Digital Libraries and the quest for information curation

UFP's Erasmus Staff Week for Librarians

Workshop on 8th April, 2014

Luis Borges Gouveia

lmbg@ufp.edu.pt

The Workshop

- Title:
  Digital Libraries and the quest for information curation

  The growing number of computers and networks place new challenges to every sector of our lives. Digital information is omnipresent in an everyday basis in any kind of activity that both individuals and organizations perform.

  As a result the needs and services provided by libraries needed to consider such a digital context with a smart move to adapt and evolve in order to provide digital information along with more traditional services.

  This presentation introduces some of the concepts related with such digital opportunities when considered along with digital libraries, and discusses some innovative uses to organize, present and curate information and knowledge, within a library context.

Workshop conducted by

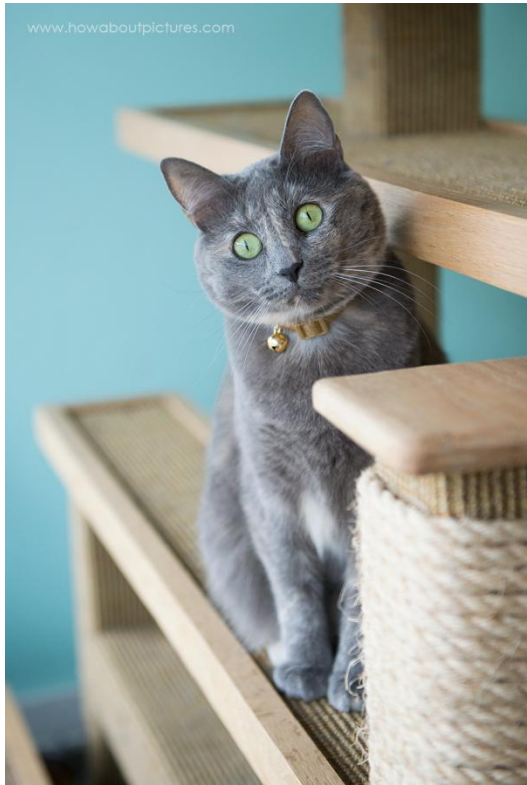
- Luis Borges Gouveia is an *Associate Professor* of Knowledge Management, Science and Technology Faculty at *University Fernando Pessoa*
- Among his research interests are *the impact of digital information and information society issues in our day-to-day life*.
- He publishes 12 books and was involved in several projects concerning the use of information in e-learning and e-government settings
  - More information is available at *http://about.me/lbgouveia*

Schedule, Tuesday 8th April

From 14:30 till 16:00:

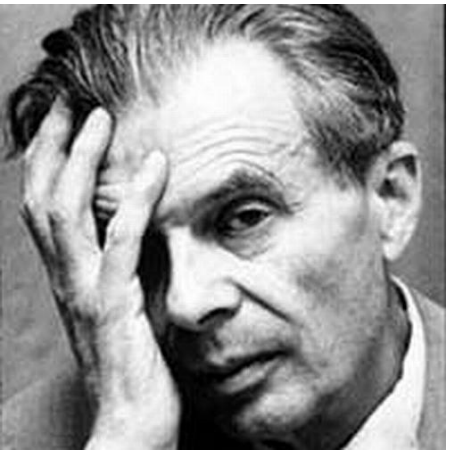
- The digital challenges
- Digital libraries
- The quest for digital curation

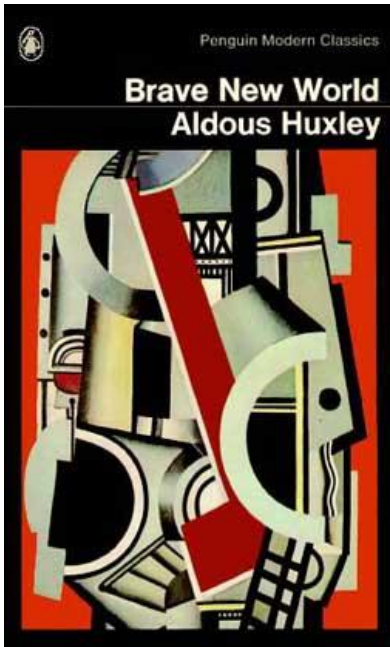
- *Be free to interact whanever wanted*
- *The main idea is to discuss*



Internet as *a brave new world*

- As stated Aldous Huxley (1894-1963), *the defence of a Paradise-enginnering world setting can not be a viable substitution for people and its emotional links* (he is in favour of universal happiness, not a perfect and pure true)
  - See more at [*http://www.huxley.net/*](http://www.huxley.net/)





The easy part…

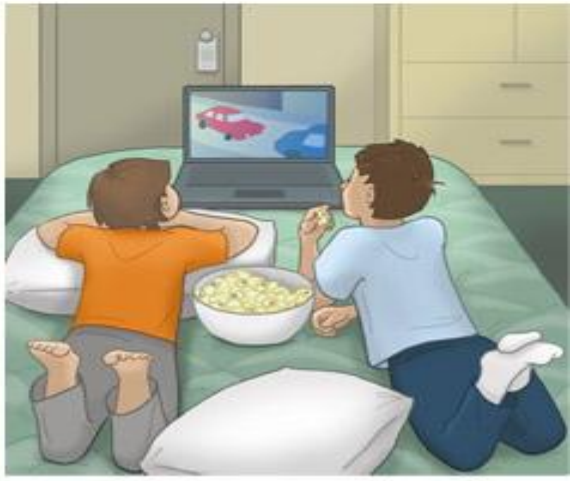
Taking notes in a class

Reading cached e-mail during a flight

Updating patient records between appointments

Browsing the web while reclining on a sofa

Giving a business presentation



Watching a movie in a hotel room

Some ideas of such a world…

- Highly networked
- Fast paced
- Ongoing change
- Mutating workplace (leisure also?...)
- Do it now, everywhere, with available tech, no time waste, and resources efficiency
- Action
  - Collaboration oriented
  - Life long learning
  - Self learning
- Be prepared for
  - share, cocreate, be creative, reuse, and stay with high mobility

Digital challenges

A definition…

- Describes any system based on discontinuous data or events.
  - Computers are digital machines because at their most basic level they can distinguish between just two values, 0 and 1, or off and on. There is no simple way to represent all the values in between, such as 0.25.
  - All data that a computer processes must be encoded digitally, as a series of zeroes and ones.
- The opposite of digital is *analog*.

  - *From http://www.webopedia.com/TERM/D/digital.html*

Spread the word (on that case, the bit)

- Bit: binary digit
- First computer, then networks
- First some data, then many data
  - On such days almost all the data that matters
- First some information, then those information we need to access
  - On such days almost all the information with value
- And knowledge?
  - We are working on it…

T. S. Eliot (1888-1965): **The Rock** (1934)

Where is the Life we have lost in living?

Where is the wisdom we have lost in knowledge?

Where is the knowledge we have lost in information?

http://www.wisdomportal.com/Technology/TSEliot-TheRock.html

Information society

- *a society where information and communication technology are the primary resource to exchange digital information, and to support interaction between individuals using practices and methods in permanent change*

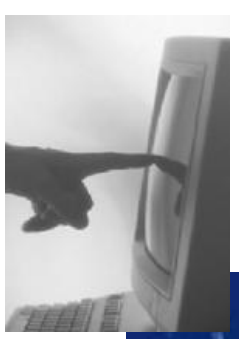
(Gouveia and Gaio, 2004)

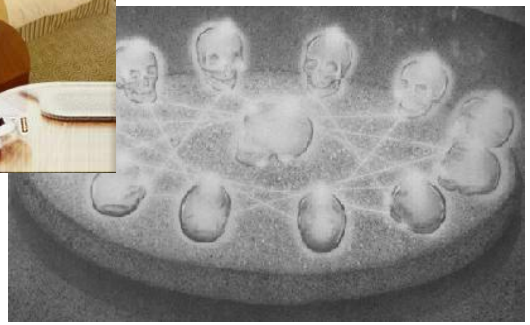
Information society

Heavy use of ICT (computers and networks)

Growing use of digital

Network organisation

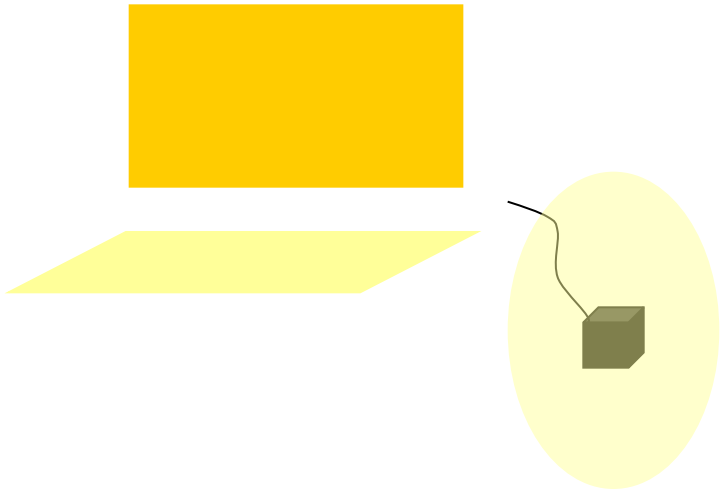




Once upon a time...

the computer

The computer without a mouse!



The computer without a mouse! ... How to use it?

The computer without a keyboard!

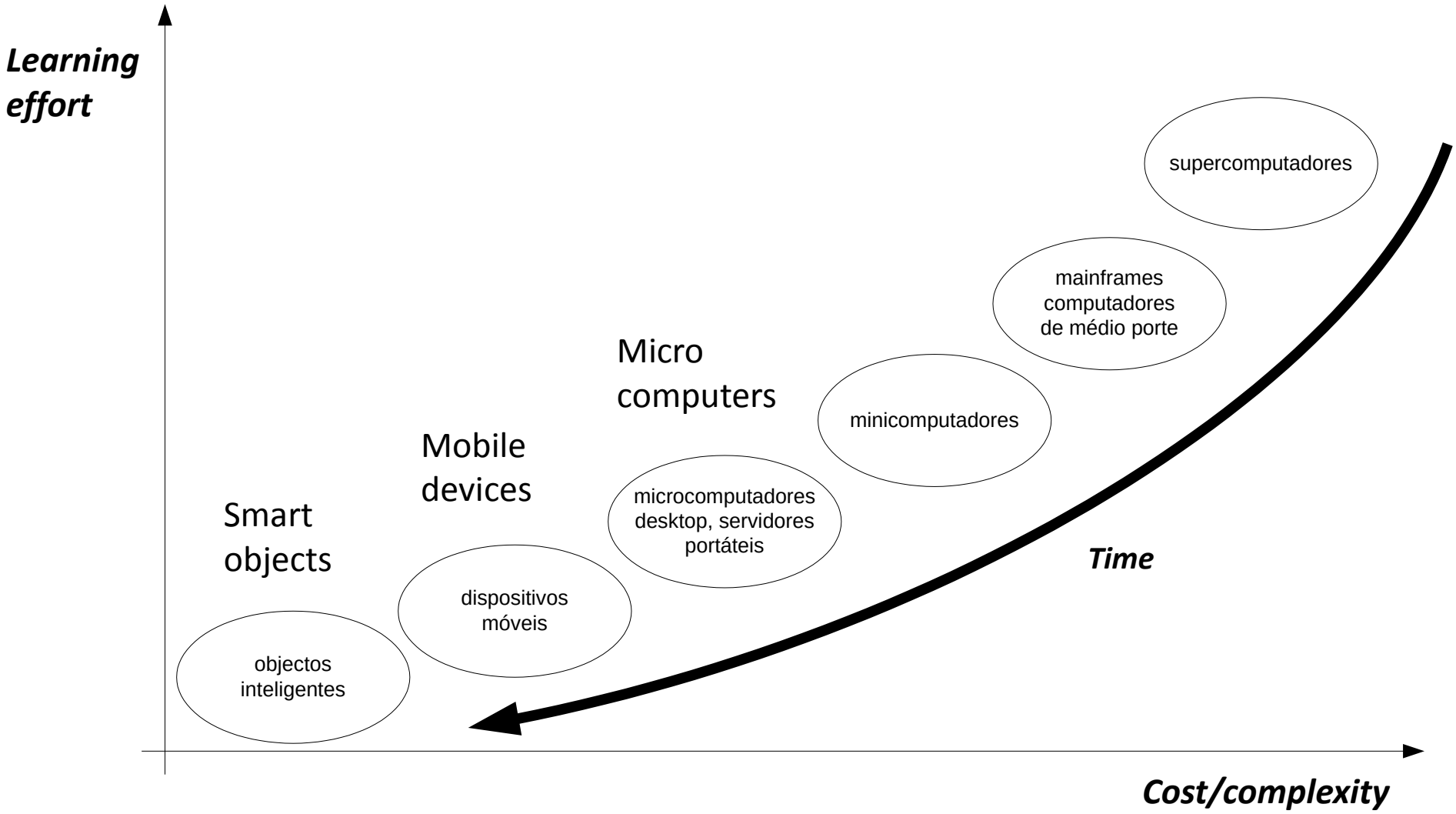
The computer without keyboard!... How to use it?

The computer without monitor!

But... Where is the computer!

But... Where is the computer!

But... Where is the computer!



But... Where is the computer!

- Networked hidden

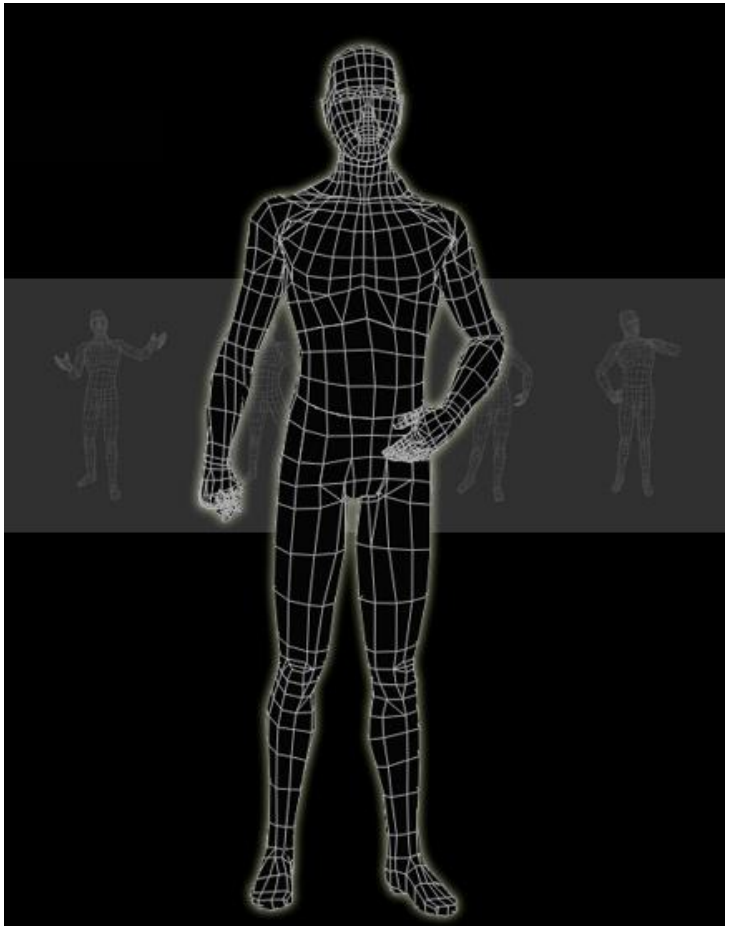
and

- Transaction related

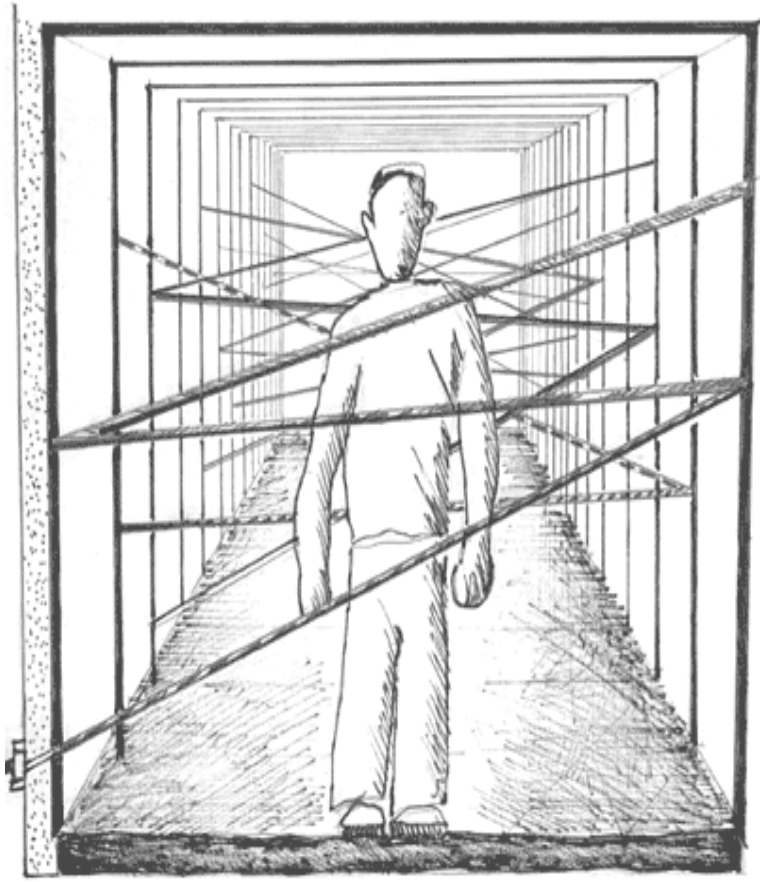
“somewhere among us”

Make us remember some religious stuff

the digital and rich information environments



The analogic and rich information environment



From the analogic to the digital “world”

- **learn**...
  - In the analogic, memorise to learn
  - In digital, forget to learn
- **work**...
  - In the analog, take time to work
  - In the digital, work without taking time

www.go2bed.com

Some remarks

- With digital support
  - Space is extended
  - Time becomes a scarse resource
  - i.e. more reach, less time
- People in organisations
  - Less government
  - More governance
  - i.e. Distance can be less a problem than time (time to know, time to be, time to do, time to react,...)
  - i.e. Networks, instead of leadership (?)

Some remarks

- Again, the digital, time-space and people...
  - i.e. Increase information flux and interactions
- At the end of the day
  - Who pays the bill
  - Who controls it
- Has we take this into consideration when design our cities* for (?)
  - Diversity
  - Proximity
  - Centrality
  - Knowledge
  - A given strategy

*schools, libraries, …

A number of challenges

- People skills
- [Social, economical, digital] gap
- Engage people
- What to know and how (wisdon?...)

But also

- Sense human presence
- Collaboration support (not just sharing or cooperation)
- Information visualisation & Knowledge representation
- Linking the real and the virtual (location devices, senses and get rid of web based as first tech proposal...)

Additional notes

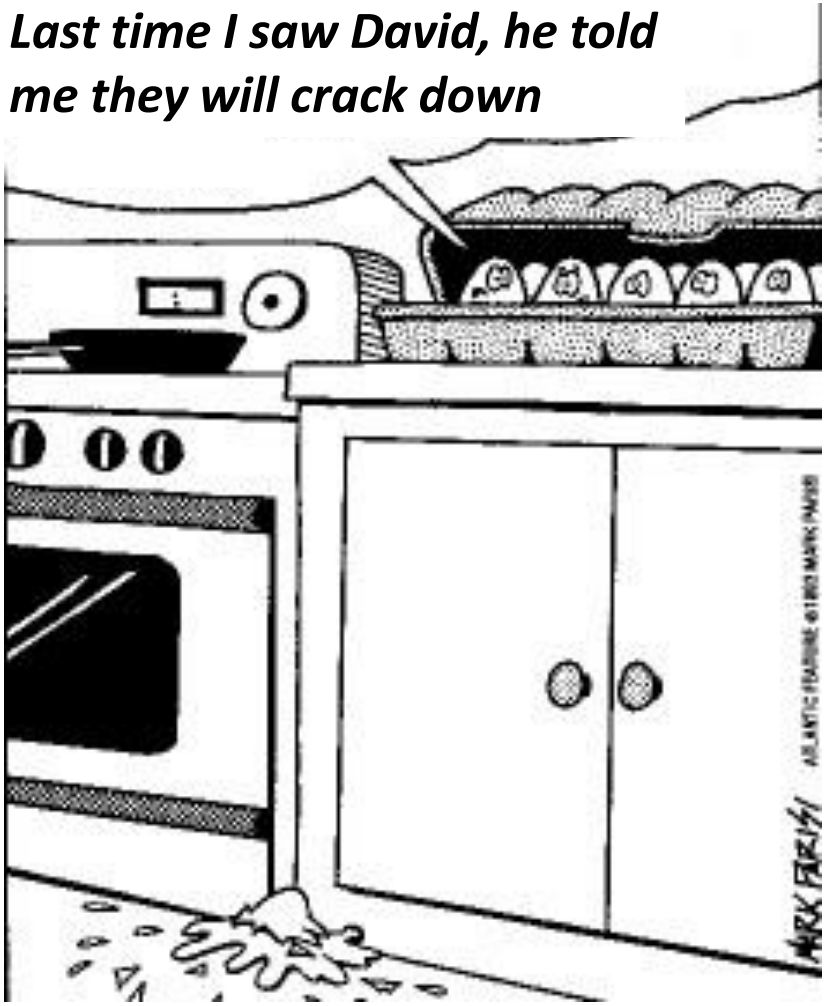
- “perform” digital it is not just related with computers and networks
- We still be analogic, but our interaction are more digital than ever

Additional notes...

- Reinvent our notions of time and space
- Reinvent the function
- Deal with information challenges

Do it networked!

Last time I saw David, he told me they will crack down



Additional notes

- Knowledge rate substitution
  - Something between 20% to 25%
  - This means that we need to replace knowledge every 4 – 5 years
- As a result, a territory must be also organised with the ability to produce knowledge
  - Need schools, research and focus
  - Additional concerns to add to energy, water and other stuff…
  - …and what is the role for libraries?

Additional notes

- A move from an offer oriented organisation to a client oriented strategy also has implications in the way we deal with information
  - New ways to deal with information discover
    - How to be discovered (offer side)
    - How to discover (client side)
  - Individuals can no more stand on their own ability; must rely mostly on their network connections
  - …again, what is the role for libraries?

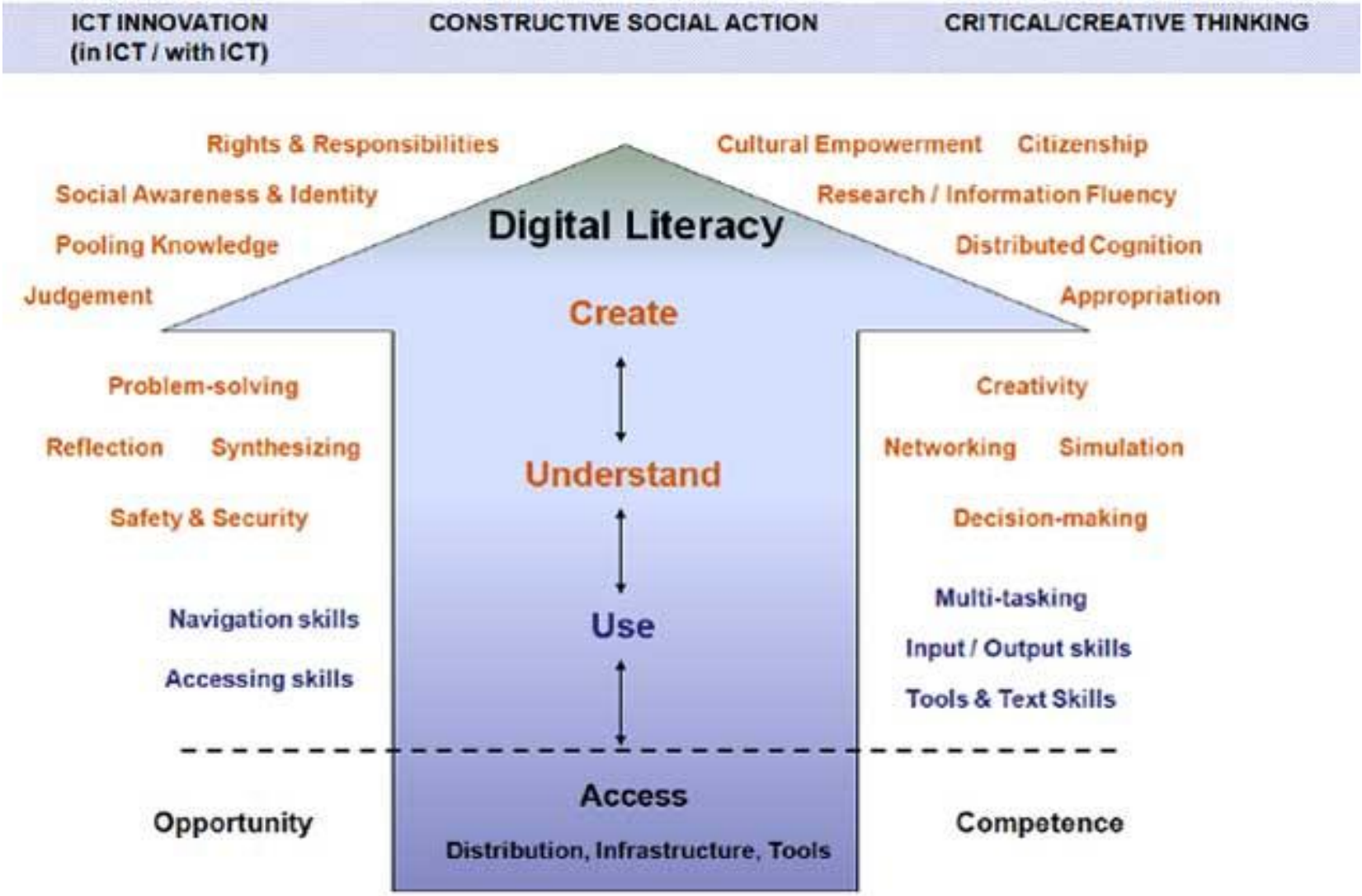
Digital literacy

"Digital literacy is the ability to understand information and —more important —to evaluate and integrate information in multiple formats that the computer can deliver. Being able to evaluate and interpret information is critical [...] you can't understand information you find on the Internet without evaluating its sources and placing it in context".

Paul Gilster

Read more at http://www.ascd.org/publications/educational-leadership/nov97/vol55/num03/A-New-Digital-Literacy@-A-Conversation-with-Paul-Gilster.aspx

Being part of a bigger and complex set



How much information?

- *"2002 could be considered the beginning of the digital age, the first year worldwide digital storage capacity overtook total analog capacity. As of 2007, almost 94 percent of our memory is in digital form"*

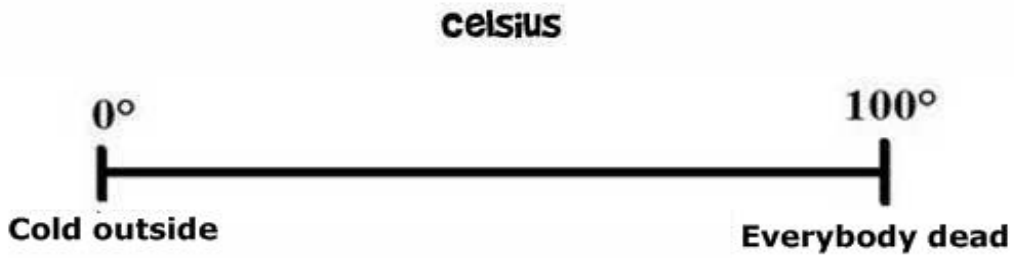
- Read more on http://news.usc.edu/#!/article/29360/How-Much-Information-Is-There-in-the-World
- Watch a video about the Martin Hilbert study: http://vimeo.com/20928251

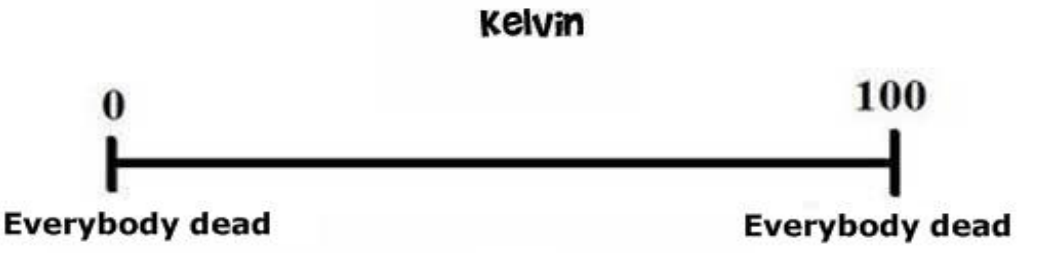
Deal with it…

- Information overload
  - Issues of quantity (to much information to cope with…)
  - Issues of quality (to complex information to cope with…)
  - Issues of synchrony and trueness (can we deal with channel diversity and confirm information trueness…)
- How to deal with information overload?
  - Management (tasks, goals, time, information, relationships, attention, …)
  - Relate, structure, prioritise
  - Say no!
  - Do not have it, link it!
  - Filter, discard
  - Share, collaborate
  - Focus and forget
  - Learn to network a lot

Turn knowledge into understandable information

- Translate it to real world relationships…
  - compare and relate
  - use multimedia and visual complexity





See

http://www.visualcomplexity.com

With Internet the issue of publication also changes…

- See this 9:29 min video from Aaron Swartz – **The Network Transformation** *http://www.youtube.com/watch?v=CzNXDdjtXQI&noredirect=1*

Internet activist
(1986, 2013)

Digital Libraries

The World Digital Library: http://www.wdl.org/en/

grassroots

- Vannebar Bush with As we May Think (1945), paper
  - http://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/
- J. C. R. Licklider with Libraries of the Future (1965), book
  - http://www.incunabula.org/2013/02/j-c-r-licklider-libraries-of-the-future-1965/
- Those authors are part of the Internet pioneers, who antecipate in many ways our today reality
  - http://www.ibiblio.org/pioneers/licklider.html
- Also William Arms in its 2000 book, provide a reference for the use of the digital libraries idea
  - http://www.cs.cornell.edu/wya/diglib/
- The idea of linking and relating information and provide a more semantic oriented access is beyond just better access and goes further into learning and understanding not just authors work but the knowledge it represents
  - The very idea of libraries…

From Licklider book (1965)

"If books are intrinsically less than satisfactory for the storage, organization, retrieval, and display of information, then libraries of books are bound to be less than satisfactory also. We may seek out inefficiencies in the organization of libraries, but the fundamental problem is not to be solved solely by improving library organization at the system level.

Indeed, if human interaction with the body of knowledge is conceived of as a dynamic process involving repeated examinations and intercomparisons of very many small and scattered parts, then any concept of a library that begins with books on shelves is sure to encounter trouble."

Developments…

- From 1965 to early 1990s
  - The building blocks: digital storage, electronic processors, computer networks, natural language processing, text formating and scanning, optical character recognition, indexing, databases, and the Internet
  - Computer and information sciences
  - Online information industry
  - Libraries, standards and automation (MARC – machine-readable cataloguing; OCLC – Online Computer Library Center)
  - Archives and other professional communities (TEI – Text Encoding Initiative; SGML – Standard Generalised Markup Language)

Early vision of digital library (1991)

- *easy, fast, and convenient access to the world's information (regardless of where that information is stored) at any time, from anywhere in the world*
- *effective storage and organization of massive amounts of text, multimedia and data beyond the bounds of what even the largest single library could provide*
- *organization and access to materials in many languages*
- *greatly improved searching and browsing capabilities*
- *interoperability enabling the cross-searching of many diverse collections at once*
- *direct, instant delivery of information and data to multiple users at the same time*
- *transformative improvements in support for research and education globally; better support for interdisciplinary work and scholarly collaboration across institutions and around the world*
- *significant cost savings over traditional (duplicative) methods for cataloging, storing and preserving analog materials.*

Drives for the change

- Solve large scale long standing challenges: organise, access and relate the growing corpus of information and knowledge available
- Time become a more and more important issue
- To much information, not to consider machine and computer support
- The growing demand from universities and scholarly communication
- The digitisation movement (fuels the way to innovative projects as Mercury, CORE, Tulip, Red Sage and many others)

The Kahn Wilensky architecture

- The general principles for the design of a digital library that is "open in its architecture and which supports a large and extensible class of distributed digital information services"
- Has Four main component types plus a "other" group:
  - **repositories**, ranging from file systems to distributed storage systems for content;
  - mechanisms to support **search** (indexing or metadata);
  - **identifier systems** for identifying and locating digital objects;
  - **user interfaces** to perform user services (for example searching, browsing, visualization, delivery)
  - Other components may include security systems for authenticating users, services to aggregate search results from multiple sources, and tools for supporting collaboration and other types of interaction

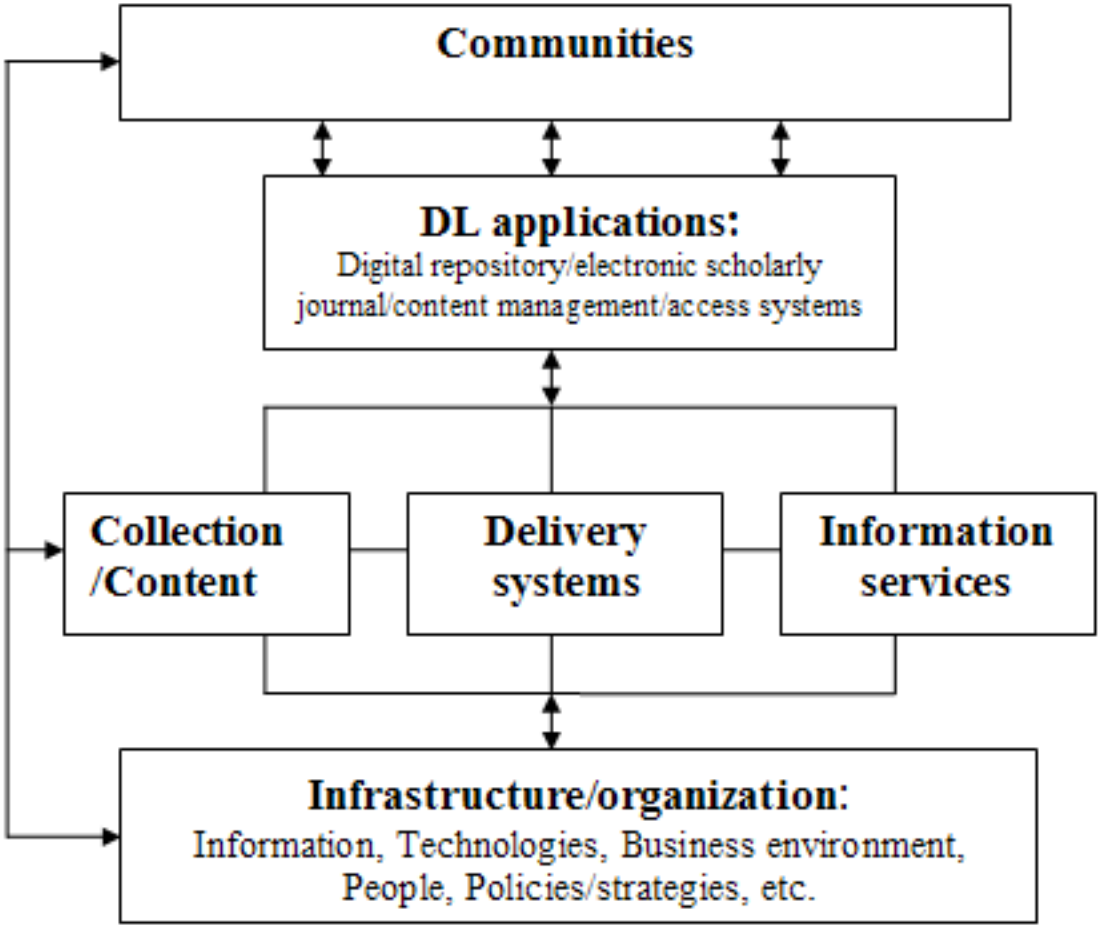
history

- Digital Library Initiative (DLI-1) this funding line of projects provides the move for the vision, aim and shift to digital libraries as we knew them today (1994)
- Other projects:
  - UK eLib Programme (eLib) (1993), http://www.ukoln.ac.uk/services/elib/
  - Digital Library Initiative (1994) (DLI phase 1) & Digital Library Initiative (1998) (DLI phase 2) http://www.dlib.org/dlib/july98/07griffin.html
  - Joint NSF/JISC international projects (EUA & UK) (1998) http://www.dlib.org/dlib/june99/06wiseman.html
  - European Commission (EC) (1991) http://cordis.europa.eu
  - Over the world, a number of projects were conducted from 1996

Some world known projects

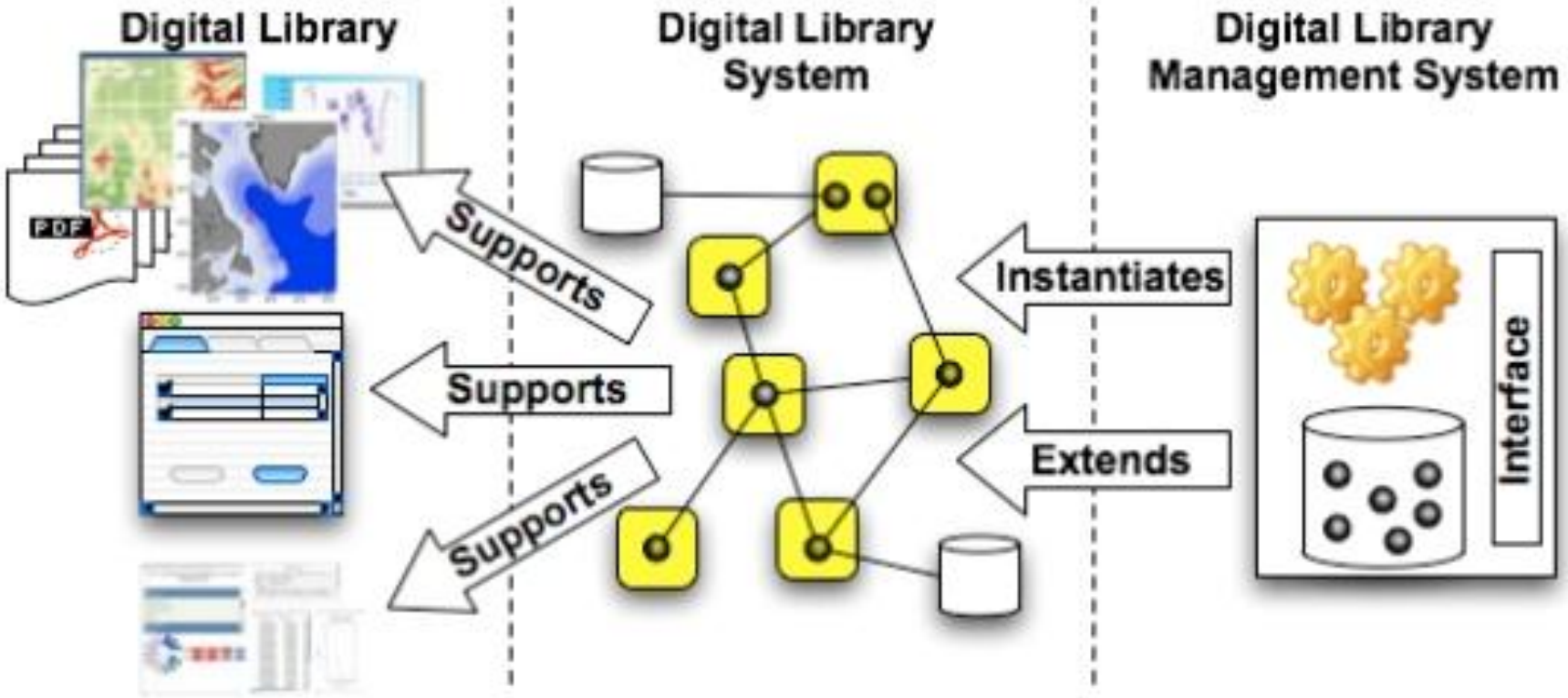
- Project Gutenberg, free books
  http://www.gutenberg.org/
- Internet Archive, archiving the Internet
  https://archive.org
- The Million books project, the universal digital library
  http://www.ulib.org/
- The World Digital Library
  http://www.wdl.org/pt/
- Europeana, Europe's digital library, museum and archive
  http://www.europeana.eu

Definition of a digital library based on a practice community (Choi and Rasmussen, 2006)



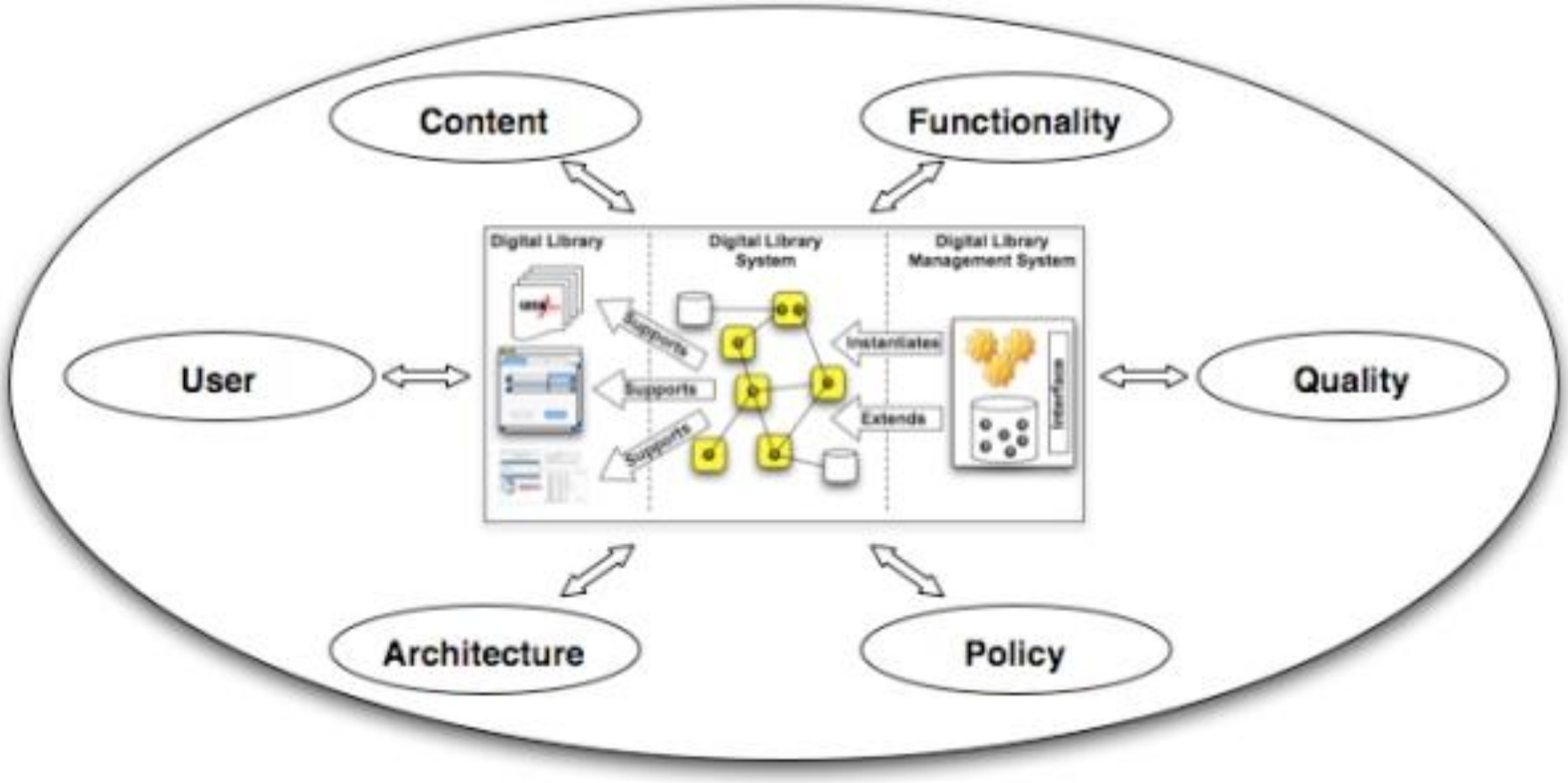
http://www.dlib.org/dlib/september06/choi/09choi.html

A three-tier framework (Candela et al, 2007)



http://www.dlib.org/dlib/march07/castelli/03castelli.html

Concepts related with DLs (Candela et al, 2007)



http://www.dlib.org/dlib/march07/castelli/03castelli.html

Definition of digital library

- Not “a definition” but many alternatives, like these:
- Borgman (2000)
  - Digital libraries are a set of electronic resources and associated technical capabilities for creating, searching, and using information. Digital libraries are constructed – collected and organized – by [and for] a community of users, and their functional capabilities support the information needs and uses of that community
- Bishop, Van House and Buttenfield, (2003)
  - Sociotechnical systems – networks of technology, information, documents, people, and practices

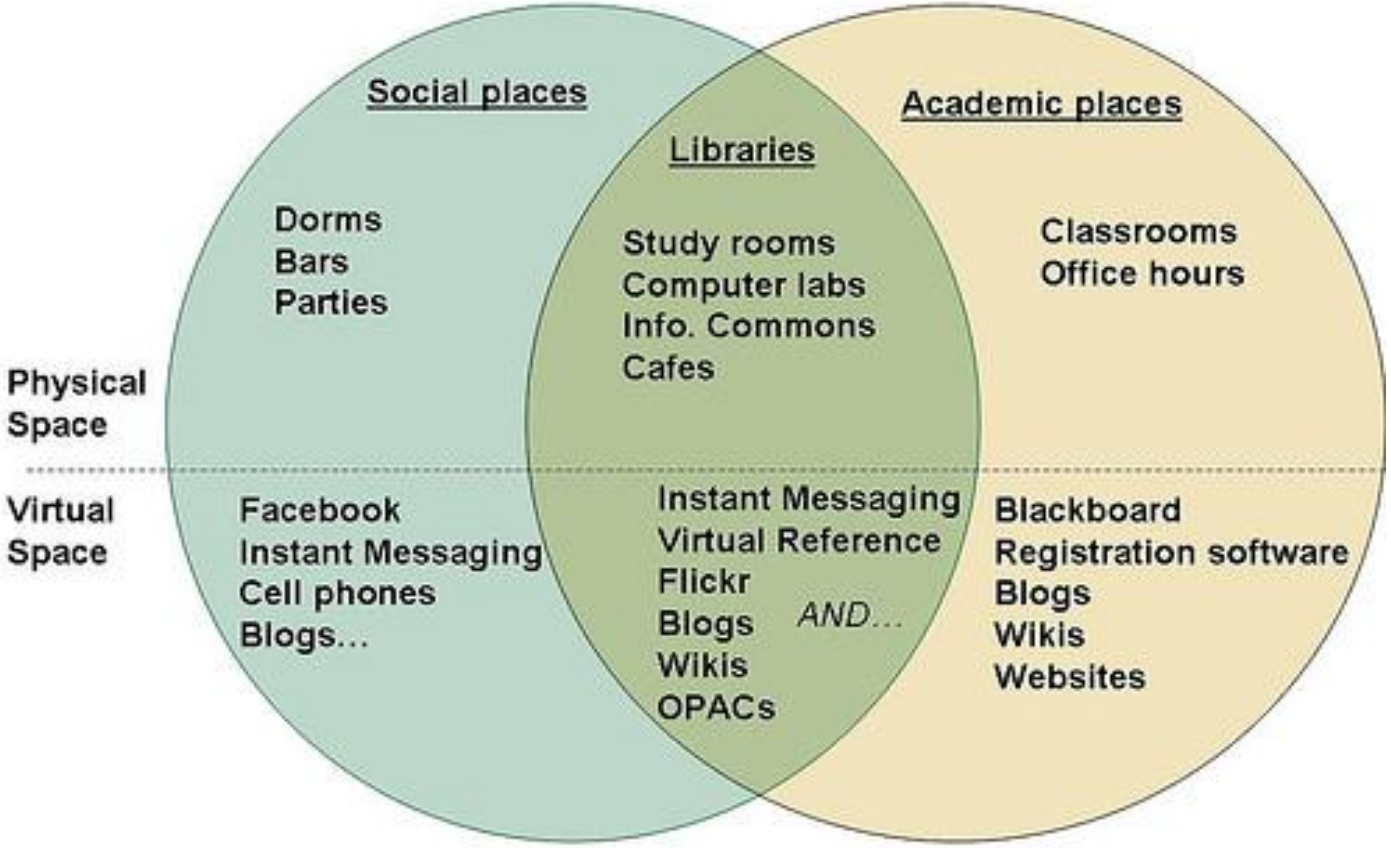
About preservation

- the curatorial role of digital libraries as managed collections, requiring that digital objects be selected, made accessible, and preserved as *long-term, stable resources*
- *Preservation is a critical and quite challenging issue* where the digital becomes a even higher challenge
  - *How to access old formats and mantain support reading devices from several generations?*

Challenges

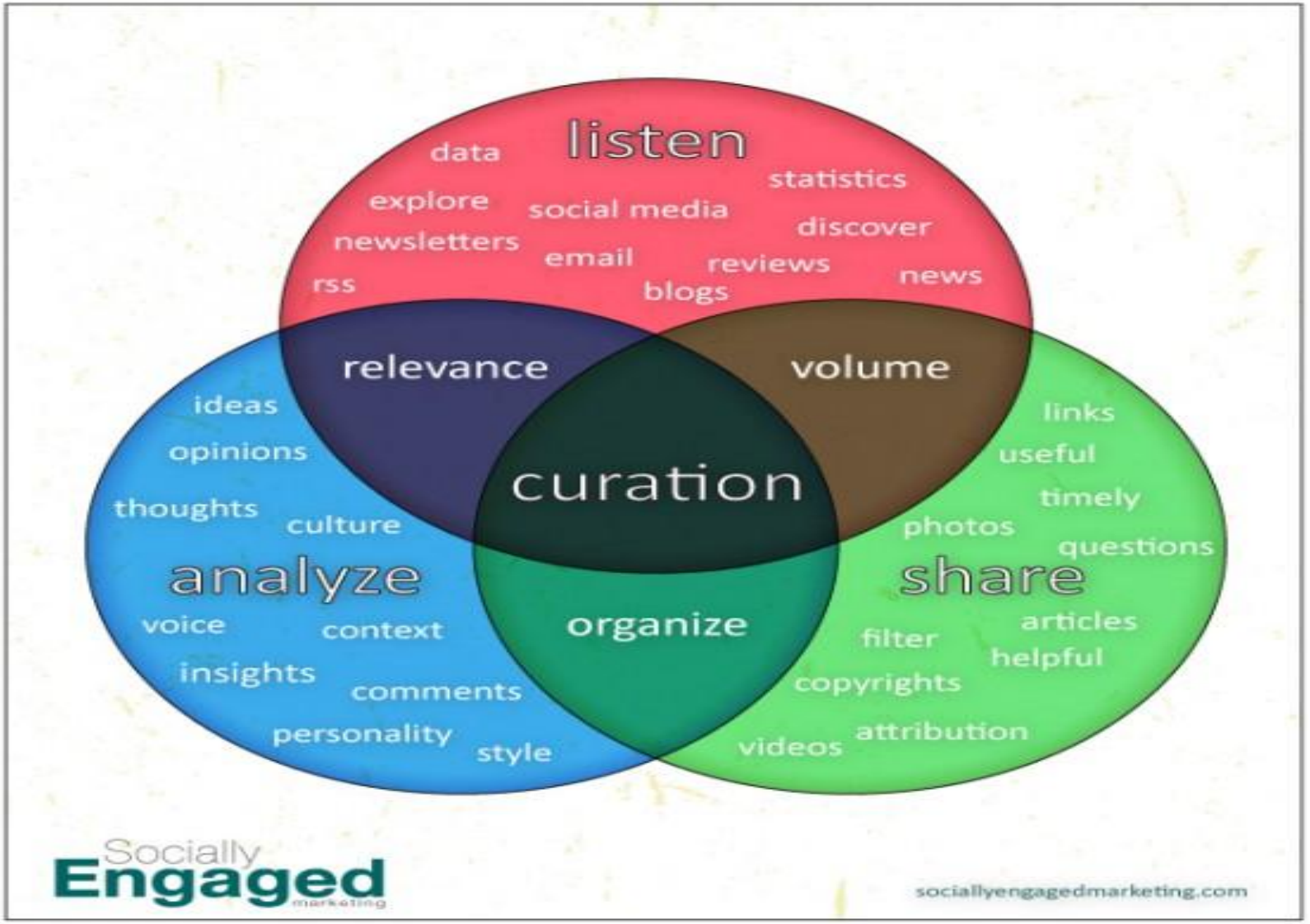
- Distributed digital libraries
- Hybrid libraries
- There is still a physical library and a digital library?
- Preservation issues
- Open access or restricted content
- Protect authorship and collaborative work
- Protect content and property
- Security issues
- Global versus locally

Digital Library as third place



http://mchabib.com/2006/10/05/digital-library-as-third-place/

Information & digital curation



Information curation

- *Curation is the act of individuals with a passion for a content area to find, contextualize, and organize information. Curators provide a consistent update regarding what's interesting, happening, and cool in their focus. Curators tend to have a unique and consistent point of view providing a reliable context for the content that they discover and organize.“*
  Steven Rosenbaum http://www.fastcompany.com/1834177/content-curators-are-new-superheros-web

- Plan and oversee the arrangement, cataloguing, and exhibition of collections. It additionally describes and analyses valuable objects for the benefit of researchers and the public

http://informationcuration.wikispaces.com/

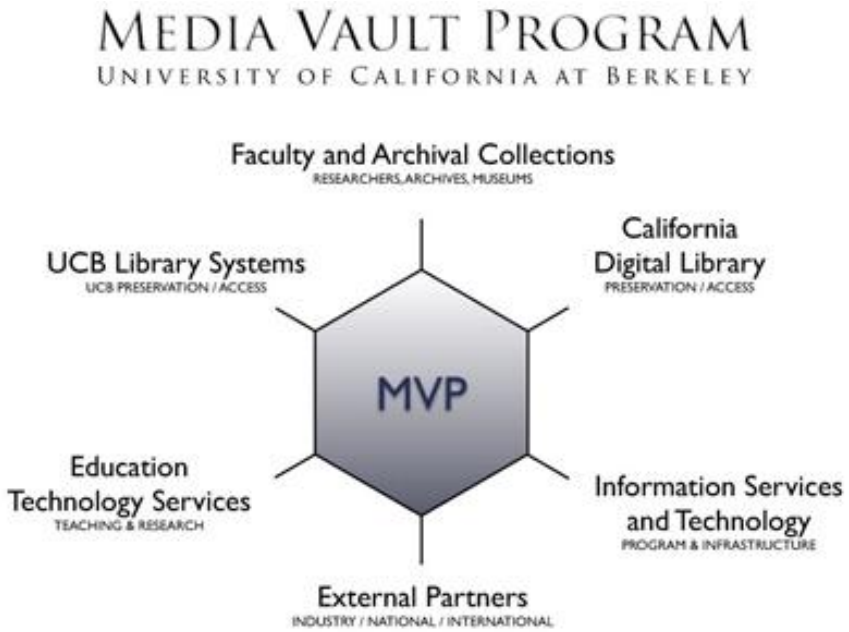
Digital curation

- General term for the activities and strategies that support the stewardship of a particular portion of library collections – digital assets
  - Context and focus
- Consists of producing, acquiring, organizing, maintaining, controlling, preserving, and securing digital assets is an enterprise effort that requires the coordination of personnel and operations putting the right information into where and when it is needed
  - Time and space requirements of information needs

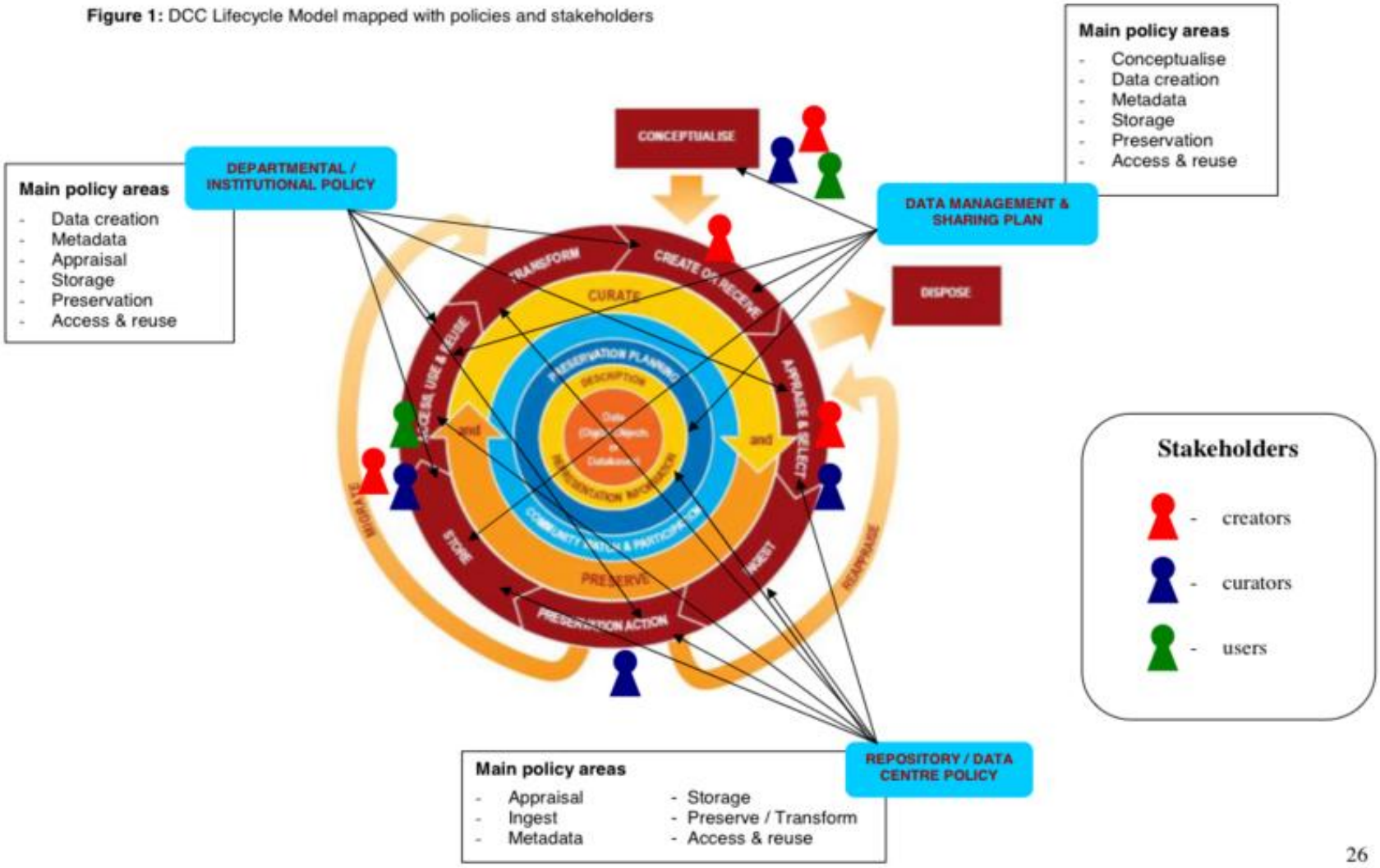
First example

University of California at Berkeley, academic information and library services

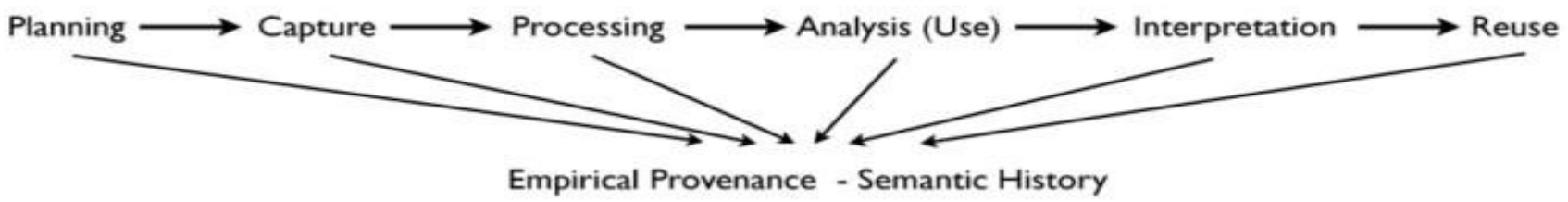
- Organising a workflow for disaster recovery in a campus context and its information and library services



Digital curation



http://mediavault.dreamhosters.com/wiki/Lifecycle



Scholarly Activity	Discover	Gather	Create	Share	Save
Publication	Go to library, search online, talk with colleagues, review own materials	capture in notebook, netbook, laptop, desktop, iPhone	Word doc, references, images, figures, PDF	Send to publisher, publish online via FTP, blog, repository	Backup locally, Send to repository via ingest process (OAIS), convert to complex digital object (METS)
Tools	Finding aids, web search, Collections	bSpace, Zotero, Portfolio, network storage, email, iPhoto	MS/Open Office Endnote, Photoshop, Illustrator, iPhoto, PDF, Omeka	MV Publish, e-Scholarship, bSpace, Calmail	Time Machine, UC Backup, Digital Preservation Repository, METS
Provider	Library, Google, MV collections	ETS, Zotero, IST, gmail/calmail	MS/Open Office, MV DAM (IST), bSpace (ETS), Omeka (Okapi)	Publishers, Flickr, Facebook, IST, Library, CDL, ETS	Apple, IST, CDL, LIbrary Services
Function	Search	Digital Asset Management	Productivity Tools	Publishing	Archiving

http://mediavault.dreamhosters.com/wiki/Lifecycle

Second example

My Digital Library – MyDL, personal academic activity

- A shorter, more low scale and replicable exercise
  - Having a number of assets on slideshare and within other information sources
  - Using wordpress to compose stories that provide context and a time/space flux

My Digital Library – MyDL

Luis Borges Gouveia

MYDL NEWS | **E-LEARNING** | INFOSEC | MY BOOKS | ABOUT

e-learning

(working in progress…)

A number of works, presentations, prjects and publications have been made around the theme of e-learning and the use of information and communication technologies for the teaching and learning process support.

Among those, there are some provocative presentations about how the digital may impact e-learning activities: some present a refletive position presented for secondary school, in April 2010 (slidesPT, 2010) or for higher education, in FEUP, July 2010 (slidesPT, 2010).

A number of projects provide technical work on learning management plaftform in early as 1999 (in which, we propose a LMS system with some of the features that come regulary in Moodle and other similar systems). The EFT-Web has an initial demo on (slidesPT, 1999) – the system concepts were presented in 4th Ocotber 1999, in Lisbon conference (slidesPT, 1999)

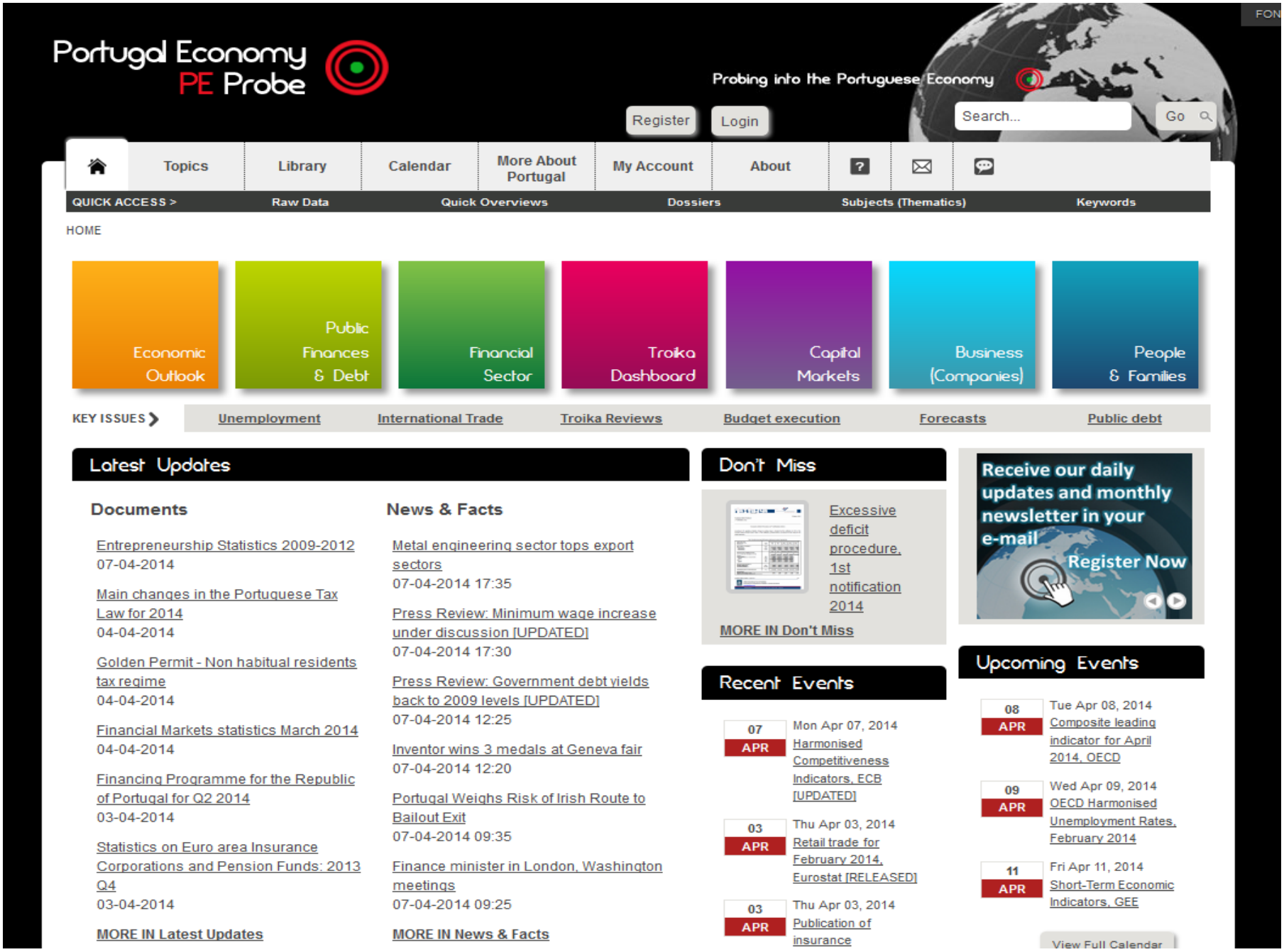
One of my main projects in the are was the development of the Virtual University initiative that starts with this presentation on the university in June, 2003 (slidesPT, 2003).

http://lmbgdigitallibrary.wordpress.com/e-learning/

Third example

Portugal Economy PE Probe, economic information

- Organising information and put it in context, based on a critical but selective use of own created materials combined with other from multiple sources
- It is a collection of economic and finantial information about Portuguese ongoing economic status and performance



http://www.peprobe.com/

Fourth example

Pordata, statistical database

- Organise statistical data and structure it in groups that can be combined into integrated statistical series
- Goes beyond data gathering as it provides a service to regulate and turn compatible statistical series from diferent sources and contexts, in order to allow to use them together

PORDATA
Base de Dados Portugal Contemporâneo

Search the various databases

Select Data

Versão Portuguesa | Personal Area

FUNDAÇÃO FRANCISCO MANUEL dos SANTOS

Search Environment | Portraits

THOUSANDS OF STATISTICS ABOUT MUNICIPALITIES, PORTUGAL AND EUROPE

Municipalities database analysed in 630 indicators divided into 13 themes

MUNICIPALITIES

Portugal database analysed in 1074 indicators divided into 16 themes

PORTUGAL

Europe database analysed in 576 indicators divided into 11 themes

EUROPE

Pordata.pt wins WSA award

Distinguished as one of the world's best projects in e-Science & Technology.

Latvia joins the Eurozone

Find out more about the 18th member of the Eurozone.

Indicator Highlights

MUNICIPALITIES

- Marriages between people of the opp...
- Resident population

PORTUGAL

- Resident population, according to C...
- Marriages

EUROPE

- GDP per capita (PPS)
- Real GDP growth rate

Pordata

PORDATA is a public service and inclusive project devised for a large number of users who share an interest in discovering more about Portugal. From today onwards, it is with

Francisco Manuel dos Santos Foundation

The **PORDATA** Contemporary Portugal Database reflects the main priority for the first years of the Francisco Manuel dos Santos Foundation. The Foundation intends to devote

http://www.pordata.pt/en/

Final remarks

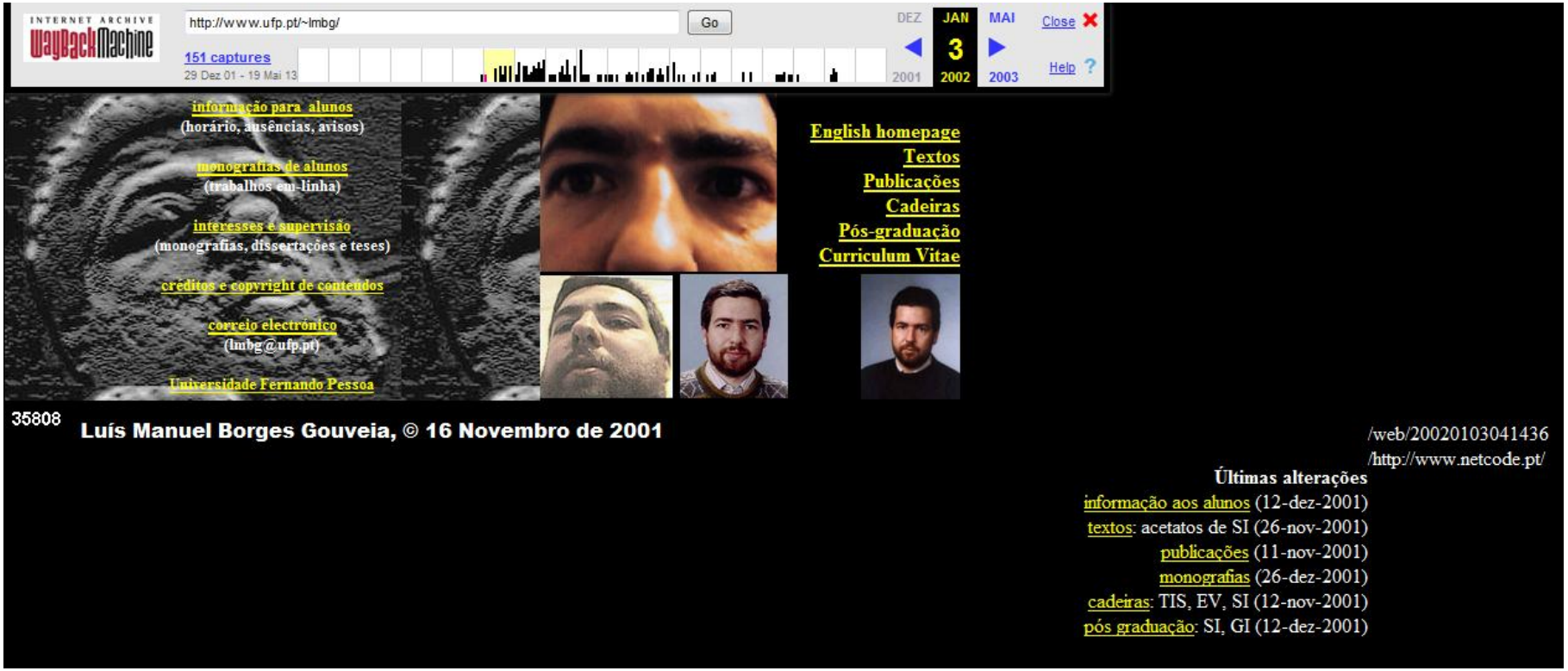
- "*We live today not in the digital, not in the physical, but in the kind of minestrone that our mind makes of the two*."
  Paola Antonelli, senior curator in the Department of Architecture and Design at The Museum of Modern Art

- "*More connections to more devices means more vulnerabilities*."
  Marc Goodman, Global Security Advisor and Chair for Policy and Law at Singularity University

50 thought-provoking quotes about libraries and librarians
http://ebookfriendly.com/best-quotes-about-libraries-librarians/

Turning Internet into an historic archive

- Internet Archive (Wayback Machine) https://archive.org/

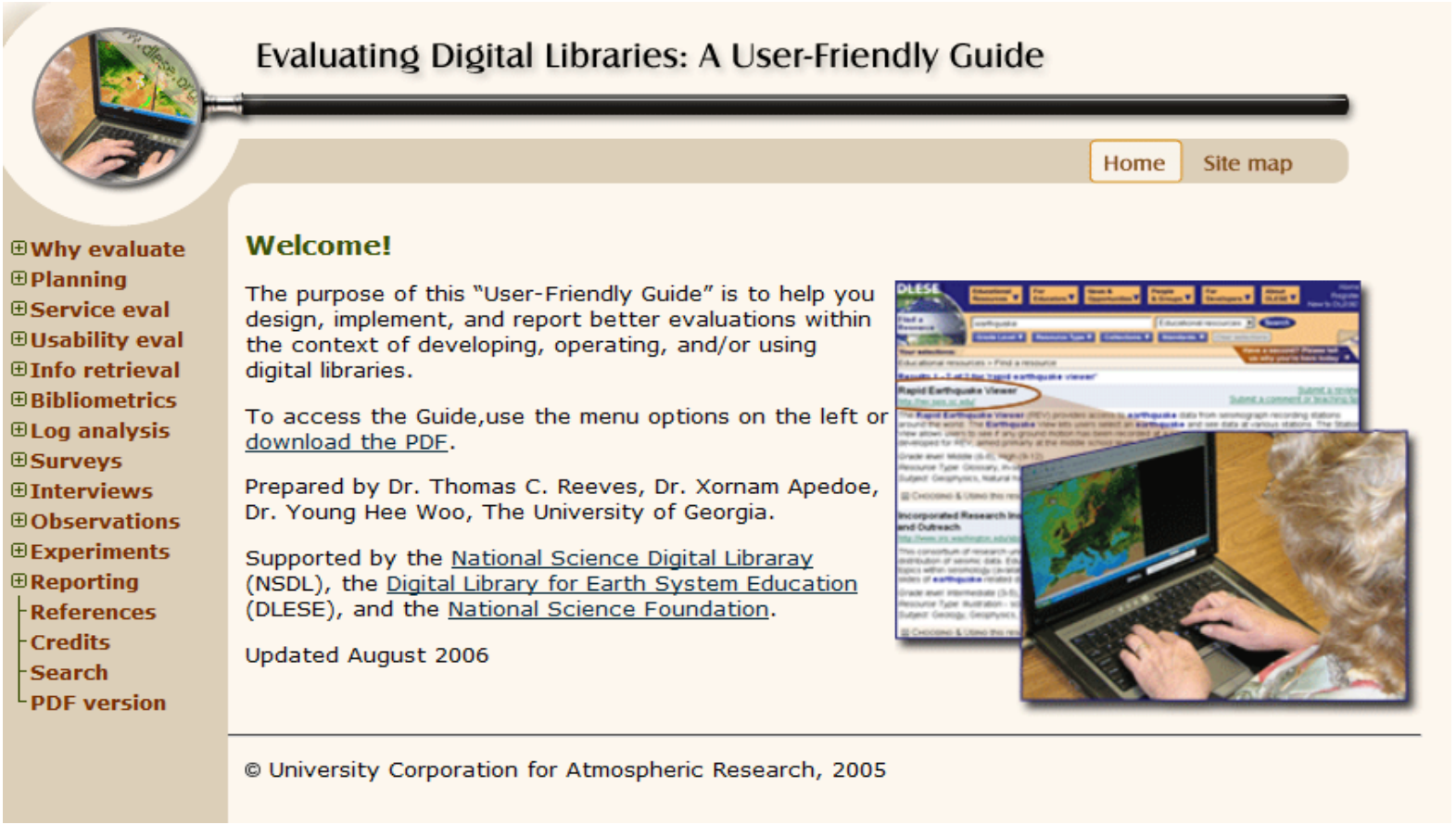


One public service digital library example

- Lancashire County Council DL (England)

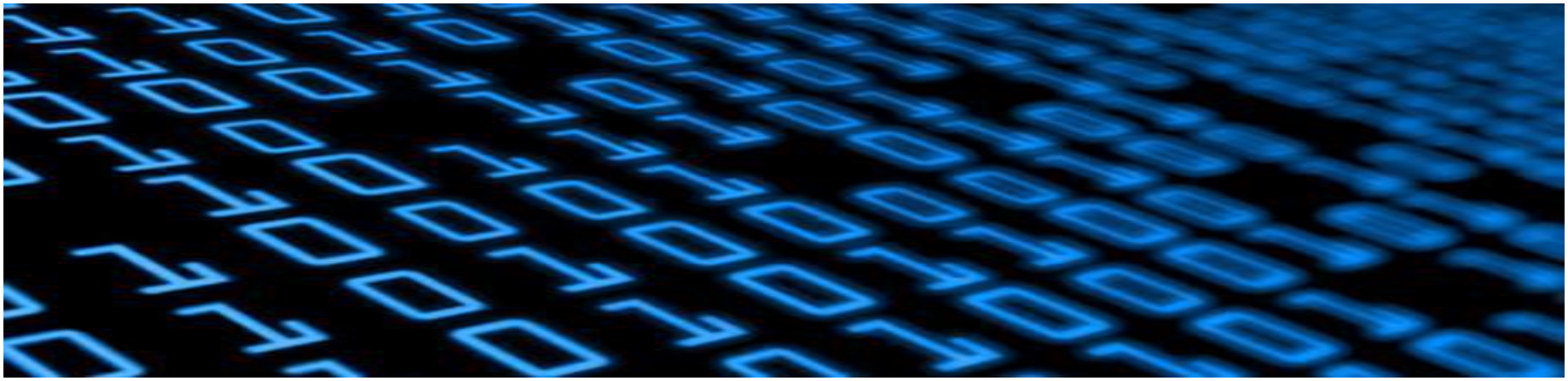
Evaluating Digital Libraries

- A user-friendly guide (12 chapter/98 page manual & website) http://www.dpc.ucar.edu/projects/evalbook/index.html



A video on DL issues

- Building (and using) Big Digital Libraries
  John Unsworth, MIT Video
  http://video.mit.edu/watch/building-and-using-big-digital-libraries-john-unsworth-11546/
- *Commoncraft: a digital library of teaching videos*
  http://www.commoncraft.com/videolist



Much more to be said…

- All a world of techno stuff need to be considered
  - Open Acess
  - Interoperability
  - Repositories
  - Technology standards
- Also a number of issues need to be dealt with
  - DL statistics
  - Directories and publishers
  - Digital libraries services
  - DL integration issues (with libraries and other systems)
  - Legal & security issues
  - DL economics and funding
  - DL initiatives and digital platforms

Some related TED & TEDx talks

- Brewster Kahle: A free digital library
  - http://www.ted.com/talks/brewster_kahle_builds_a_free_digital_library
- Jeff Skoll: Media with meaning
  - http://www.ted.com/playlists/21/media_with_meaning
- Jamais Cascio: Tools for a better world
  - http://www.ted.com/talks/jamais_cascio_looks_ahead
- Jim Wales: on the birth of Wikipedia
  - http://www.ted.com/talks/jimmy_wales_on_the_birth_of_wikipedia.html
- Kevin Kelly: The next 5,000 days of the Web
  - http://www.ted.com/talks/kevin_kelly_on_the_next_5_000_days_of_the_web

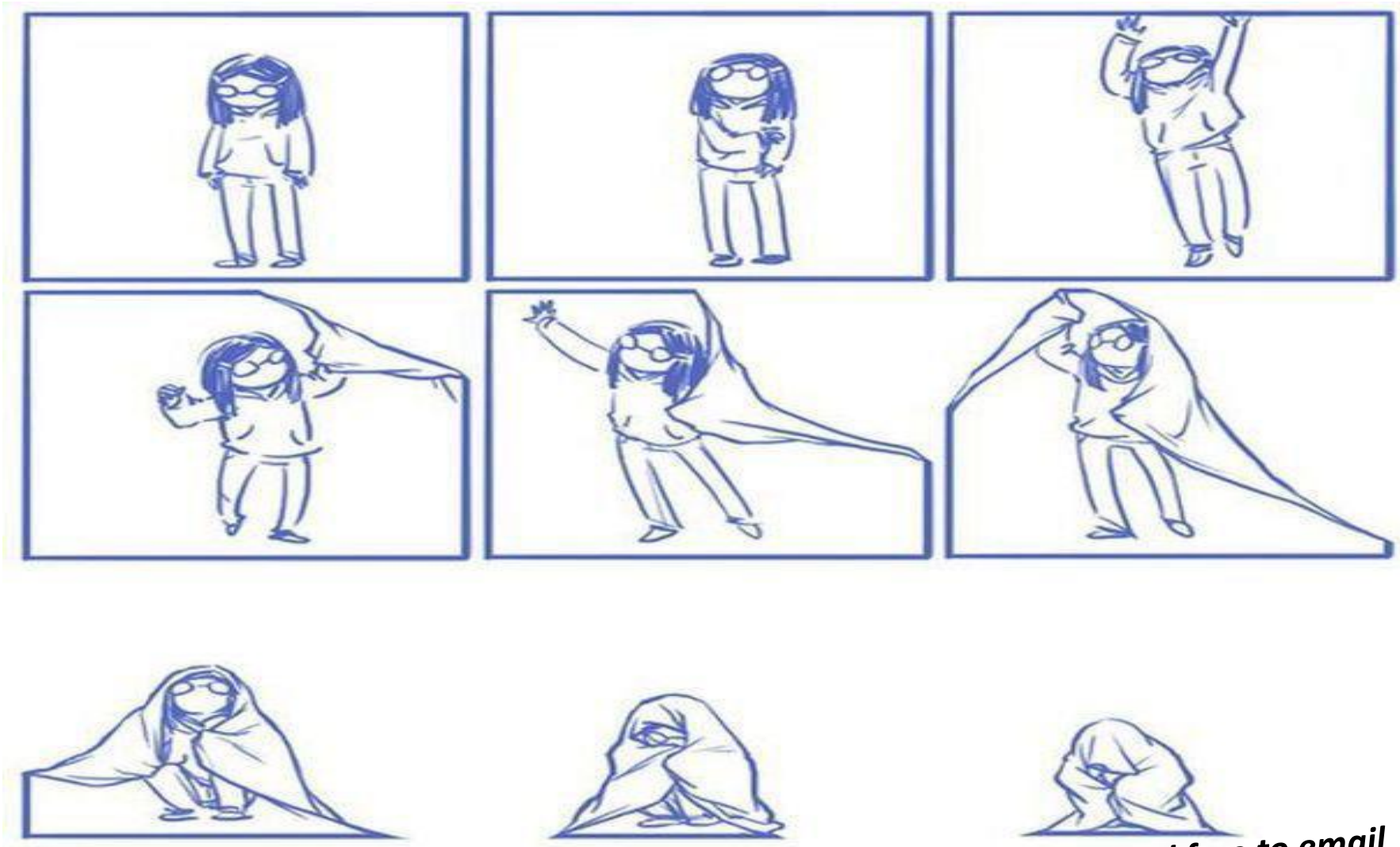
Suggested readings

- James Gleick. (2011). *The Information. A History, a Theory, a Flood*. Fourth Estate. *http://around.com/the-information/*
- G. G. Chowdhury, Schubert Foo. (2012). *Digital Libraries & Information Access*. Facet Publishing. http://www.facetpublishing.co.uk/title.php?id=8217&category_code=301
- Karen Calhoun. (2014). *Exploring Digital Libraries*. Neal-Schuman Publishers. http://www.alastore.ala.org/detail.aspx?ID=4247
- John J. Burke. (2013). *The Neal-Schuman Library Technology Companion*. Fourth Edition. Neal-Schuman Publishers. http://www.alastore.ala.org/detail.aspx?ID=4298
- William Y. Arms. (2000). *Digital Libraries*. The MIT Press. http://www.cs.cornell.edu/wya/diglib/

Online references

- Digital libraries bibliography
  http://onlinebooks.library.upenn.edu/webbin/book/browse?type=lcsubc&key=Digital%20libraries%20--%20Bibliography
- World Digital Library Project Site
  http://project.wdl.org/
- D-Lib Magazine
  http://www.dlib.org/
- IFLA and the Information Society
  http://www.ifla.org/information-society
- The old IFLA site on Digital Libraries
  http://archive.ifla.org/II/diglib.htm
- Digital Library Federation
  http://www.diglib.org/

Hope the workshop provide some value



Feel free to email

…and enjoy Porto!