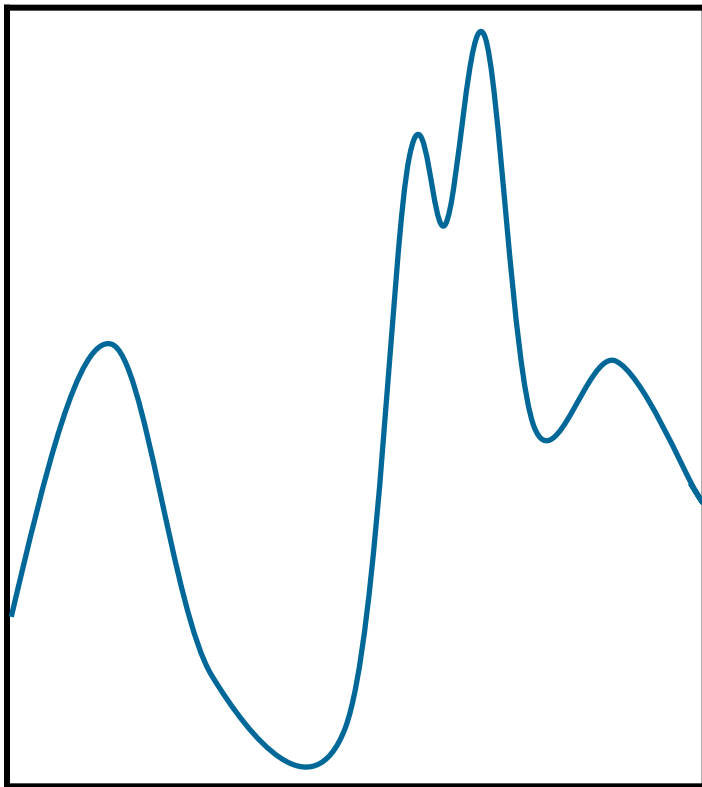


Digital

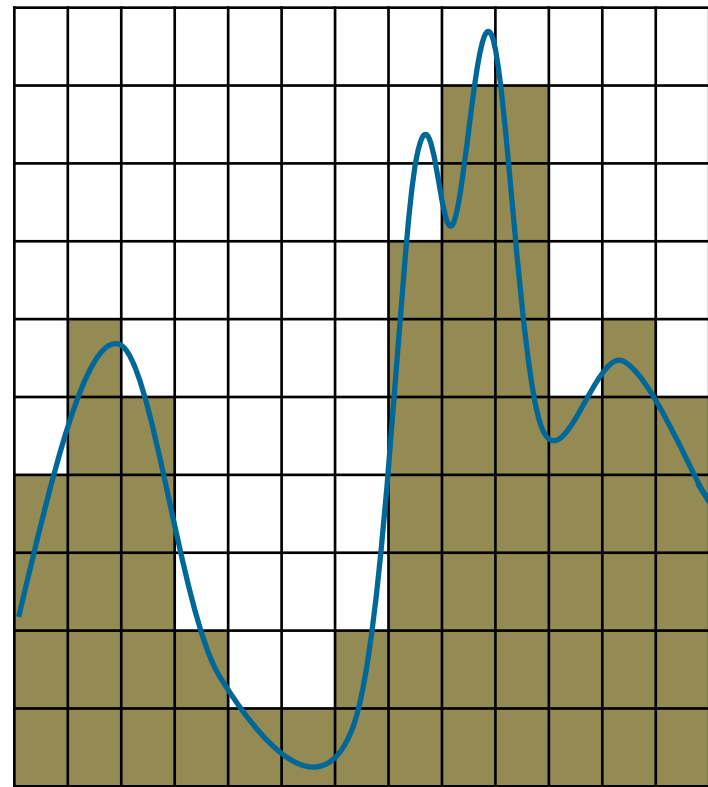


Converting analogue to digital signals (5)

Analogue

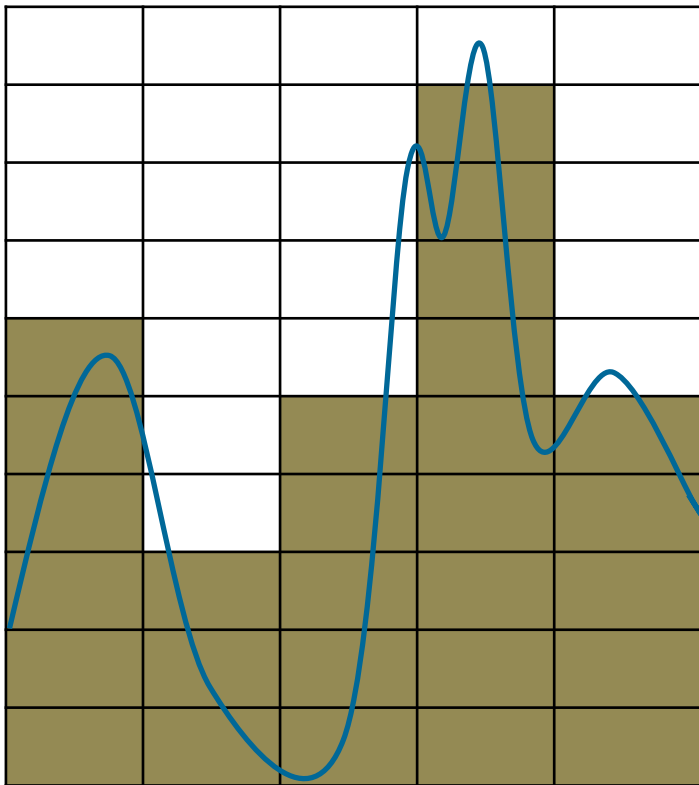


Digital

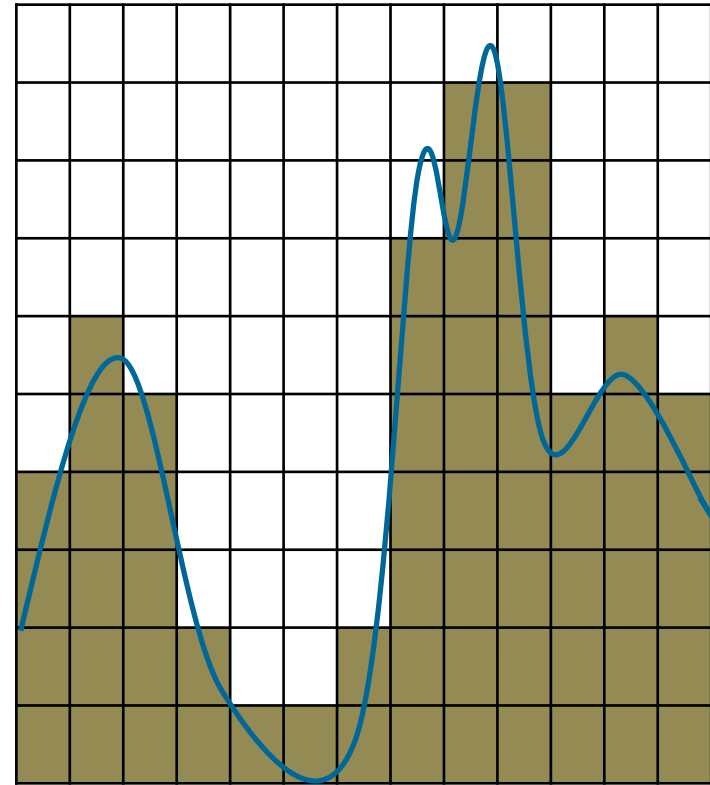


Converting analogue to digital signals (6)

Digital 1



Digital 1

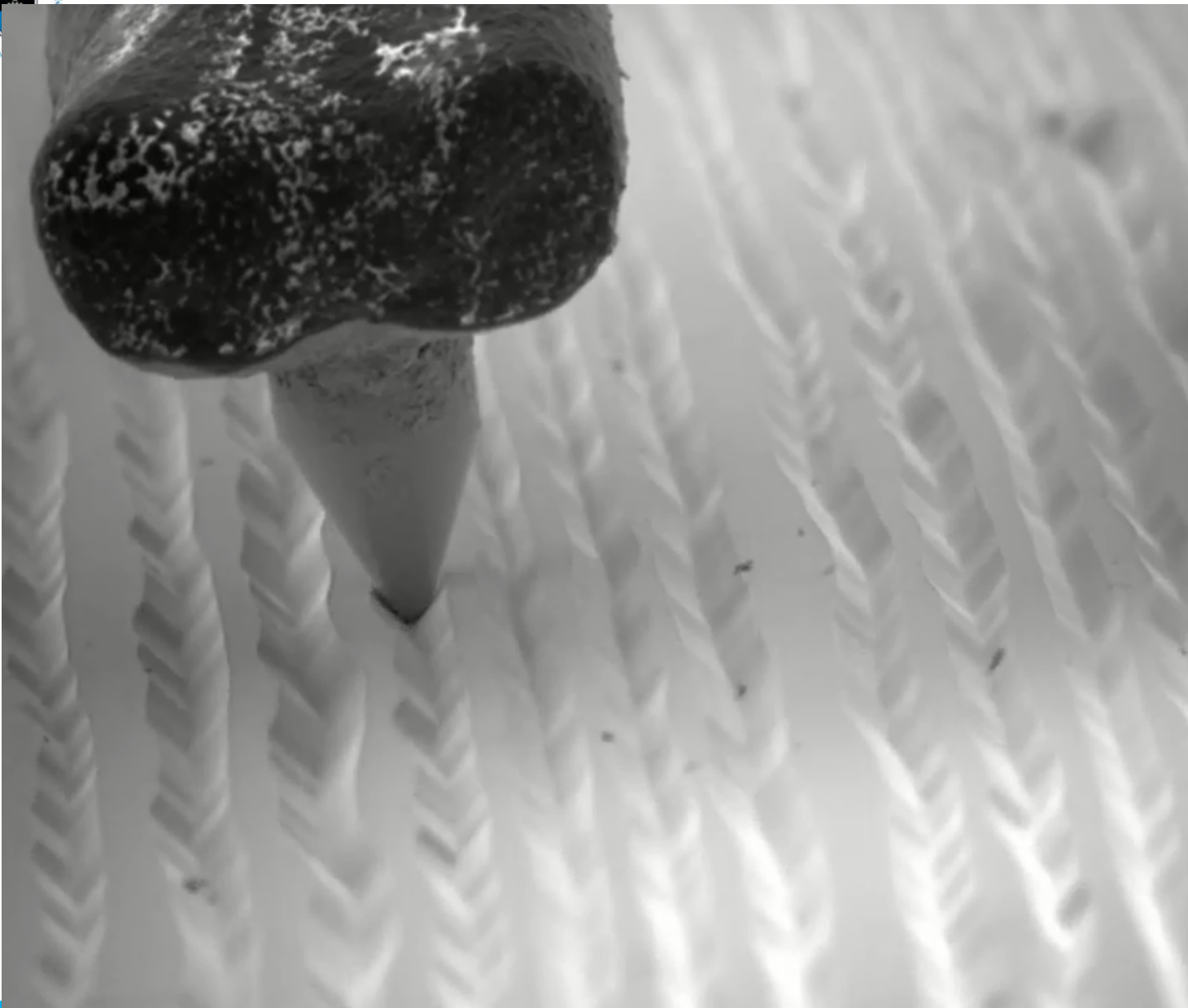


Loss

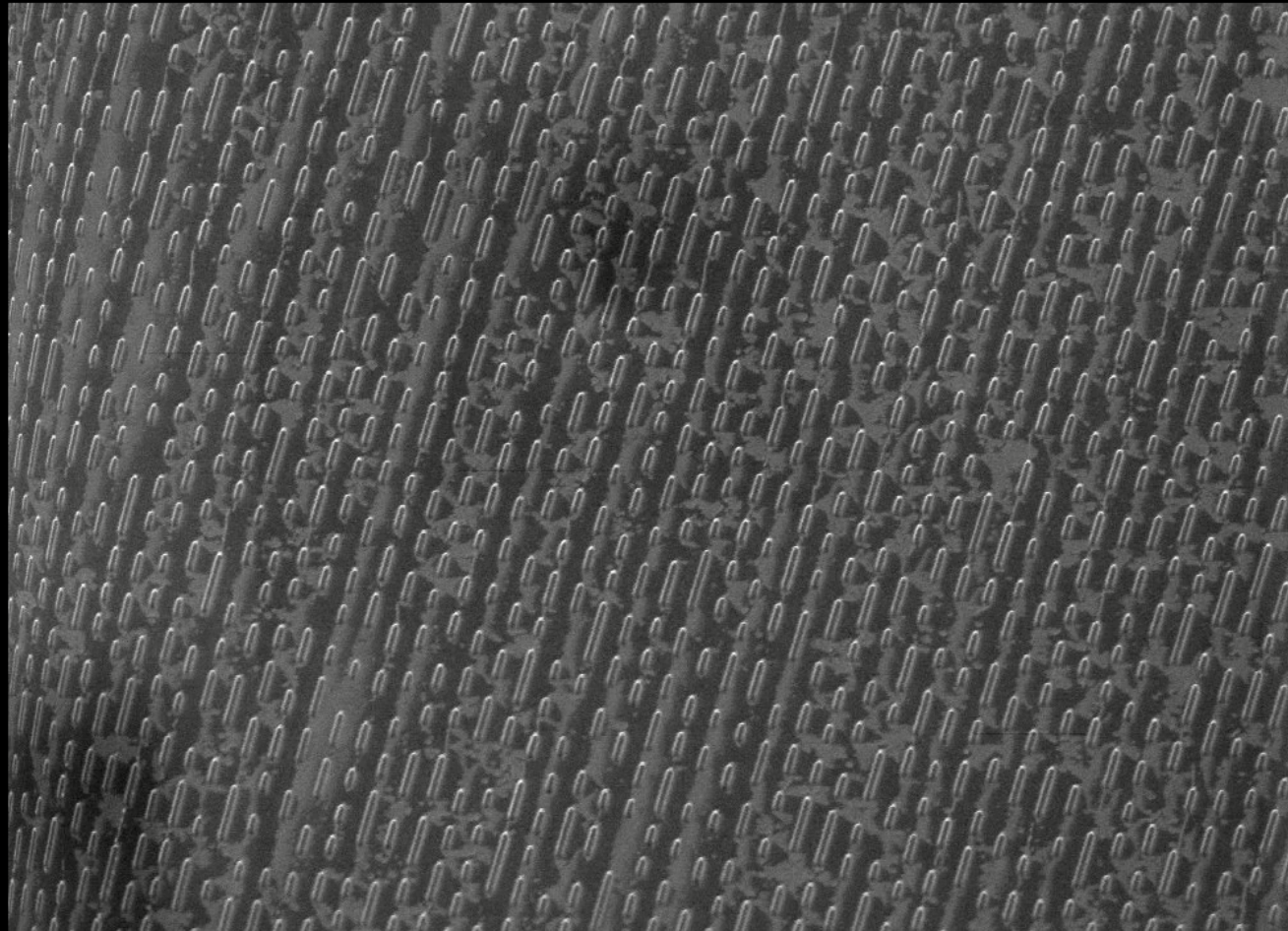
- (We'll come back to this)



© Source: www.dailysabah.com



Phonograph record and stylus under an electron microscope (magnification unknown)



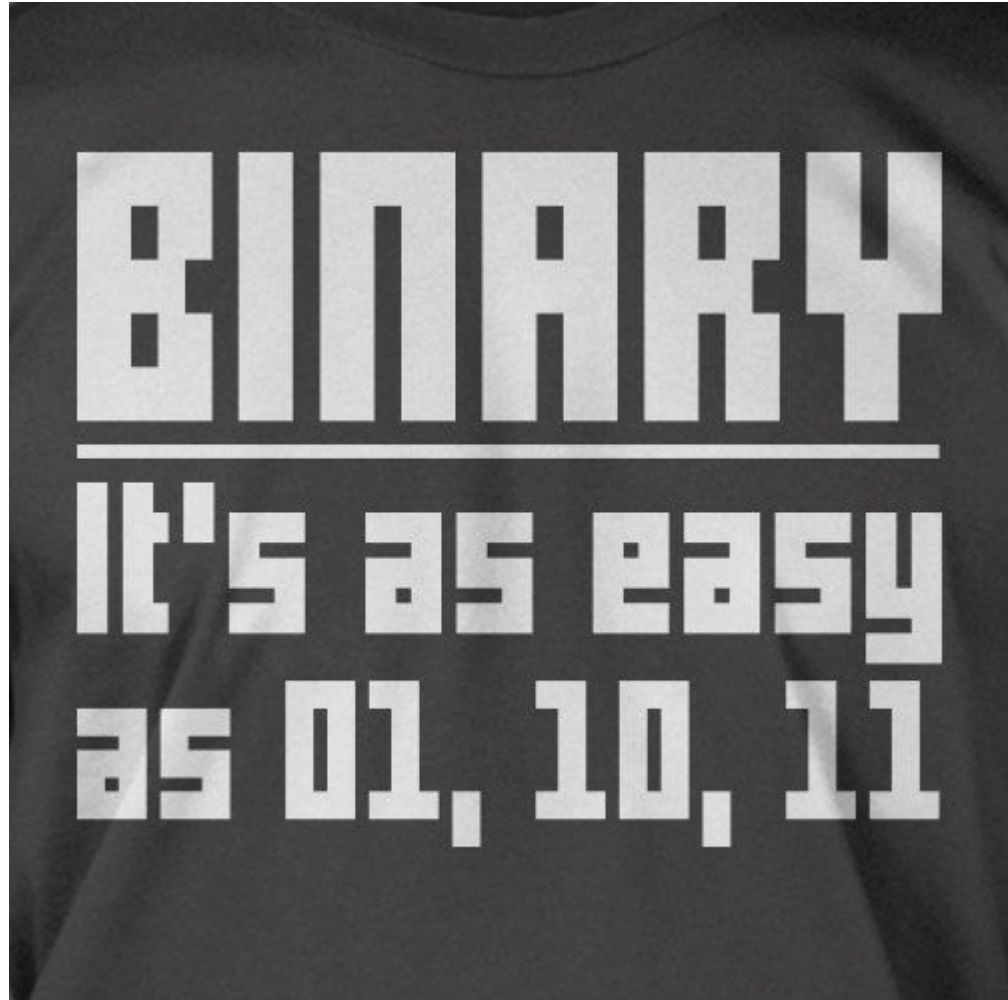
Surface of a CD-ROM under a scanning electron microscope (magnification unknown)



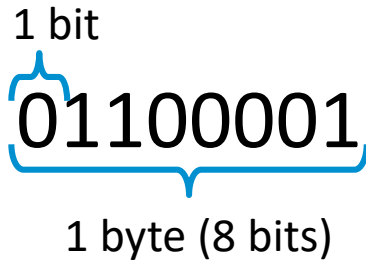
COMPUTERS

How computer encoding works

- Signal?
- Medium?
- Encoding:
 - Binary



Binary code

- The character “a”: 
- Exercise: 01100001 01110010 01100011
01101000 01101001 01110110 01100101
- This is the way computers store all information: text, images, sound, programs....

Computer storage media

- Hard drives
- Solid state/Flash

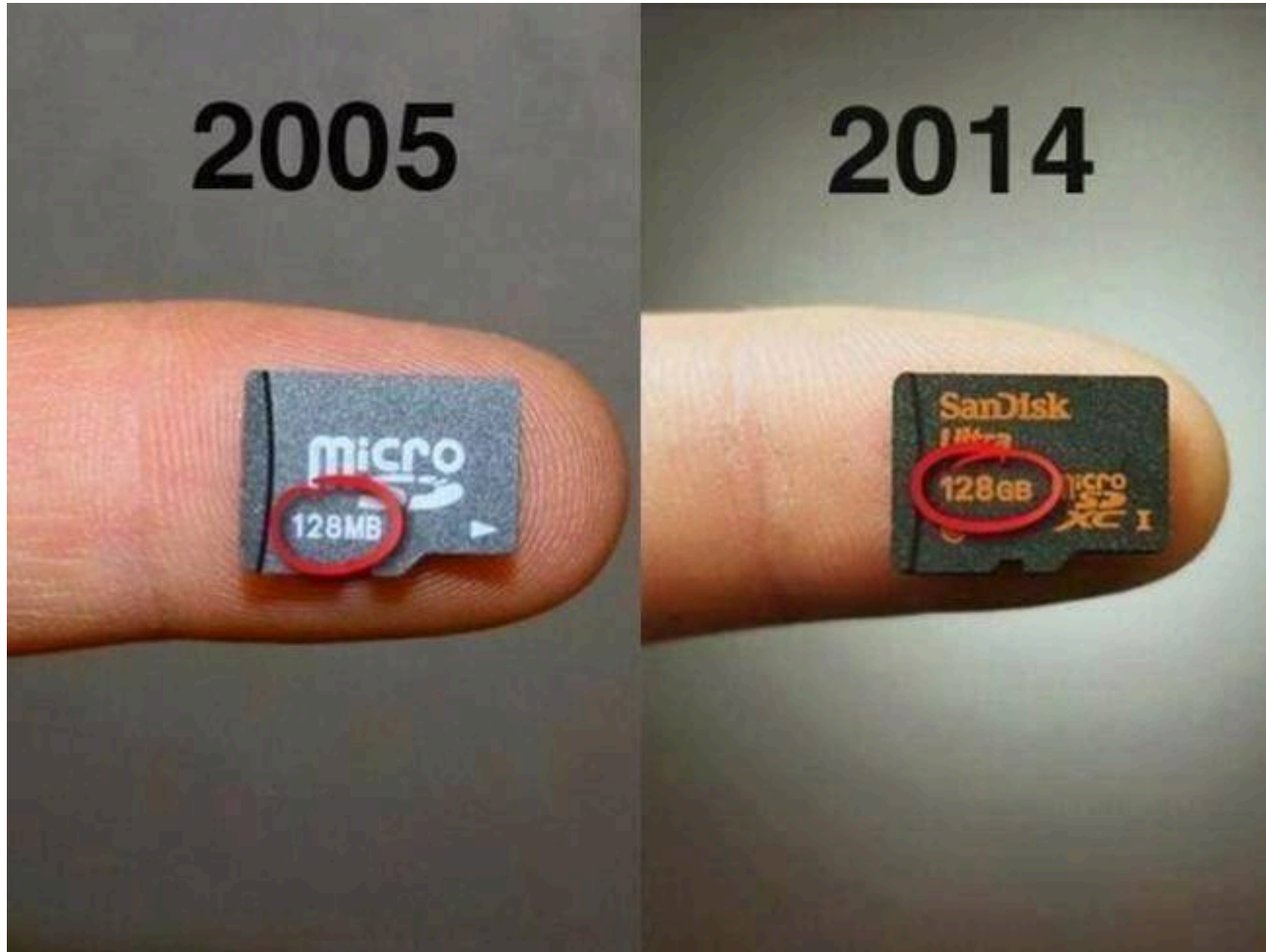
 How many more can you think of?



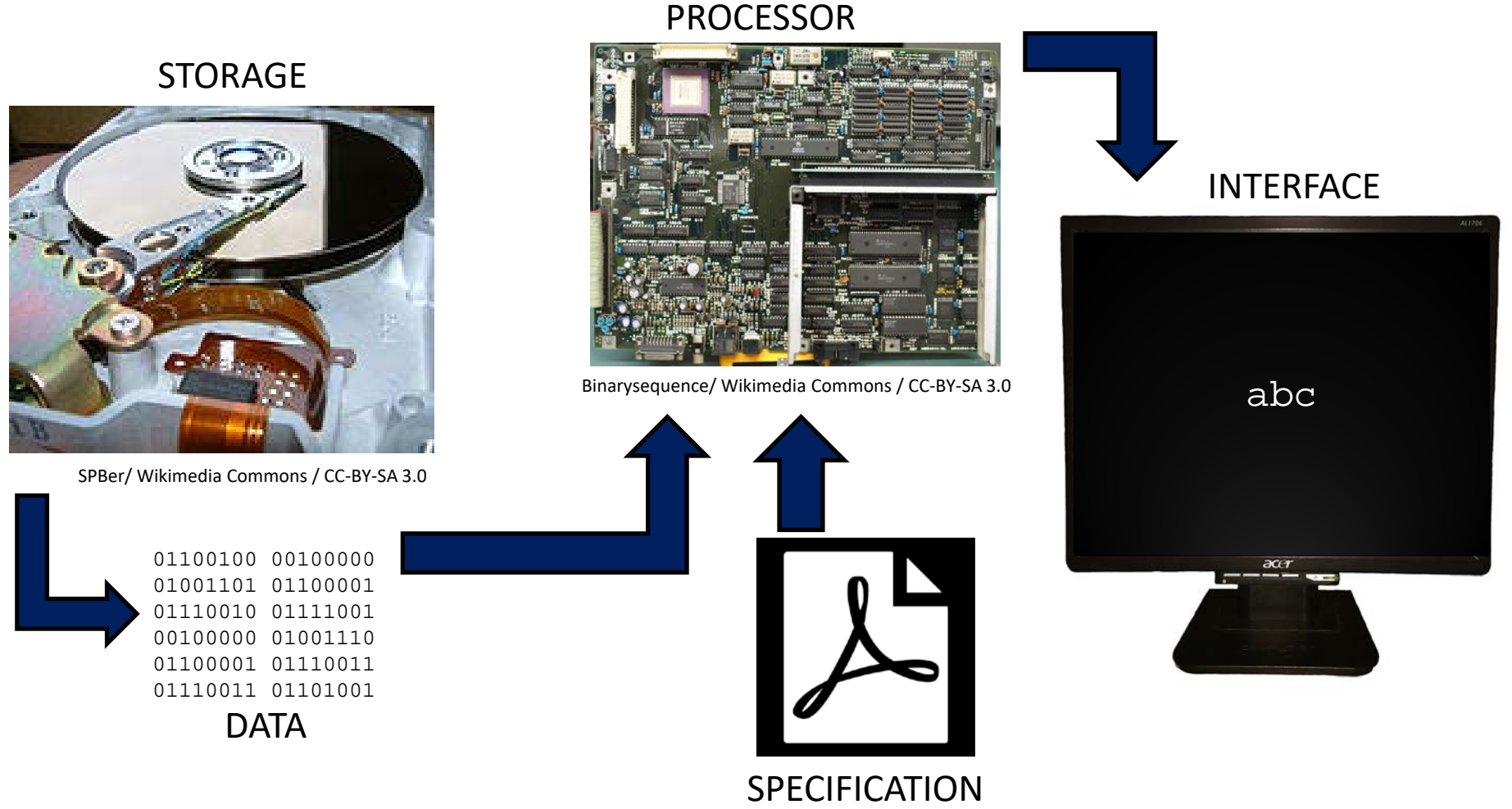
H. Rival / Wikimedia Commons / CC-BY-SA 3.0



Storage



How computer formats work





CURATION

Curation: what does it mean?



Bishop: "I'm afraid you've got a bad egg, Mr Jones";

Curate: "Oh, no, my Lord, I assure you that parts of it are excellent!"

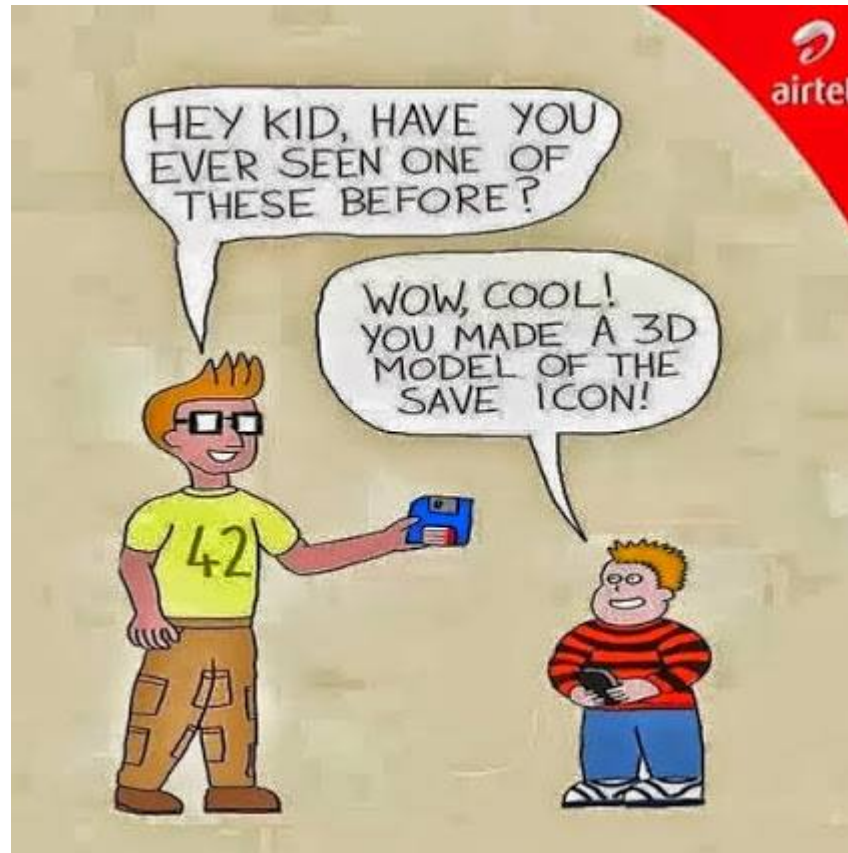
'True Humility' by George du Maurier, originally published in *Punch*, 9 November 1895.



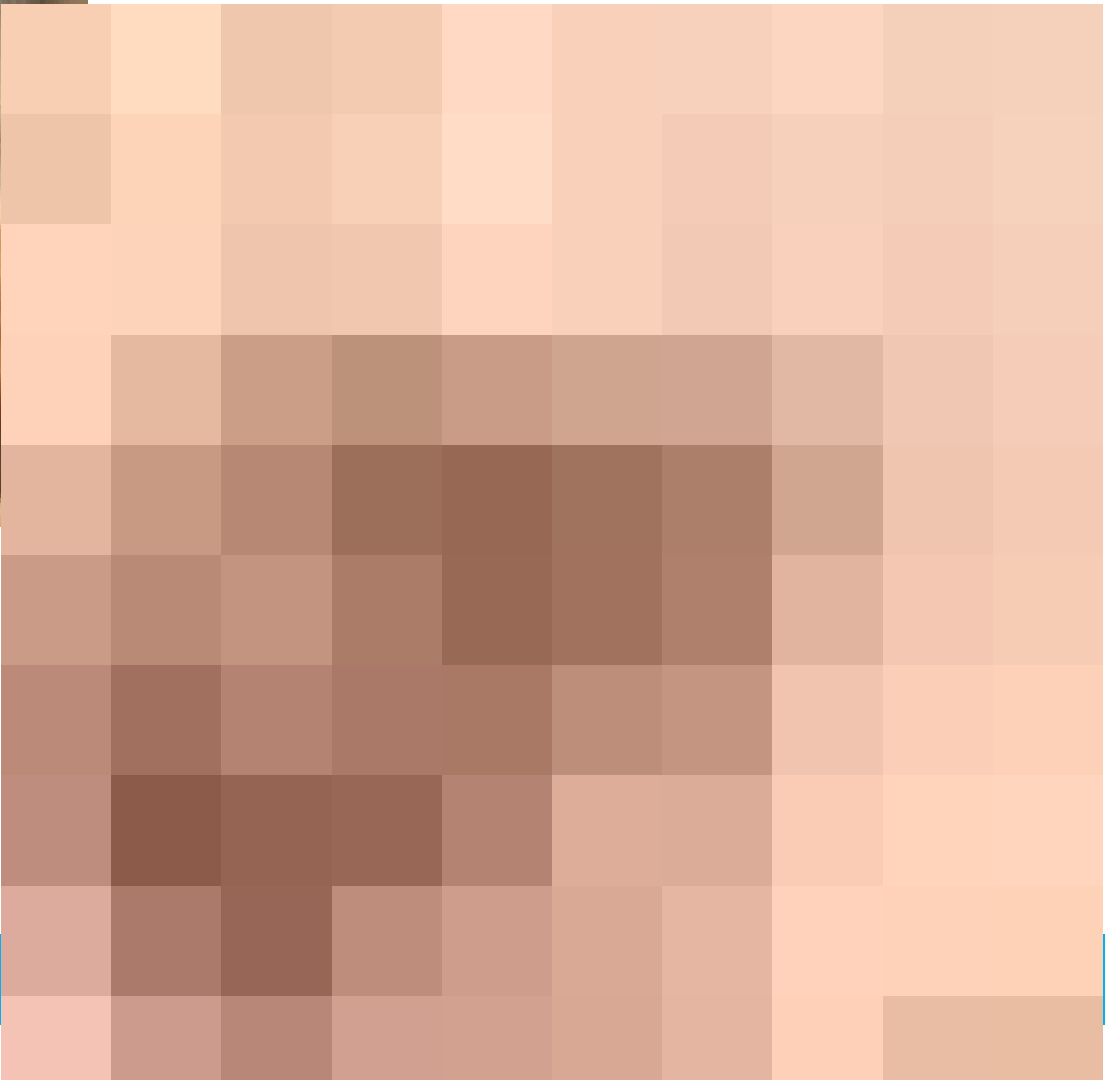
Digitisation and Loss

- Demo: Let's play reality

Obsolescence



Resolution, Quality



THE LIFESPAN OF STORAGE MEDIA

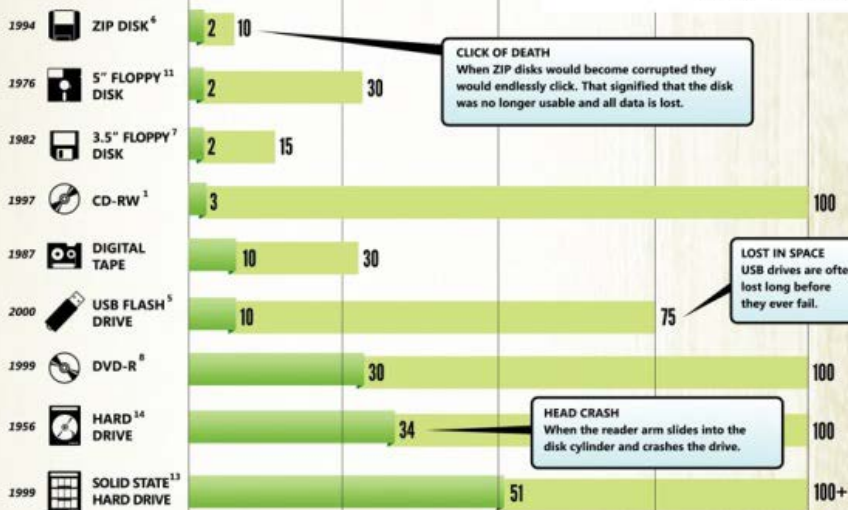
FOR AS LONG WE HAVE BEEN INVENTING WAYS TO CAPTURE AND PRESENT DATA, we have been developing media on which to store it. With each new innovation, we've seen greater convenience, increased storage capacity, and longer lifespans. But at some point, even with the utmost care, media fails... leaving us with a useless piece of plastic, tape, or metal.

The Cloud, with its extensive network of server farms, leverages redundant storage media, ensuring that in the event of hardware failure, the data lives on, stored safely in another part of the cloud. By way of comparison, we've assembled some lifespan statistics for storage media throughout the digital ages.

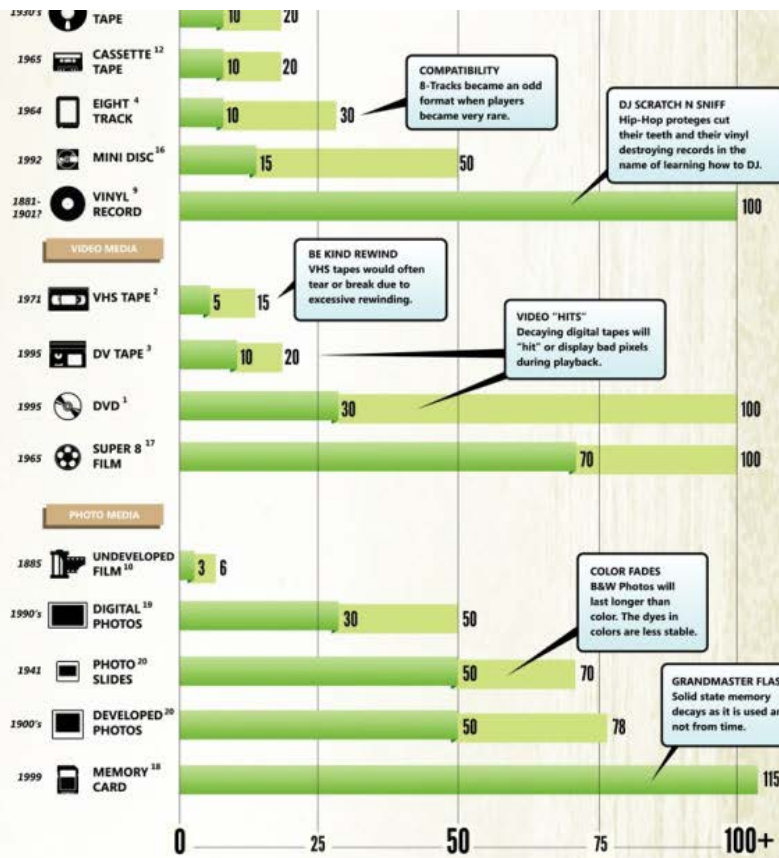
FAST **REUNDANT** **RELIABLE** **ACCESSIBLE ANYWHERE** **FOREVER**



COMPUTER MEDIA



AUDIO MEDIA



ADDITIONAL WAYS MEDIA CAN FAIL



[HTTP://CRASHPLAN.COM/MEDIA Lifespan](http://crashplan.com/medialifespan)

SOURCES:

1 archive.gov, esta.org 2 videotransferboston.com/bell-life-of-video-tape/ www.audiokarma.org 3 http://www.csi.org/foia/reports/pub54906_expectancy.html 4 www.dreasure.com 5 6,7,8,9 webdev.cac.edu/talkin/Storage_Quick_Reference.pdf 10 <http://webdev.cac.edu/talkin/storage.htm> 11 <http://webdev.cac.edu/talkin/storage.htm> 12 <http://www.musicstark.com/articles/can-vinyl-last-forever-eps-encore-vinyl-records> 13 http://engineering.dartmouth.edu/microeng/processimg/_N893000P1.pdf 14 http://www.atarimagazines.com/creative/Vol12/205_Floppy_disk_handling_and_php 15 www.audiokarma.org 16 <http://www.storagegawards.com/standby-the-endurance.html> 17 The Western Digital MTBF specification is 300,000 hours. 18 <http://www.whatisthebest.com/showthread.php?t=1619> Reel-to-Reel Tape Lifespan 19 <http://forums.sony.com/topic/21963-minidis-lifespan/> 20 <http://www.bosonlin.com/super8valty.asp> 21 http://www.whatisthebest.com/equipment/technic/51364/Understanding_memory_cards.html 22 <http://printers.tprnews.com/photos-only/how-to-make-your-photos-last-longer.htm> 23,24 <http://www.caps-project.org/cache/DigitalMediaLifeExpectancyAndCare.html>

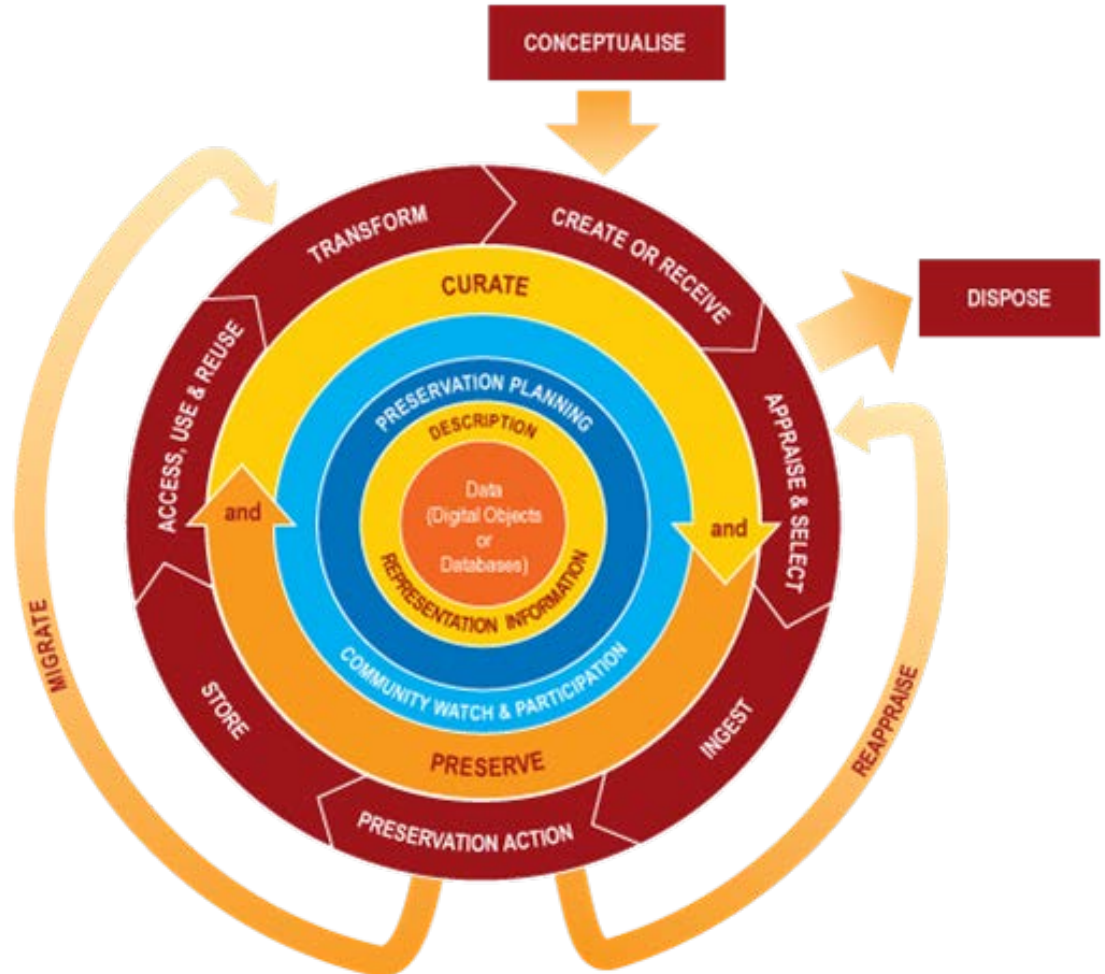


©2012 CODE 42 SOFTWARE, INC., ALL RIGHTS RESERVED



The Circle of Life

- Collection
- Assessment
- Preservation
- Dissemination
- Disposal



Source: Digital Curation Centre UK

Handling noise



Source: patchinpictures.com



Mediating Meaning







THE TEAM: UCT students Kopano Maroga, Nyiko Maroleni, Brian Kamanzi and Nombuso Mathibela are contributing to the local government elections edition of the Cape Argus.

PICTURE: DAVID RITCHIE



THE TEAM: UCT students Kopano Maroga, Nyiko Maroleni, Brian Kamanzi and Nombuso Mathibela are contributing to the local government elections edition of the Cape Argus. PICTURE: DAVID RITCHIE

Drops in the ocean...

“With the ivory tower of UCT as our present reluctant home we sit at the intersection of many identities and as a result we relate to the self, community, city and state in different ways. These voices rep-



THE TEAM: UCT students Kopano Maroga, Nyiko Maroleni, Brian Kamanzi and Nombuso Mathibela are contributing to the local government elections edition of the Cape Argus. PICTURE: DAVID RITCHIE

Drops in the ocean...

“With the ivory tower of UCT as our present reluctant home we sit at the intersection of many identities and as a result we relate to the self, community, city and state in different ways. These voices represent drops in the ocean of popular sentiment that fill the spaces



Helen Zille ✓

@helenzille

[Follow](#)

If this woke bunch hate being UCT students so much, pls help them out of their misery and withdraw their funding.

9:05 PM - 3 Aug 2016

↩️ ↻️ 452 ❤️ 318

Storify MG_Reporter Helen Zille faces wrath of Twitter after suggesting UCT students have funding cut a day ago

Former Democratic Alliance leader and professional pot stirrer Helen Zille has found herself in hot Twitter water once again. This after she tweeted a picture of a recent Cape Argus and insisted someone do the student authors a favour and "withdraw their funding".



THE TEAM: UCT students Kopano Maroga, Nyiko Maroleni, Brian Kamanzi and Nombuso Mathibela are contributing to the local government elections edition of the Cape Argus. PICTURE: DAVID RITCHE

Drops in the ocean...

"With the ivory tower of UCT as our present reluctant home we sit at the intersection of many identities and as a result we relate to the self, community, city and state in different ways. These voices represent drops in the ocean of popular sentiment that fill the spaces

 **Helen Zille** 
@helenzille

 Follow

If this woke bunch hate being UCT students so much, pls help them out of their misery and withdraw their funding.

9:05 PM - 3 Aug 2016

  452  318

Click here
to find out how



How Eldorado Park's dreams and
twitter.com/helenzille/status/760914391961305088/photo/1

The authors, who apparently are involved in the #FeesMustFall movement, wrote about the political environment in South Africa and whether or not they should cast their ballots. It would seem it was their referral to UCT as their reluctant home



METADATA

Metadata

- “Data about data” what does that mean?
- For librarians: a complete bibliographic record
- For information technologists: database schema or definitions of data elements
- For archivists?

Main categories of Metadata

Descriptive

Resource description for:

- Discoverability (information retrieval)
- Identification

Structural

Physical & logical structure for

- Resources with various parts
- Compound objects

Administrative

- Management – acquisition, appraisal
- Technical - records the technical infrastructure used to create/access data
- Use/rights – access rights & usage
- Digital provenance/preservation



← Back to search



Bernt Rostad + Follow

Bibliotheca Alexandrina

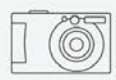
The Bibliotheca Alexandrina, the new Library of Alexandria, was inaugurated in 2003 near the site of the old library. It's located on the shore of the Mediterranean Sea, in the heart of Alexandria. The library's architecture is striking. The main reading room stands beneath a 32-meter-high glass-panelled roof, tilted out toward the sea like a sundial, and measuring some 160 m in diameter.

The Old Library of Alexandria was once the largest library in the world. It is generally thought to have been founded at the beginning of the 3rd century BC, during the reign of Ptolemy II of Egypt, after his father had built what would become the first part of the library complex, the temple of the Muses—the Musesion (hence we get museum). It was destroyed in antiquity though historians disagree on the cause: Was it Caesar when he set fire to the Egyptian fleet in 48BC, was it the attack of Aurelian in the 3rd century AD or perhaps the decree of Theodosius in 391 AD? (Source: [Wikipedia](#))

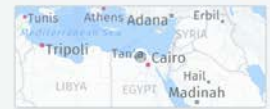
1,572 views 2 faves 0 comments

Taken on March 31, 2006

All rights reserved



Canon Digital IXUS 500



Bab Sharq District, Alexandria, Alexandria

- f/2.8
- 1/320
- Flash (off) (did not fire)
- 7.4 mm
- ISO 60
- Show EXIF

This photo is in 1 album



DSE Expedition 2006 - Egypt [1] 131 items

Tags

- travel
- egypt
- alexandria
- library of alexandria
- bibliotheca alexandrina
- outdoor

Additional info

- Viewing privacy
- Safety level

Public Safe

SOURCE:

<https://www.flickr.com/photos/brostad/232588053/>



Exercise

What metadata do you think is required for archives?

Metadata standards

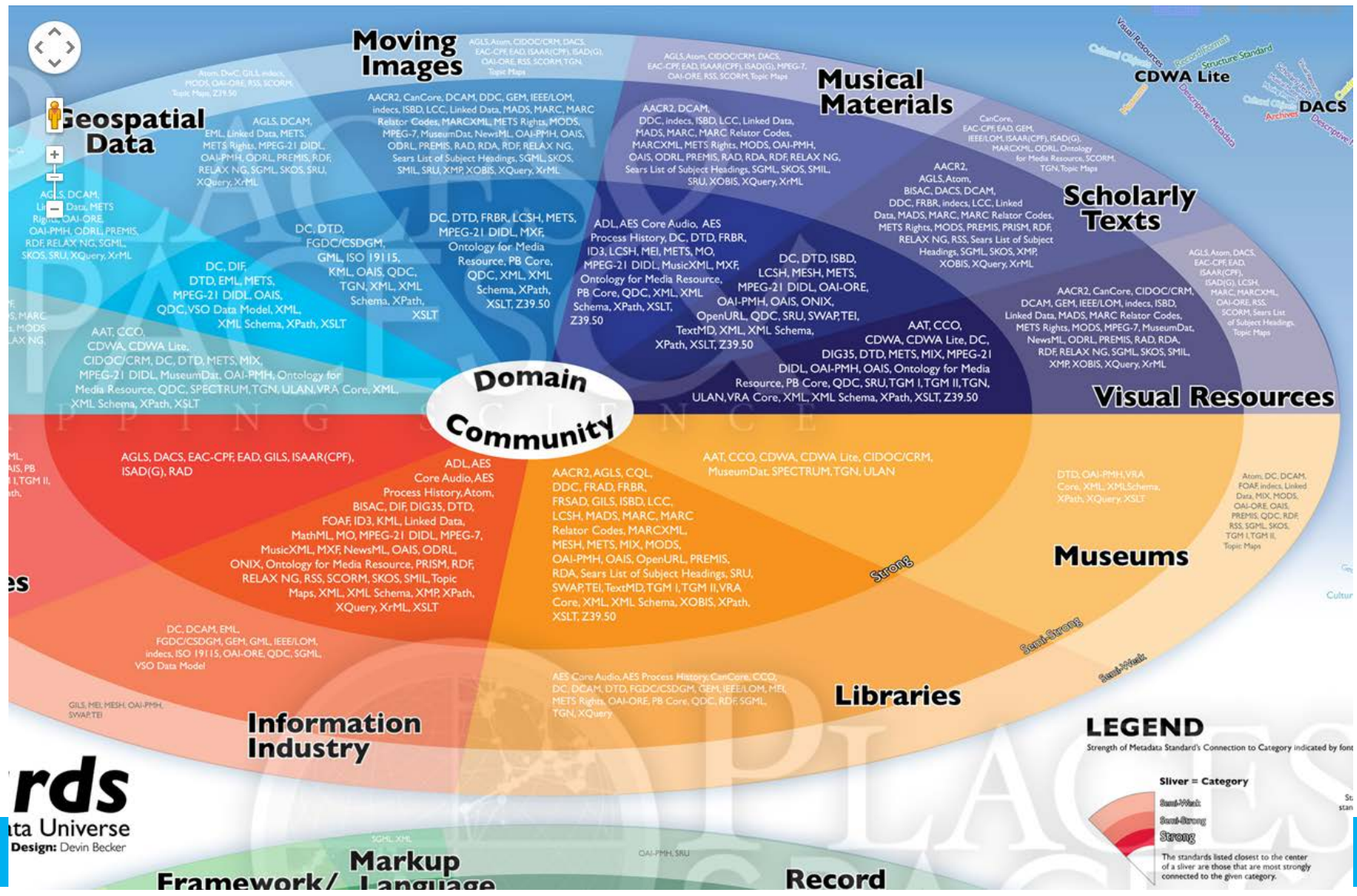
- RAD
- MARC-AMC...
- Encoded Archival Description
- Dublin Core
- Domain standards
- Controlled vocabularies

Metadata standard / metadata schema?

For more:

<http://www.dcc.ac.uk/sites/default/files/documents/resource/curation-manual/chapters/archival-metadata/archival-metadata.pdf>

Metadata standards





...AND (NEARLY) EVERYTHING ELSE

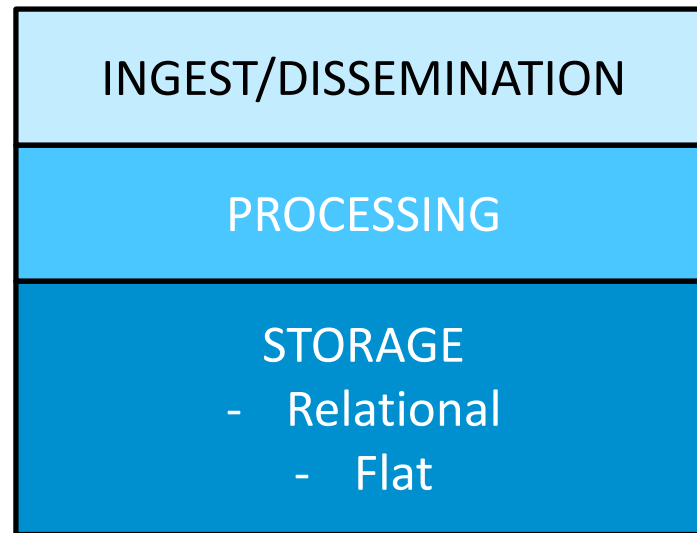


Intellectual Property

- The myth of the “public domain”

Websites and Repositories

- Repositories
 - The 3-tier model



- Trusted Digital Repositories



Optical Character Recognition (OCR)

- Fonts, characters
- Columns, page numbers, footnotes
- Noise
- The human factor: checking is essential

Derivatives and Surrogates

- What is a difference?



Igor Novikov, 'Miss Switzerland', 2010

On the nature of a surrogate



Born Digital vs Digitised

- Do/should we treat them differently?
- Maintaining the relationship between original and derivative artefacts

Archive and Access Copies

- Archive: for preservation
 - Standards-compliant
 - Backup, migration, checksums
- Access:
 - Lower resolution: faster downloads, easier to send
 - Protect intellectual property

Governance and policy

- Access
- Metadata
- Supporting technology
- Retention / disposal
- Digitisation procedures
- Preservation / migration



Principles: *Respect des fonds*

- Provenance
 - In the digital world?
- Original order
 - What does this mean? How can you tell?



Collections in dialogue

- Interoperability
- Content / metadata harvesting
- Portals
- Overlays (e.g. GIS)

Approaches

- Back-scanning / Scan-on-demand
- OCR
- Outsource / Insource?
- Sustainability and reliability
- Retain relationship between originals & derivatives

REMEMBER: Digitisation is a *capability*,
not a *project*



Questions?

Richard Higgs
richard.higgs@uct.ac.za
Twitter: @higgsri