

# Cutting edges topics in Digital Libraries

**RAFFAELLA BERNARDI**

UNIVERSITÀ DEGLI STUDI DI TRENTO

P.ZZA VENEZIA, ROOM: 2.05, E-MAIL: BERNARDI@DISI.UNITN.IT

# Contents

1	Main topics of the course .....	4
2	The Future of Digital Libraries 1998 .....	5
2.1	The Future of Digital Libraries 2005 .....	6
2.2	The Future of Digital Libraries 2009 .....	7
3	Today cutting edge library keywords .....	8
3.1	Open Access: at its birth .....	9
3.2	Open Access: today .....	10
3.3	Linked Data .....	11
3.3.1	Linked Data: an example .....	12
3.3.2	Resource Description Framework (RDF) .....	13
3.3.3	Linked Data from 2007 to 2009 .....	14
3.3.4	Linked Data in 2010 .....	16
3.3.5	Cross-domain Linked Data sets .....	17
3.3.6	Search and Linked Data .....	18
3.4	Linked Data and Libraries .....	19
3.4.1	Library and Linked Data RDF example .....	20
3.4.2	Library Linked Data Incubator .....	21

3.4.3	Next Generation Library Catalogs? .....	22
3.5	E-Books .....	23
3.5.1	E-book history .....	24
3.5.2	Third Generation .....	26
3.5.3	Why aren't ebooks more popular .....	27
3.5.4	Friday 20th of May 2011 News .....	28
3.5.5	Google and E-books .....	30
3.6	Book .....	31

# 1. Main topics of the course

- History
  - Libraries: From catalog card to OPAC to Digital Libraries.
  - CS: From big computers with small memory, to small chips with large memory.
  - Web: http protocol, web services, semantic web, and Open linked data.
- Search
  - Search in DB: from query languages to Information Retrieval (VSM)
  - Federated Search: search in more DB.
- Evaluation:
  - Gold standards: guidelines, annotation and Inter annotator agreement
  - Eval. measures: relevance, ranking, position
- Language Technologies
  - NLP, thesauri, NER, etc.
  - CLIR

## 2. The Future of Digital Libraries 1998

Michael Lesk <http://www.lesk.com/mlesk/> (Understanding DL's author)

“I will talk on how I see value coming from digital information, and what kinds of changes I see coming about as a result. To summarize in advance,

1. Digital libraries are now economically efficient, and the area is booming;
2. Digital technology offers great advantages for libraries;
3. The adoption of digital information will mean *changes in the role of libraries*, and in how we *manage* them.

In summary, it is clear that digital libraries are coming, both in free and commercial versions, and in both image and ascii formats. This is an opportunity, not a threat. Digital information can be more effective for the users and cheaper for the librarians. *Access will become more important than possession*. But this must be used to encourage *sharing*, not competition. The real asset of a library is the people it has who know how to find information and how to evaluate it. It must emphasize its skills in training, not in acquisition. If we think of information as an ocean, the job of libraries is now to provide navigation. It is no longer to provide the water.”

More at <http://www.lesk.com/mlesk/kanazawa/kanazawa.html>

## 2.1. The Future of Digital Libraries 2005

“Where do we go from here?” by Clifford Lynch

A good deal of what will happen here I would characterize as *advanced technology deployment in production systems* rather than pure research, much like what takes place in the deployment of high performance computer communications networks to support the research and education community. It is work that combines production engineering and research in complex and delicate balances.

Perhaps the overarching theme here, and it is one that may point to a major direction for research that follows on the last decade of progress in digital libraries, is *connecting and integrating digital libraries with broader individual, group and societal activities*, and doing this across meaningful time horizons that recognize digital libraries and related constructs as an integral and permanent part of the evolving information environment. The next decade for digital libraries may well be characterized most profoundly by the transition from technologies and prototypes to the ubiquitous, immersive, and pervasive deployment of digital library technologies and services in the broader information and information technology landscape.

More at <http://www.dlib.org/dlib/july05/lynch/07lynch.html>

## 2.2. The Future of Digital Libraries 2009

“To employ the fullest potential of digital technology in partnership with users by enabling seamless and *open access* to all types of information without limits to format or geography, and to enhance the ability of libraries, archives and museums to *collaborate* among themselves and with others to offer the broadest and most complete service possible.”

This statement includes the main conclusions: a) technology is not enough b) we need *cooperation with users* c) we need international cooperation with *cultural institutions and partnership with others* (publishers, et al.). [IFLA, 2009]

IFLA (The International Federation of Library Associations)

### 3. Today cutting edge library keywords

- Open Access
- Linked Data
- E-book

**Mashup era:** a mashup is a Web page or application that uses and combines data, presentation or functionality from two or more sources to create new services. (e.g. Google maps are used in all different contexts.)

In DL, enrich catalogues with information from Wikipedia, Amazon etc.

### 3.1. Open Access: at its birth

“Timeline of the Open Access Movement” the open access movement’s ‘constitutional convention’ was in December 2001

To achieve open access to scholarly journal literature, we recommend two complementary strategies.

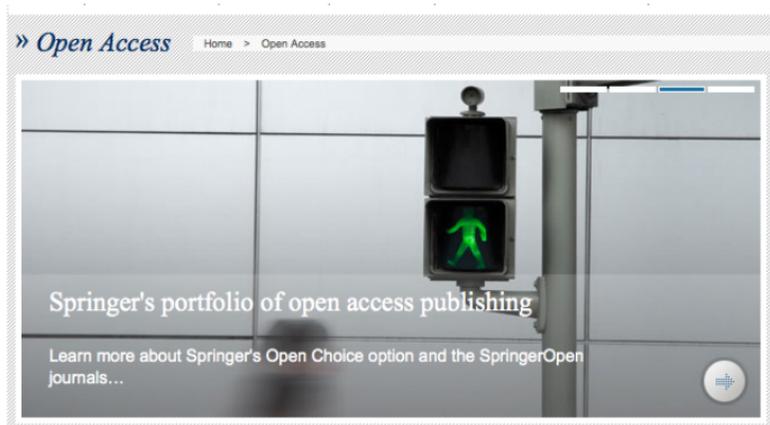
1. Self-Archiving: refers to making “e-prints” available on the Web. An eprint is either a digital preprint or a postprint.
2. Open-access journals

This requires new Business Models; new licenses (e.g. Creative Commons License).

“The open access movement has gained considerable traction in the last six years. It has become the most successful scholarly publishing reform movement in modern times, and it has begun to transform the scholarly communication system.” [Bailey, 2006]

## 3.2. Open Access: today

Example: SpringerOpen



“All SpringerOpen journals are fully open access and the entire content is freely and immediately available online on SpringerOpen.com as well as on Springer’s online platform SpringerLink. All articles are published under the Creative Commons Attribution license, enabling authors to retain copyright and comply with open access mandates.”

### 3.3. Linked Data

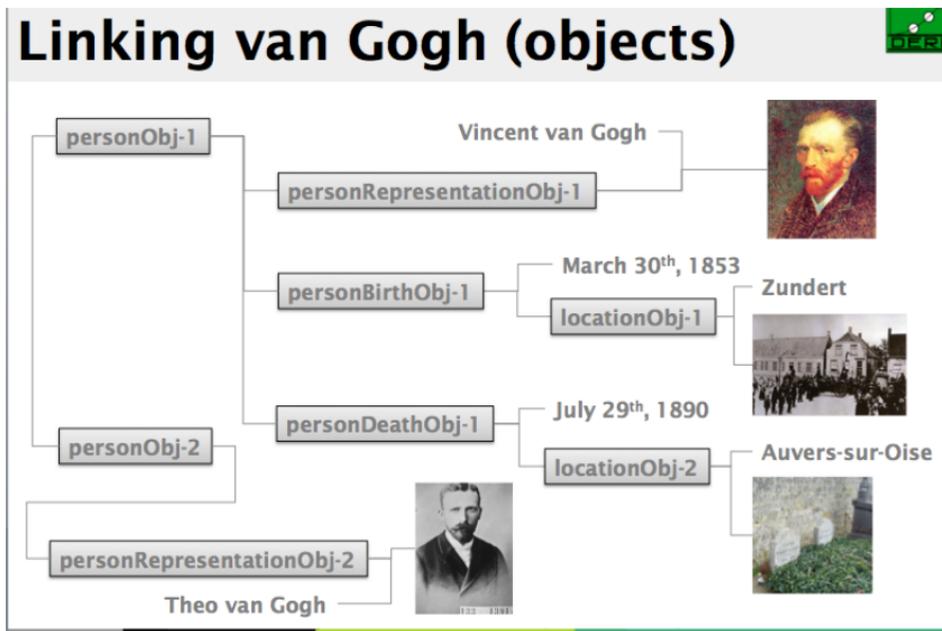
Tom Heath and Christian Bizer “Linked Data. Evolving the Web into a Global Data Space”. Morgan & Claypool Publishers.

Linked Data: Set of principles and technologies to enable data sharing and reuse on a massive scale.

First sub-goal (need): to have well structured data. The more regular and well-defined the structure of the data the more easily people can create tools to reliably process it for reuse.

While most Web sites have some degree of structure, the language in which they are created, HTML, is oriented towards structuring textual documents rather than data: it is hard for software applications to extract snippets of structured data from HTML pages.

### 3.3.1. Linked Data: an example

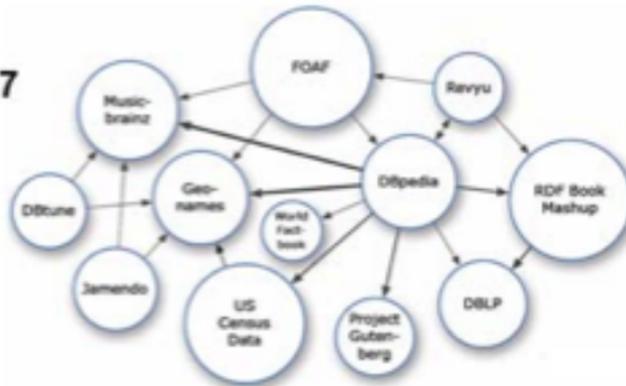


**3.3.2. Resource Description Framework (RDF)** This mechanism is provided by the Resource Description Framework (RDF): The key things to note at this stage are that RDF provides a flexible way to describe things in the world such as people, locations, or abstract concepts and how they relate to other things.

Increasing numbers of data providers and application developers have adopted Linked Data. In doing so they have created this global, interconnected data space - the Web of Data or Semantic Web.

### 3.3.3. Linked Data from 2007 to 2009

May 2007





### 3.3.4. Linked Data in 2010

Domain	Data Sets	Triples	Percent	RDF Links	Percent
Cross-domain	20	1,999,085,950	7.42	29,105,638	7.36
Geographic	16	5,904,980,833	21.93	16,589,086	4.19
Government	25	11,613,525,437	43.12	17,658,869	4.46
Media	26	2,453,898,811	9.11	50,374,304	12.74
Libraries	67	2,237,435,732	8.31	77,951,898	19.71
Life sciences	42	2,664,119,184	9.89	200,417,873	50.67
User Content	7	57,463,756	0.21	3,402,228	0.86
	203	26,930,509,703		395,499,896	

### 3.3.5. Cross-domain Linked Data sets

**DBpedia:** a data set automatically extracted from publicly available Wikipedia dumps. Things that are the subject of a Wikipedia article are automatically assigned a DBpedia URI, based on the URI of that Wikipedia article.

**Geonames:** is an open-license geographical database that publishes Linked Data about 8 million locations.

**BBC** has published RDF description of every episode of every TV or radio programme broadcast across the BBCs various channels.

### 3.3.6. Search and Linked Data

#### [Raffaella Bernardi | web site | at DISI](#)

[disi.unitn.it/~bernardi/](http://disi.unitn.it/~bernardi/)

Web pages of R. Bernardi at UniTN. Home page. my details. **Raffaella Bernardi**, Assistant Professor, DISI · Faculty of Arts and Humanities, University of Trento ...

#### [Publications](#)

Raffaella Bernardi and Michael Moortgat Continuation ...

#### [Teaching](#)

Teaching Activities. Committees| PhD Courses | BSc and MSc ...

#### [Other scientific activities](#)

Organizational Activities of Scientific Events. Member of ...

[More results from unitn.it »](#)

#### [Contact me](#)

Contact me at UniTN. Surface address. Raffaella Bernardi DISI ...

#### [Raffaella Bernardi was born ...](#)

Raffaella Bernardi was born on Augustus 9, 1971 in Pescara ...

#### [Talks](#)

Talks. Invited Talks| Other Talks. Invited Talks ^. From words to ...

Search system exploiting link data: Visinav: <http://visinav.deri.org/>: it can answer queries like “give me the URLs of all blogs that are written by people that Tim Berners-Lee knows!”

## 3.4. Linked Data and Libraries

This field has seen some significant early developments which aim at integrating library catalogs on a global scale by:

- interlinking the content of multiple library catalogs, for instance, by topic, location, or historical period;
- interlink library catalogs with third party information (picture and video archives, or knowledge bases like DBpedia);
- and at making library data easier accessible by relying on Web standards.

**Examples** the American Library of Congress and the German National Library of Economics which publish their subject heading taxonomies as Linked Data; while the complete content of LIBRIS and the Swedish National Union Catalogue is available as Linked Data. Similarly, OpenLibrary project.

### 3.4.1. Library and Linked Data RDF example

Linked data is often manifested as a stream of RDF/XML. By querying Freebase (freebase.com) I was able to download an RDF file describing Henry David Thoreau's book called Walden. [1] Upon closer examination of the file you will see that it is made up of "triples" -- subjects, predicates, and objects. For example, the subject of the whole thing is:

```
<fb:common.topic rdf:about="http://rdf.freebase.com/ns/en.walden">
```

There are then a number of predicate/object combinations. Here's one for the author:

```
<fb:book.written_work.author  
rdf:resource="http://rdf.freebase.com/ns/en.henry_david_thoreau"/>
```

Here's one for a subject:

```
<fb:book.written_work.subjects  
rdf:resource="http://rdf.freebase.com/ns/en.house"/>
```

Notice how the author name and subject term are not really values but pointers. Instead they are HTTP-actionable URIs:

- \* [http://rdf.freebase.com/ns/en.henry\\_david\\_thoreau](http://rdf.freebase.com/ns/en.henry_david_thoreau)
- \* <http://rdf.freebase.com/ns/en.house>

### 3.4.2. Library Linked Data Incubator

In order to provide a forum and to coordinate the efforts to increase the global interoperability of library data, W3C has started a Library Linked Data Incubator Group

“The mission of the Library Linked Data incubator group is to help *increase global interoperability of library data on the Web*, by bringing together people involved in Semantic Web activities – focusing on Linked Data – in the library community and beyond, building on existing initiatives, and identifying collaboration tracks for the future.”

“Moreover this activity is no longer isolated in the library environment. Actors from a broader set of domains (*libraries, archives, museums, publishing, the Web*) now share common issues and standards. Their capacity to jointly build interoperable solutions is key to the development of cultural and scientific content across the Web.”

<http://www.w3.org/2005/Incubator/lld/>

**3.4.3. Next Generation Library Catalogs?** “Next-generation library catalogs should implement linked data concepts. Such things will allow computers to create relationships between items much faster and more accurately than humans can. These relationships might not be ”better”, but they will make it easier for humans to evaluate the end result.” “Finally, the use of linked data is yet another example of how librarianship needs to change its methods. We still need to describe materials, but we need to do it differently.” [Eric Lease Morgan Head, Digital Access and Information Architecture Department University of Notre Dame]

**Libris** Around 200 libraries, and also archives and museums. The catalogue is based on Linked Data. <http://libris.kb.se/>

## 3.5. E-Books

**Definitions:** “ebook. noun a book composed in or converted to digital format for display on a computer screen or handheld device” Merriam-Webster.com

“e-book. noun a book that is published in electronic form, for example on the Internet or on a disk, and not printed on paper” Cambridge.org

Now is often used when referring to the methodology of reading just about any mid-to-long-length electronic document that is viewable on a screen, including research papers, short stories and even magazines.

### 3.5.1. E-book history

Catherine C. Marshall “Reading and Writing the Electronic Book”. Morgan & Claypool Publishers.

- 1st Generation: Hypermedia
- 2nd Generation: ebook hardware (term start to be used, around 1998 and reading software that generally aimed at providing functionality that would allow the user to annotate, bookmark, look up words, search within the text, and interact with the material in basic ways. But:

Reading from digital devices . . . suffers in several areas among them light, resolution, speed, and impact on the reader and there has been essentially no improvement in any of these areas in the last five years. . . Its just too hard to read from a computer, and it doesnt seem likely to get a lot easier (Crawford 1998).

But:

Thus the second generation of electronic books rolled in with considerable fanfare and re- ceded much more quietly, all but disappearing from the record.<sup>7</sup> All the

while, a revolution was taking place unnoticed on the sidelines as many people simply started reading newspapers, magazines, blogs, and other born-digital content on the screen, ignoring eBooks, and making do with whatever hardware they had in hand.

**3.5.2. Third Generation** Take advantage of a convergence of enabling technologies (low-power bistable displays, capacious storage, lighter hardware, the availability of digital content, and ubiquitous wireless) with warming consumer attitudes toward reading on the screen and purchasing unbundled content (using an iPod-like model of what a reading device might be like).

Think of portable personal digital libraries, not portable electronic books, as the future role of these appliances (Lynch 2001)

**3.5.3. Why aren't ebooks more popular** The arguments seem to fall into three categories:

- Tactile loss : I like the feel and smell of paper / I can't read on a screen
- Lending loss : I want to lend my books to who I want
- Library loss : I want a library that will last forever

The one big difference between the music and ebook markets is there is no “mp3” of ebooks there is no universal ebook file format.

### 3.5.4. Friday 20th of May 2011 News

MEDIA

Consiglia 548

## Kindle batte il libro di carta su Amazon vince l'e-book

Ogni 100 volumi cartacei, il sito vende 105 versioni digitali. Un sorpasso editoriale completo, l'ultimo di una serie, che segna il successo delle pubblicazioni elettroniche. E dei dispositivi per leggerle. Bezos: "Non ce lo aspettavamo così in fretta"



**ROMA** - Questa volta ci siamo. Dopo annunci più o meno verosimili sul sorpasso del libro elettronico sui volumi di carta, ora i numeri ci sono tutti. Non ci sono più distinzioni tra edizioni pregiate ed economiche ad influire sul computo come in passato, e Amazon comunica ufficialmente che gli e-book sorpassano ormai nelle vendite i libri tradizionali, e senza considerare contenuti gratuiti o promozionali. Il rapporto è 105 a 100 per titolo, uno scarto ancora leggero, ma di grande significato.

Era destino che accadesse, complice anche il calo del prezzo del Kindle, il lettore di e-book di Amazon, e l'introduzione di una versione ancora più economica, supportata dalla pubblicità. Già lo scorso luglio le vendite di libri per Kindle avevano superato le vendite di libri con copertina rigida, ora hanno superato anche quelle delle edizioni economiche, le "paperback".

Jeff Bezos, fondatore e amministratore delegato di Amazon, commenta il successo dei libri elettronici e sembra quasi stupito dal sorpasso sulle edizioni cartacee: "I clienti preferiscono i libri per Kindle ai libri tradizionali.

Pensavo che ciò potesse avvenire rapidamente, ma non ci aspettavamo che potesse accadere così presto". Amazon

vende libri di carta online da 15 anni, e versioni elettroniche da meno di quattro. Ma negli ultimi tempi, la

**3.5.5. Google and E-books** <http://books.google.com/help/ebooks/overview.html>

## 3.6. Book

<http://video.repubblica.it/tecno-e-scienze/signore-e-signori-vi-presentia-52721/52001>