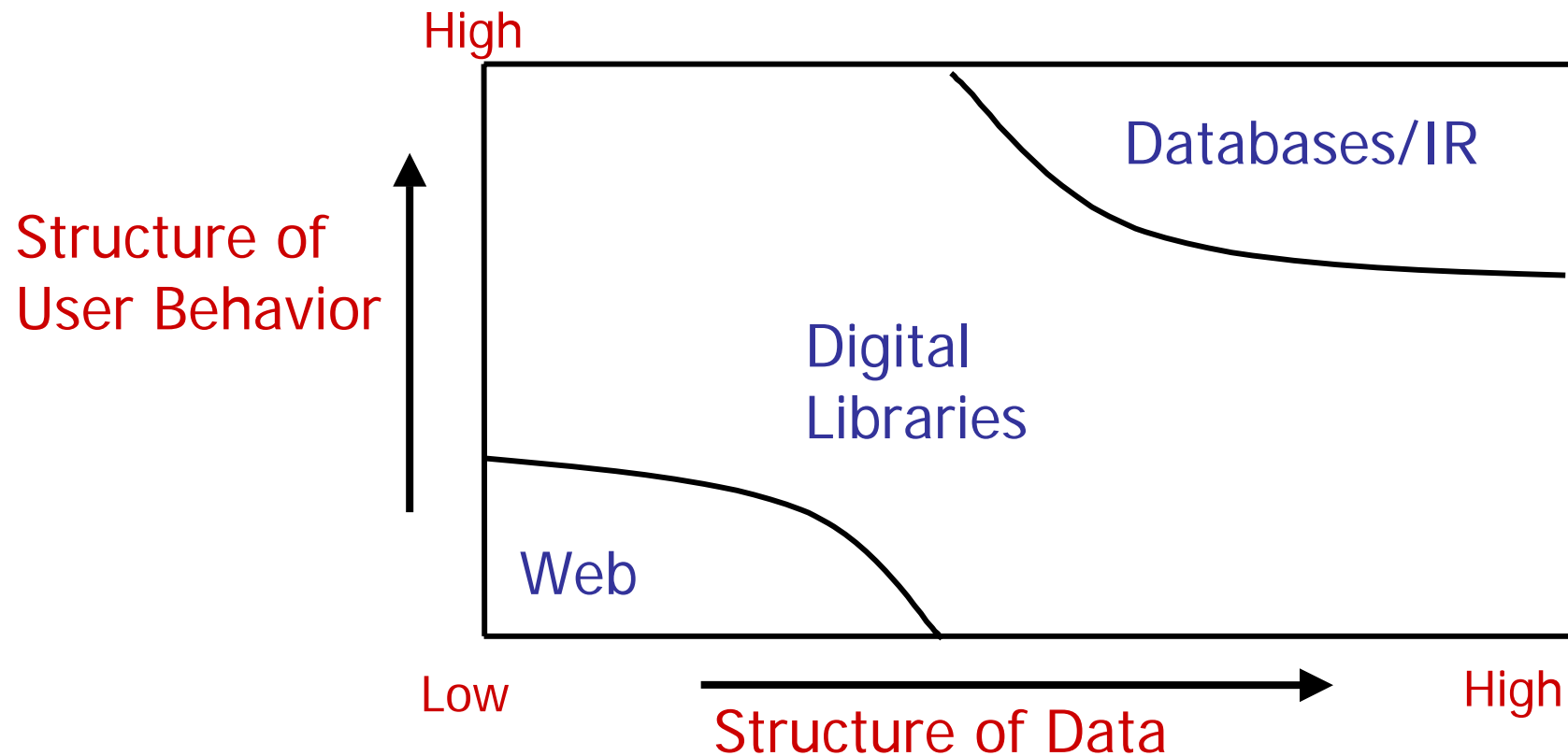


The Web versus the Digital Library

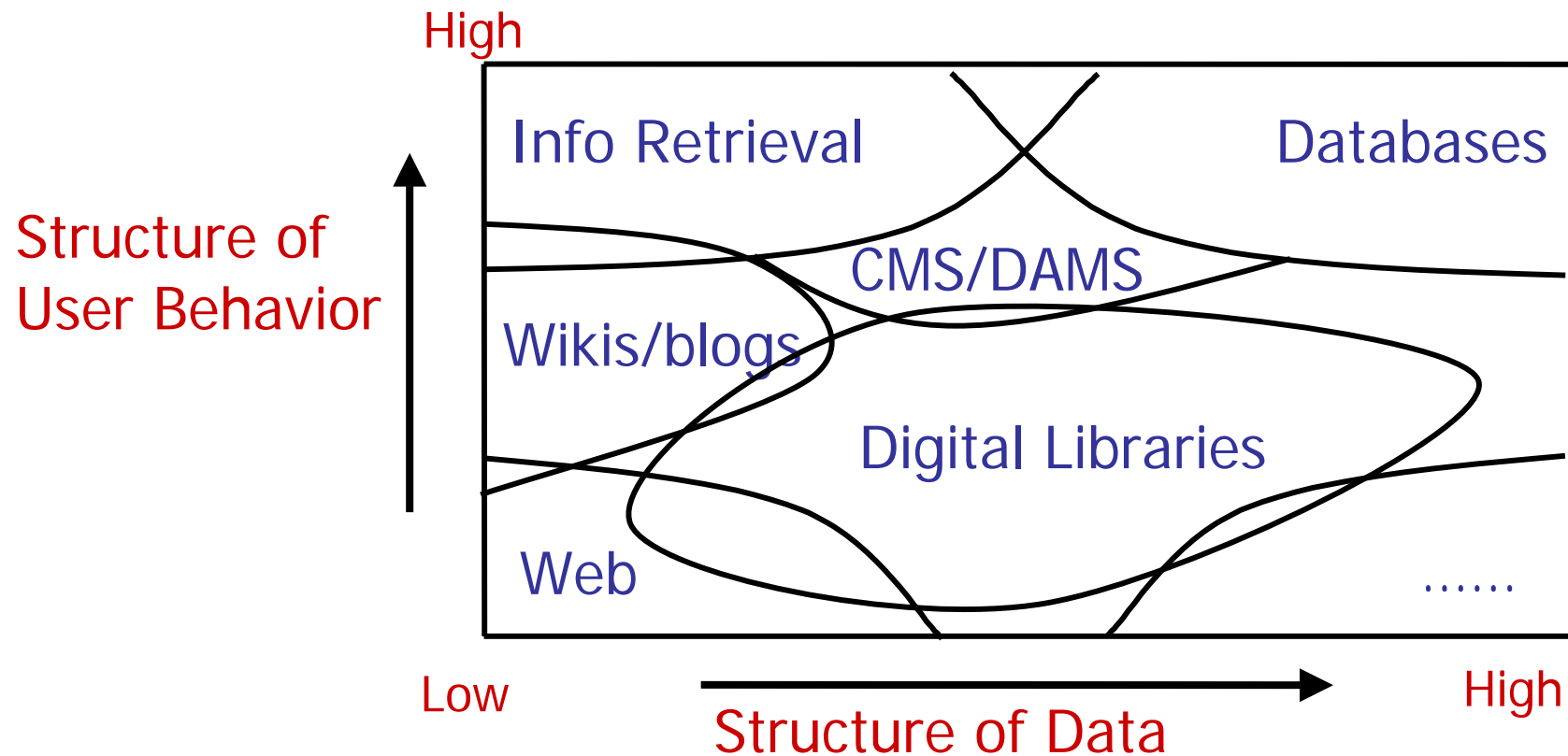
Vittore Casarosa
University of Parma and
ISTI-CNR, Pisa, Italy

NMPLIS Summer School
5-16 July 2010, Tbilisi, Georgia

Digital Libraries in the Information Space



Digital Libraries in the Information Space



Traditional role of libraries

- Mediators between information and users
- Selection
 - Definition of collections
- Acquisition
 - Physical objects
- Description
 - Catalogs
- Access
 - Shelves
- Preservation
 - Controlled environment

digital

Google

crawlers, spiders, bots, etc.

Dublin Core

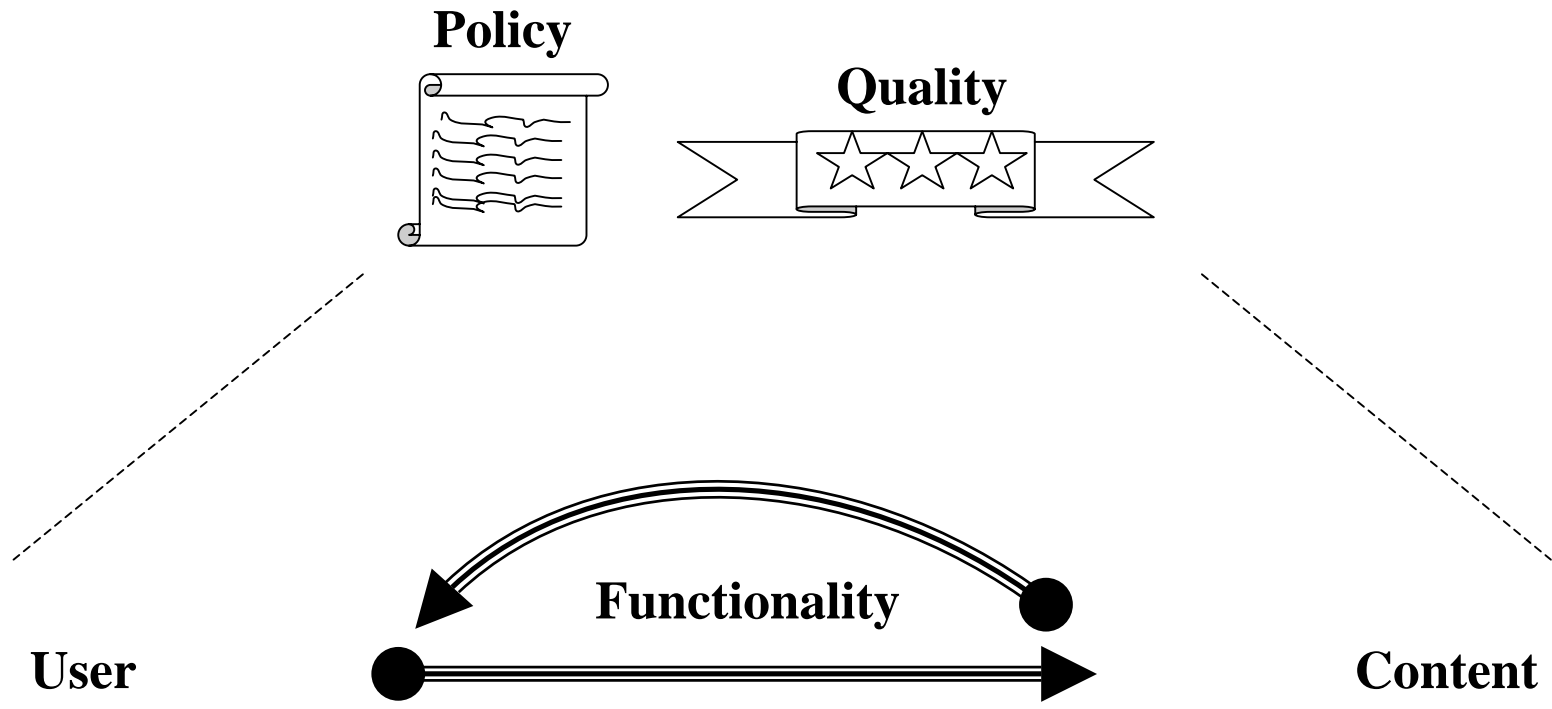
the Web

a) forever
b) the next five years
whichever comes first

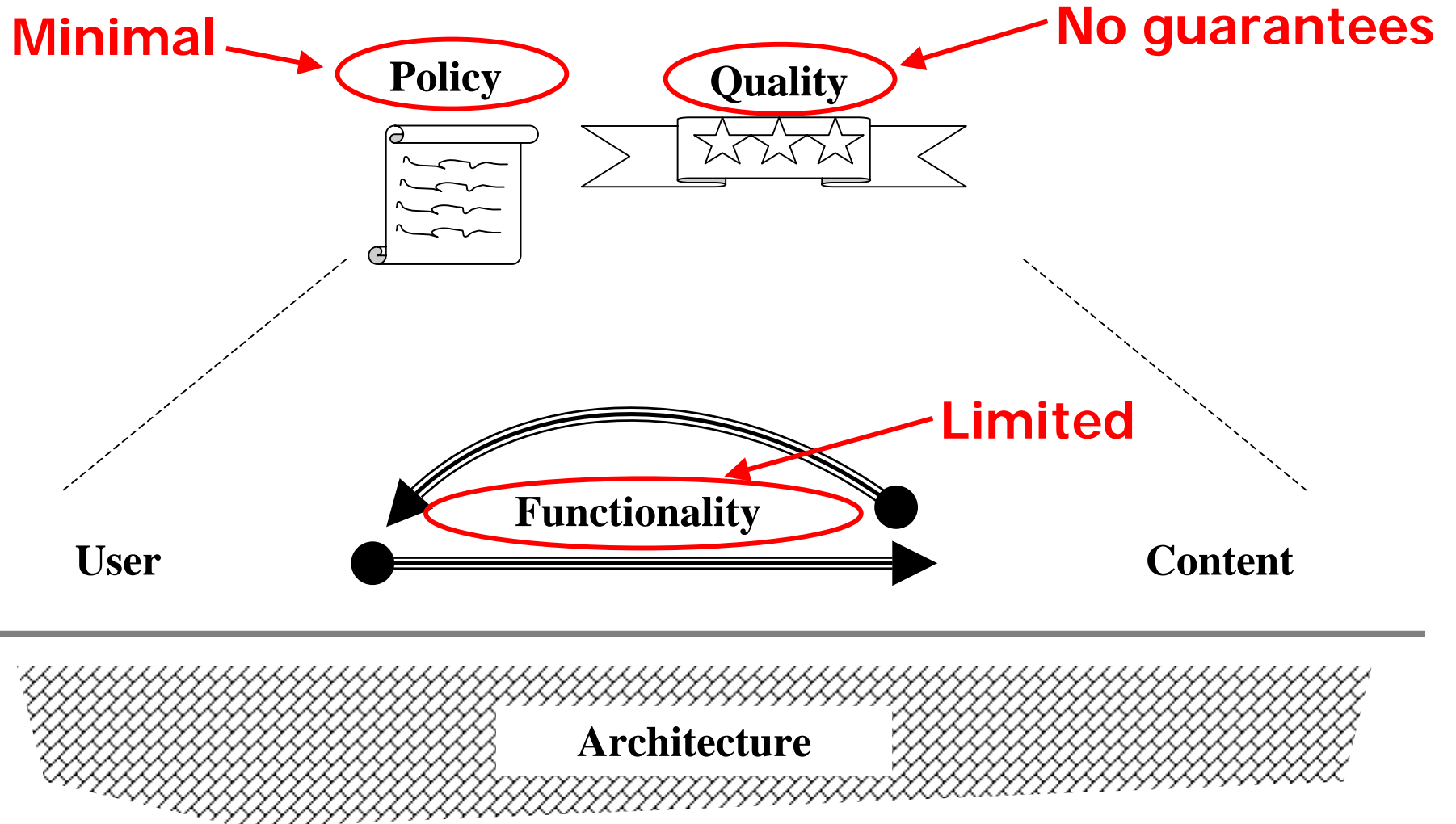
The Web vs. Digital Libraries

Opinion n. 1

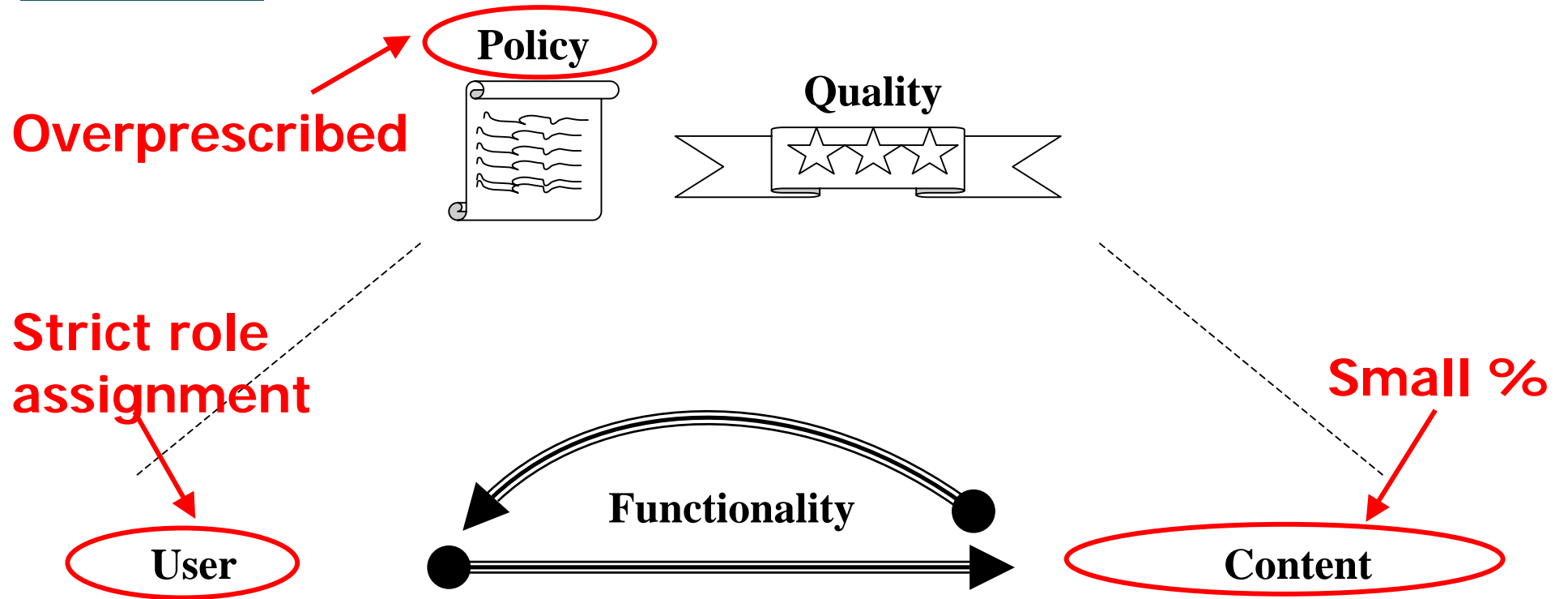
Basic concepts of an Info Mgmt System



The Web as an Info Mgmt System



Libraries as Info Mgmt Systems



No barriers to knowledge exchange

- More and more of the world's info/knowledge lives in specialized digital libraries that
 - Have content added/created by members of a community
 - Are curated by specialists of that library's topic
 - Are maintained by (designated members of) the community
- No strict separation between producer, curator, and consumer roles wrt which actors plays which
- Advanced services: annotation, personalization, contextualization, preservation, collaboration, etc.

Hopeful Conclusion: DLs vs. the Web

- They are not going to fight or replace each other, but in the end they are going to complement each other.

Hm!

Yes!

Probably!

**With lots of
new technology help
for both**

The Web vs. Digital Libraries

Opinion n. 2

Will the Web be the Digital Library ? Library ?

- Use(fullness)
- Structure
- Content

Use(fullness):

Visits to libraries:

- Approx 10 / year
- Of which “private/touristic” visits: approx 10 / year
- Visits to Digital Libraries:
 - Several times per week
- Visits to the Web:
 - Several times per day
 - Usually finding what is needed, specifically for professional purposes
- People increasingly use simple Google-style interfaces for search rather than more complex DL-like interfaces, not to mention library catalogues

Structure (Information and Metadata)

Libraries:

- very high, professional
- but inflexible: one sorting, rest in (digital) catalogues
searching these is actually like a DL without content
- Digital Libraries:
 - Very high, manually cared for, sometimes community-driven
 - Flexible ways of interaction
- Web:
 - None, but:
 - Increasingly created automatically: CiteSeer, DBWorld, GoogleBooks, Genre Classification, Topic analysis, Named Entity Detection,...

Content:

- Libraries:
 - High-quality, selected content
 - but increasingly: not the content I need: proceedings, papers are first/only in DL before they make it to the library catalogue
- Digital Libraries:
 - Sometimes the only place where I can find certain material (sufficiently easily)
 - Increasingly also older content digitized
- Web:
 - Sometimes even more comprehensive than the DLs
 - What's not on-line does not exist (cf. CiteSeer)
 - Increasing amounts of traditional, high-quality content on the Web (or in a DL?) (e.g. Internet Archive)

So:

Looks like (parts of) the Web will turn into a

- Digital Library, which will eventually replace the
 - Conventional Library
- Requiring only
 - More content (will come)
 - More structure (will be provided by better computer programs), which will lead to
 - More users using this, assisted by better interfaces

The Web vs. Digital Libraries

Opinion n. 3

Google's Mission

Google's mission is to **organize** the **world's information** and make it **universally accessible** and **useful**.

Organize

- By vertical/property: Scholar, Book Search, Product Search, News, Maps, etc
- By search

World's information

- What we can reach through the web
- What we license

Universally accessible

- Via internet
- Internationalized and localized

Useful: focussing on and meeting our users' needs

**Is Google(*) the world's
Digital Library?**

*** put here your favourite
search engine**

A Digital Library's Mission

A Digital Library 's mission is, for a **selected user community**, to **organize** that **community's information** and make it **universally accessible** and **useful** to that community.

Organize

- According to the needs of the user community (art, photographs, scientific data, ...)

Community's information

- Information (including data) generated by the community
- That can be reached through the web
- That can be licensed (or purchased)

(Universally) accessible

- Via internet (including via web search)
- Internationalized and localized

Useful: focussing on and meeting users' needs within the selected user community

Major Differences in Missions

- Scale
- Information: broad versus deep coverage
- General versus specific communities (and therefore needs)
- Organizing principles (can be very different)
- Services provided: how we “add value” to information/data

Other considerations

- Profit
- Quality, conservation and preservation
- Authority

Conclusions (opinion 3)

- Web Search (Google) and Digital Libraries share **similar but complementary missions**
- Celebrate the **diversity of missions**, and **concentrate on strengths** whether as web search engine or digital library
 - Search engines: scale, universal delivery, universal services
 - Digital libraries: specialized collections, specialized services, “library” services
- Focus on **delivering value to users** through useful and relevant (web) services (“Focus on the user and all else will follow”)
- **Web search** is a service that **Digital Libraries** should exploit to ensure **universal access** to information and services