

# Wikisource

Where we are

Where we want to go

Andrea Zanni  
Wikimedia Italia

Wikimania 2012



# The library of Babel

*“The universe (which others call the library)...”*

J. L. Borges

[http://en.wikipedia.org/wiki/The\\_Library\\_of\\_Babel](http://en.wikipedia.org/wiki/The_Library_of_Babel)



# What are digital libraries?

Nobody really knows, but many agree on requirements:

- 1. Collection**
- 2. Metadata**
- 3. Services**
- 4. People**



# Collection

- Reliability
- Readability
- Curation
- Quality



# It's a kind of magic..

Metadata are used for

- cataloging
- indexing
- retrieving
- archiving
- communicating
- preserving information.

If we want to deal with information, we need **metadata**.



# It's a kind of magic..

Metadata are used for

- cataloging
- indexing
- retrieving
- archiving
- communicating
- preserving information.

If we want to deal with information, we need **metadata**.



# Metadata

On Wikisource metadata contains information about books and authors

- in simple text
- human-readable
- no standard
- not interoperable

*... no magic :-)*



# Services

Everything that is beyond books:

- reference
- (all kind of) categories
- lists
- links
- context
- disambiguation
- redirects





# People

Librarians (and users) form the community (we are not Google books!)

- curation → books, project, policies
- empowerment → from users to librarians



# 5th law of Library Science

*“The library is a growing organism”*

S. R. Ranganathan

[http://en.wikipedia.org/wiki/Five\\_laws\\_of\\_library\\_science](http://en.wikipedia.org/wiki/Five_laws_of_library_science)



# Hyperlibrary: Xanadu 0.1

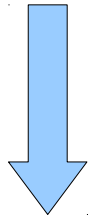
*“Can you imagine that they used to have libraries where the books didn't talk to each other?”*

Marvin Minsky  
[citation needed]



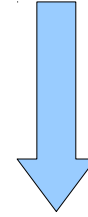
# “Collaboratory”

read



library

write



laboratory

- Tools → framework → other tools
- MediaWiki, js, templates, python, bot, API, toolserver, ...



# The future



# Interoperability

- Bibliographic data from OCLC, Open Library, catalogs.
- Disseminate metadata and full text (OAI-PMH)
- Wikisource API



# ePub

Fresh generated ePub on the rocks (via ePub converter)

- outreach
- eReader apps



# Classification

Potential of MediaWiki categories:

- Colon classification
- subjects from National Libraries
- thesauri and ontologies





# Microcontribution

*“the more simple and small task is, the wider audience you get”*

- Citizen science (Galaxy Zoo, Ancient Lives)
- from *page* unit to *word* unit

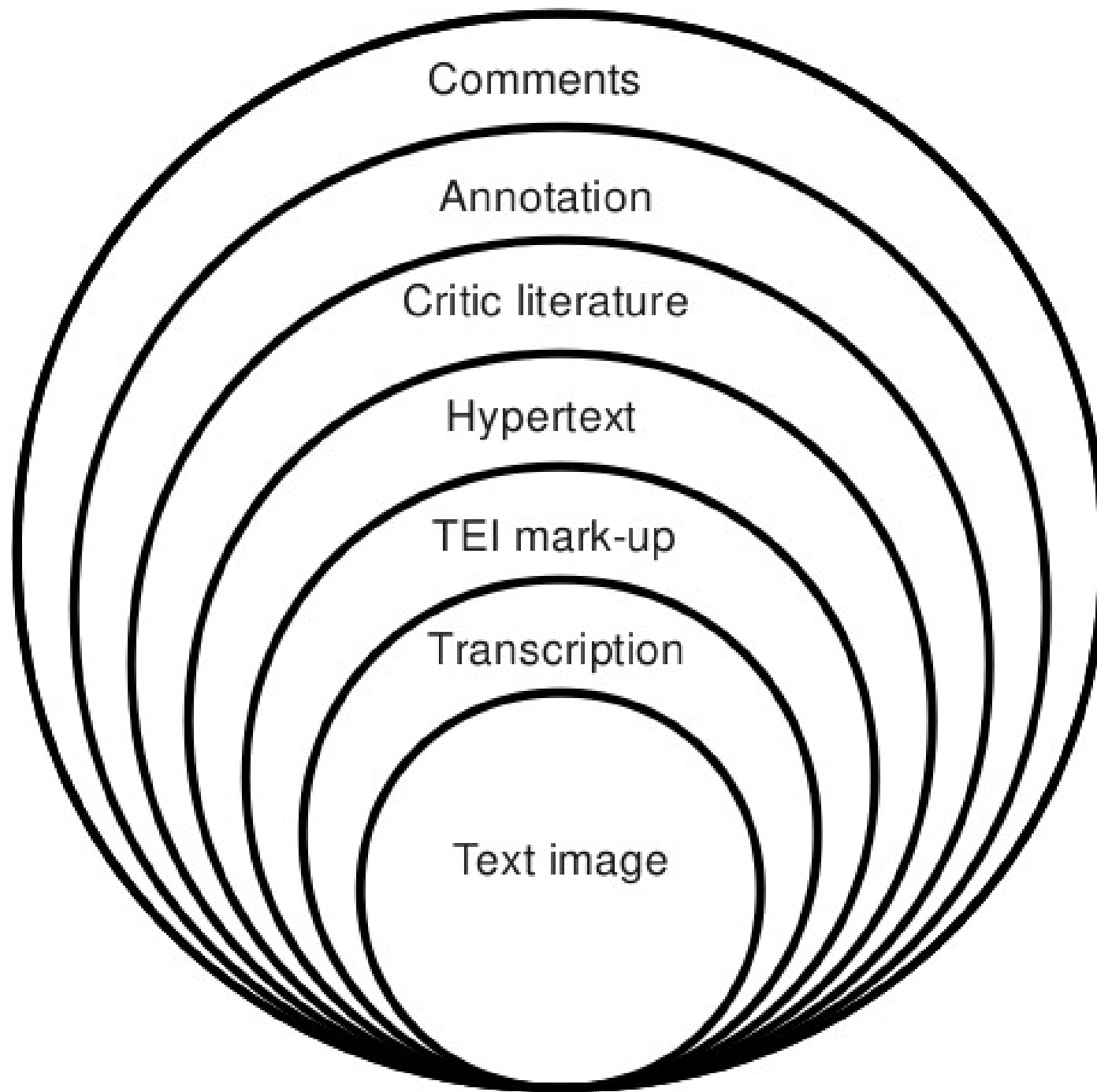
**More:** WikiCaptcha (next presentation same room!)



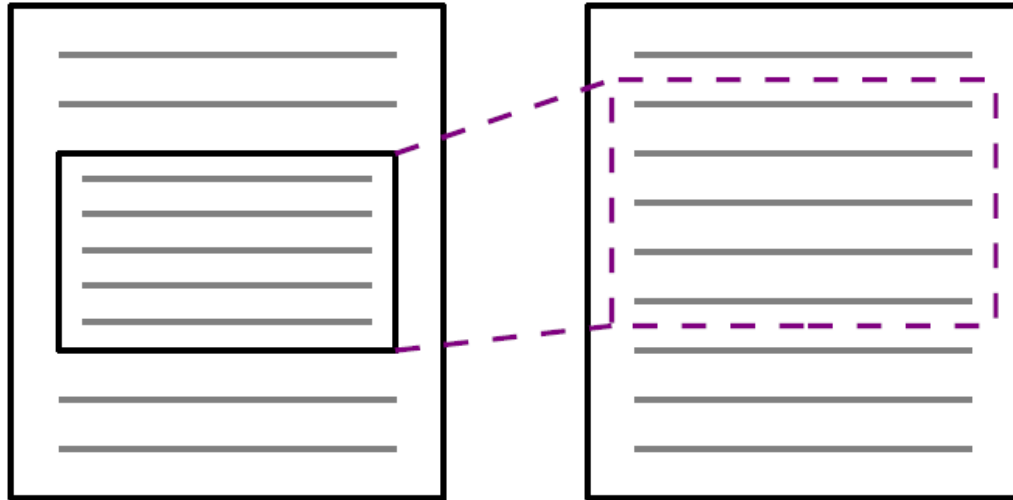
# New architecture on djvu

- in-line transcription
- high granularity
- save text directly on djvu
- multiple layers





# Xanadu 0.2



Systematic use of transclusion

- Interwiki
- Wikiquote
- Wikipedia
- Blogs, websites, etc.

# Born-digital documents process

No specific process: must pass through the whole process for digitized files

- 1) Djvu
- 2) OCR
- 3) Commons
- 4) Transcription

Collaboration with repositories and digital libraries (scientific articles, thesis, free documents).

# Feedbacks?

- Email: [andrea.zanni@wikimedia.it](mailto:andrea.zanni@wikimedia.it)
- Nickname: Aubrey
- Skype: aubreymcfato