Herbert Van de Sompel
Cornell University
Computer Science – Digital Library Group

CNI FALL 2000 – San Antonio, Texas – December 8th 2000
Closing Keynote Address
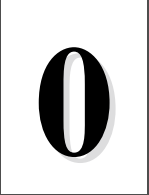
herbert van de sompel

Concrete part   Preview of the Open Archives Metadata Harvesting protocol

Speculative part   Back to the preprint roots of the OAI:
- What does the OAMH protocol mean?
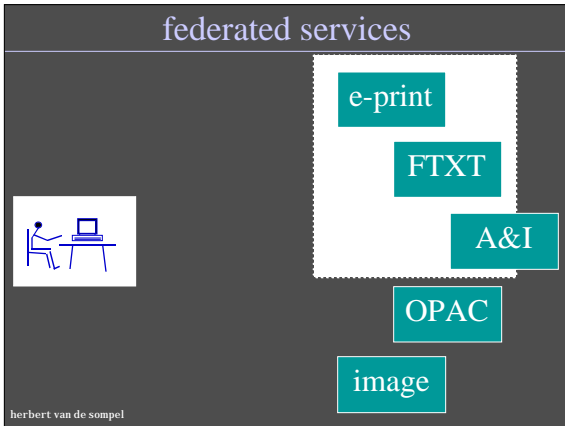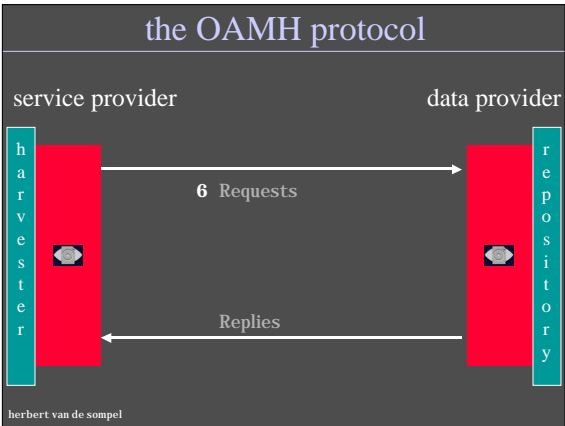- What lies beyond the OAMH protocol?

herbert van de sompel

0

the concrete part

herbert van de sompel
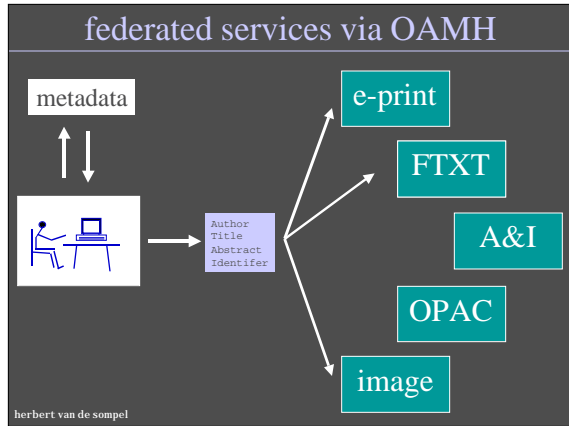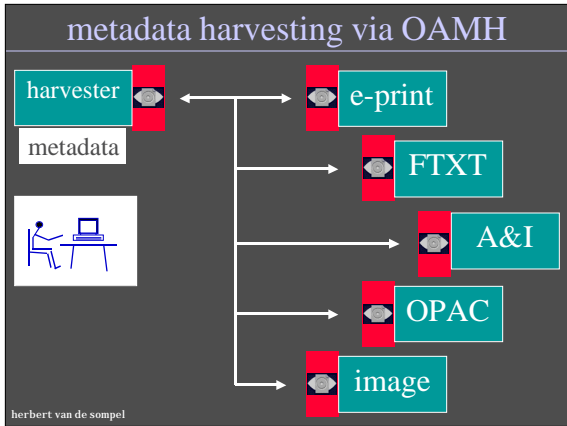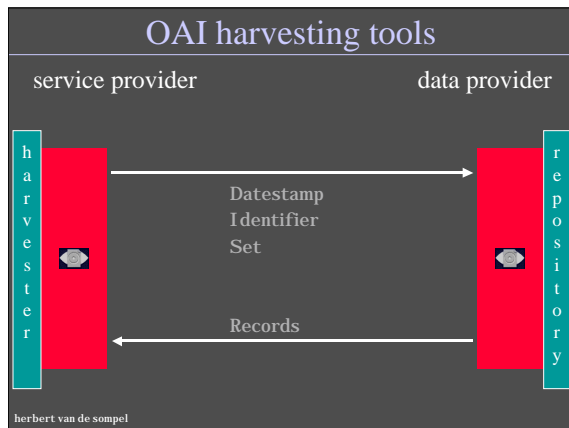

The OAMH protocol is a low-barrier interoperability specification for the recurrent exchange of metadata between systems

herbert van de sompel

---

the OAMH protocol

service provider   data provider

harvester   **6** Requests   repository

Replies

herbert van de sompel

---

federated services

e-print

FTXT

A&I

OPAC

image

herbert van de sompel

metadata harvesting via OAMH

| harvester | ⬧ | ⬧ | e-print |

metadata

⬧ FTXT

⬧ A&I

⬧ OPAC

⬧ image

herbert van de sompel

federated services via OAMH

metadata

e-print

FTXT

Author
Title
Abstract
Identifer

A&I

OPAC

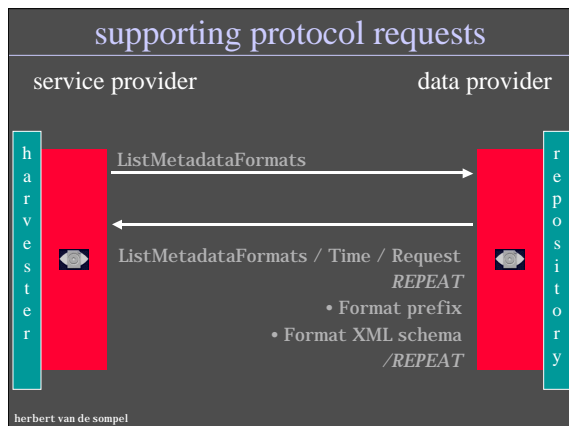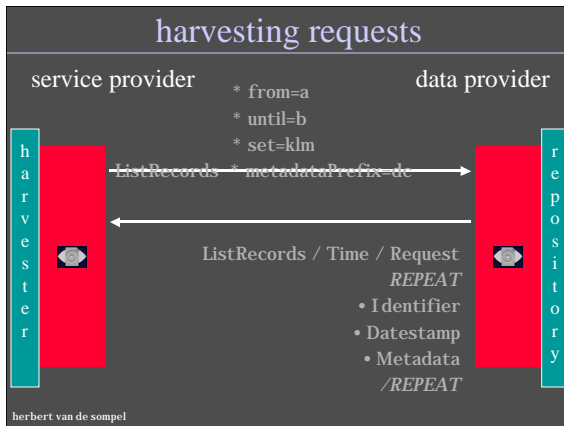image

herbert van de sompel

Core concepts in OAMH

- low-barrier interoperability
- data-provider & service-provider model
- metadata harvesting model   OAMH protocol
                                 HTTP based
                                 Reply • XML Schema

                                       • Self contained
- shared metadata format and parallel, community-
specific metadata formats   Dublin Core

- acceptable use   Community specific

herbert van de sompel

OAI harvesting tools

service provider   data provider

harvester ⬧   repository

Datestamp
Identifier
Set

Records

herbert van de sompel

OAI harvesting tools

service provider   data provider

harvester ⬧   Supporting protocol requests: repository
                • Identify
                • ListMetadataFormats
                • ListSets
                Harvesting protocol requests:
                • ListRecords
                • ListIdentifiers
                • GetRecord

herbert van de sompel

supporting protocol requests

service provider   data provider

harvester ⬧   ListMetadataFormats   repository

                ListMetadataFormats / Time / Request
                                        REPEAT
                                • Format prefix
                                • Format XML schema
                                        /REPEAT

herbert van de sompel

harvesting requests

service provider   data provider

h a r v e s t e r

r e p o s i t o r y

* from=a
* until=b
* set=klm
ListRecords * metadataPrefix=dc

ListRecords / Time / Request
REPEAT
• Identifier
• Datestamp
• Metadata
/REPEAT

herbert van de sompel

Applications of the OAMH protocol?

• federated services [S&R, SDI, alerting, linking, …]
• database synchronization
• harvesting the deep Web
• …

herbert van de sompel

1

the speculative part

herbert van de sompel

OAI roots: advance interoperability of preprints as a means to promote their global acceptance

• What does the OAMH protocol mean in the preprint context?
• Are there any steps to be taken beyond the OAMH protocol in the preprint context?

herbert van de sompel

What follows:

• Scholarly communication systems
• The paper and PDF implementation
• The attractiveness of preprints
• An interoperable, decomposed scholarly communication system built around preprints?

herbert van de sompel

Market of scholarly communication {Roosendaal & Geurts}

registration	claiming a new finding
certification	certifying the claim
awareness	ensure information throughput
archiving	preserving the heritage

herbert van de sompel

Market of scholarly communication {Roosendaal & Geurts}

registration	claiming a new finding
certification	certifying the claim
awareness	ensure information throughput
archiving	preserving the heritage
rewarding	*evaluating & rewarding performance*
accessibility	*availibility and searchability*

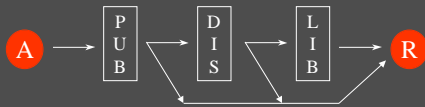herbert van de sompel

The paper version {the journal system}

registration	publisher
certification	publisher
awareness	library [selection, service, support]
archiving	library

A → PUB → DIS → LIB → R

herbert van de sompel

The PDF version {the journal system}

registration	publisher
certification	publisher
awareness	publisher, distributor, library
archiving	

A → PUB → DIS → LIB → R

herbert van de sompel

{the journal system}

registration	publisher
certification	publisher
awareness	
archiving	
rewarding	
accessibility	

herbert van de sompel

{the journal system}

registration	publisher
certification	publisher
awareness	
archiving	
rewarding	
accessibility	*low availibility [serials crisis]*

herbert van de sompel

{the journal system}

registration	publisher
certification	publisher
awareness	
archiving	
rewarding	*citation databases*
accessibility	*low availibility [serials crisis]*

herbert van de sompel

Slide 1

{the journal system}

registration	publisher
certification	publisher
awareness	publisher, distributor, library
archiving	
rewarding	*citation databases*
accessibility	*low availibility [serials crisis]*

herbert van de sompel

Slide 2

{the journal system}

registration	publisher
certification	publisher
awareness	publisher, distributor, library
archiving	???
rewarding	*citation databases*
accessibility	*low availibility [serials crisis]*

herbert van de sompel

Slide 3

It is -- at least -- legitimate to reflect on the possibility of a digital system for scholarly communication that is not merely a scanned copy of the paper system.

In order to free our minds: let's forget about who has been doing what and how in the existing system.

herbert van de sompel

Slide 4

A preprint in a digital scholarly communication system

registration	yes
certification	no
awareness	yes
archiving	no
rewarding	*no*
accessibility	*high availibility*

herbert van de sompel

Slide 5

A preprint in a digital scholarly communication system

registration	yes
certification	no
awareness	yes
archiving	no
rewarding	*no*
accessibility	*high availibility*

herbert van de sompel

Slide 6

Registration via preprints suggests the feasibility of a decomposed system for scholarly communication.

herbert van de sompel

awareness | certification | *rewarding*

A

R

registration | archiving

herbert van de sompel

Communication via preprints suggests the possibility of preprints being the starting point of a new value chain in which the raw material -- the non-certified preprint -- is in open access.

herbert van de sompel

The Innovator's Dilemma {Christensen}

• sustaining versus disruptive technologies;
• disruptive technologies:
  • somehow perform worse than established ones
  • not accepted by core customer base
  • but: convenient, cheap, ...
• disruptive technologies can create competition in an existing value network by creating a new one first.
=> preprints as a disruptive technology
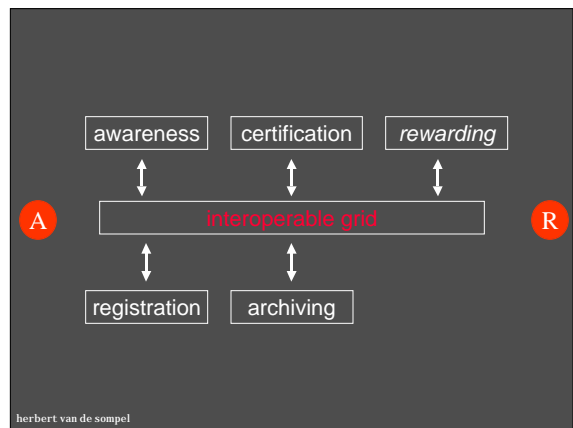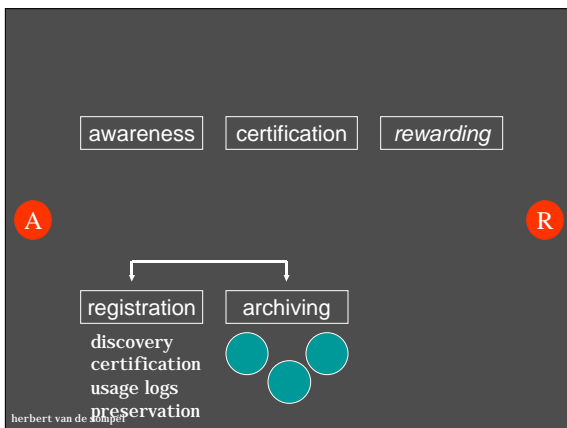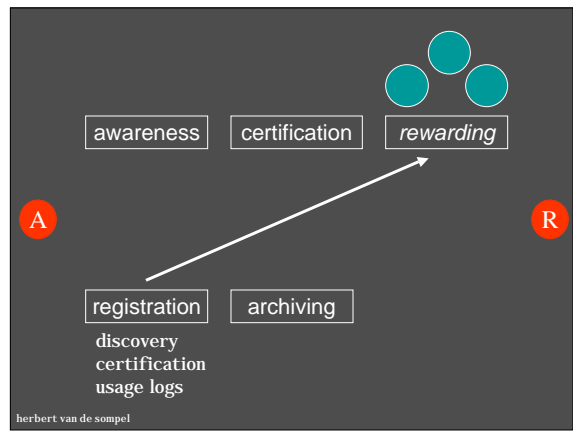
herbert van de sompel

---

awareness | certification | *rewarding*

A → new value chain → R

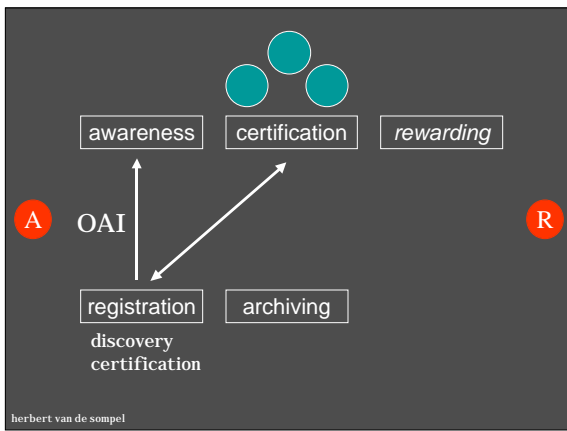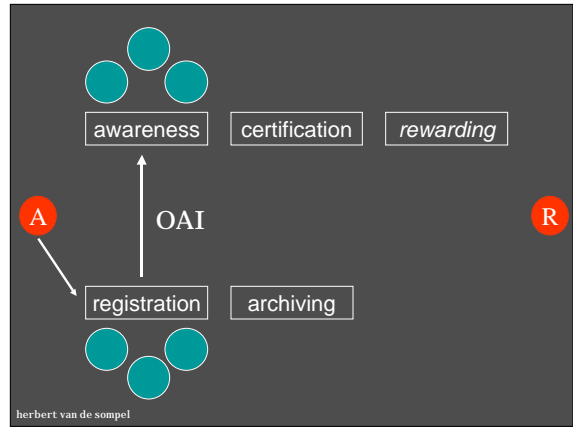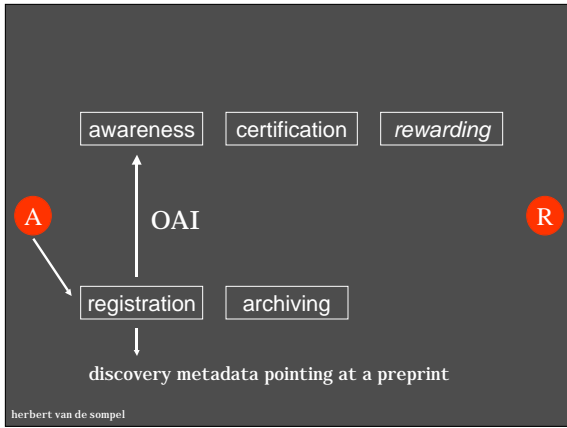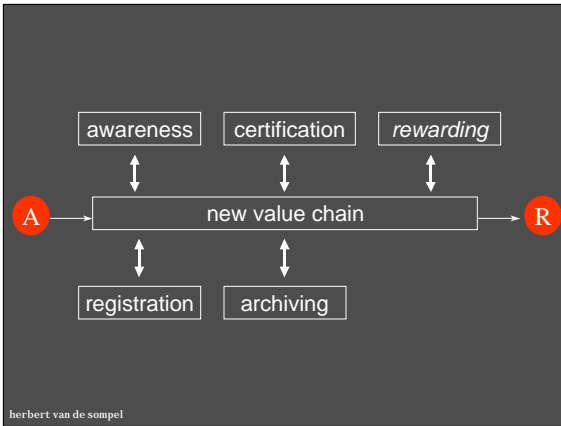registration | archiving

herbert van de sompel
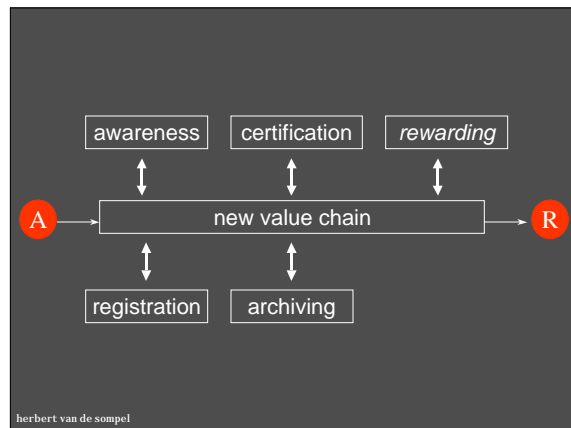
This urges us to think about:

• how the functions of a scholarly communication system can be implemented in a digital environment

• how the functions of a scholarly communication system can interoperate disregarding their implementation

herbert van de sompel

---

economy

establishing a technological basis that

technology ← sociology

allows addressing the other issues.

law

herbert van de sompel

6

7

awareness   certification   *rewarding*

A → new value chain → R

registration   archiving

herbert van de sompel
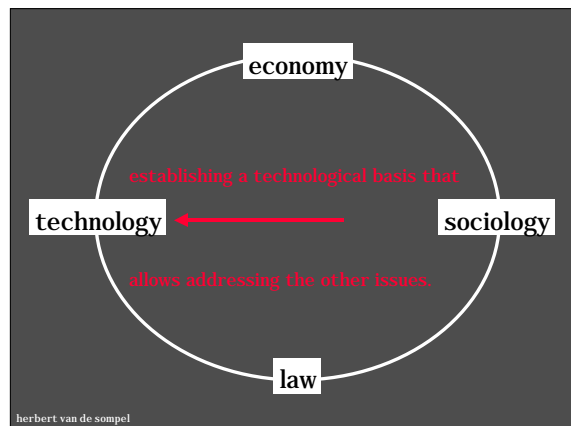
So what about the library?
[back to the who does what
question]

herbert van de sompel

Within the existing system, Libraries are
trying hard to optimize the output of a
system with far from optimal input.

herbert van de sompel

It has become increasingly difficult for
Libraries to fullfill their fundamental role of
safeguarding equity of access.

herbert van de sompel

In the PDF version of the information chain,
Libraries are aggregating the aggregators.

That is a lot of aggregating for a digital world.

herbert van de sompel

At the core of the problems that Libraries are
facing is the total dependancy on information
held upstream in the information chain.

herbert van de sompel

As such, there are numerous incentives for Libraries:
• to rethink themselves
• to be pro-active in exploring alternative mechanisms for scholarly communication

herbert van de sompel

Libraries: the good news

• Libraries are close to authors:
  • a great position to fullfill the registration function i.e. obtain institutional material
  • a great position to archive institutional material
• Libraries are fast at embracing new technologies
• Libraries have very knowledgeable people
• Libraries provide a level of redundancy in services that is no longer required in a digital environment
• The Library as an institution that safeguardes equity of access has global representation

herbert van de sompel

Libraries: the bad news

• As organizations Libraries are slow movers, hosted by slowly moving institutions
• Libraries are slow to recognize the fact that a new technology may allow [or beg] for a new mode of operation
• The -- information -- world runs on Internet time

herbert van de sompel