



The Open Archives Initiative



Simeon Warner (Cornell University)
simeon@cs.cornell.edu

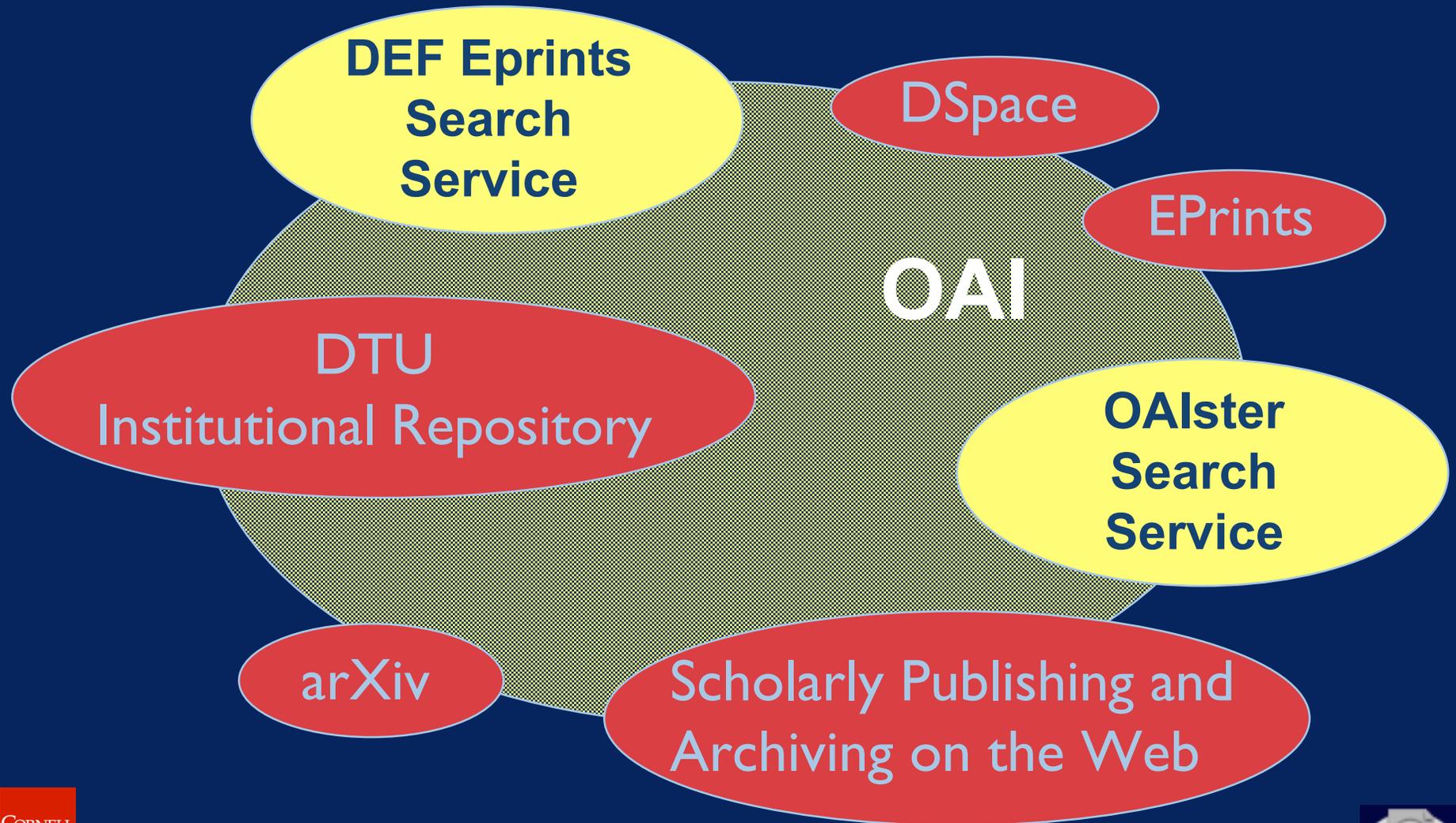


CORNELL

Open Archives seminar “Facilitating Free and Efficient Scientific Communication”, DEF/DTV/DTU, Copenhagen, Denmark, 18 February 2004.



Where does the OAI fit?



Origins of the OAI

"The Open Archives Initiative has been set up to create a forum to discuss and solve matters of interoperability between electronic preprint solutions, as a way to promote their global acceptance. "

(Paul Ginsparg, Rick Luce & Herbert Van de Sompel - 1999)



What is the OAI now?

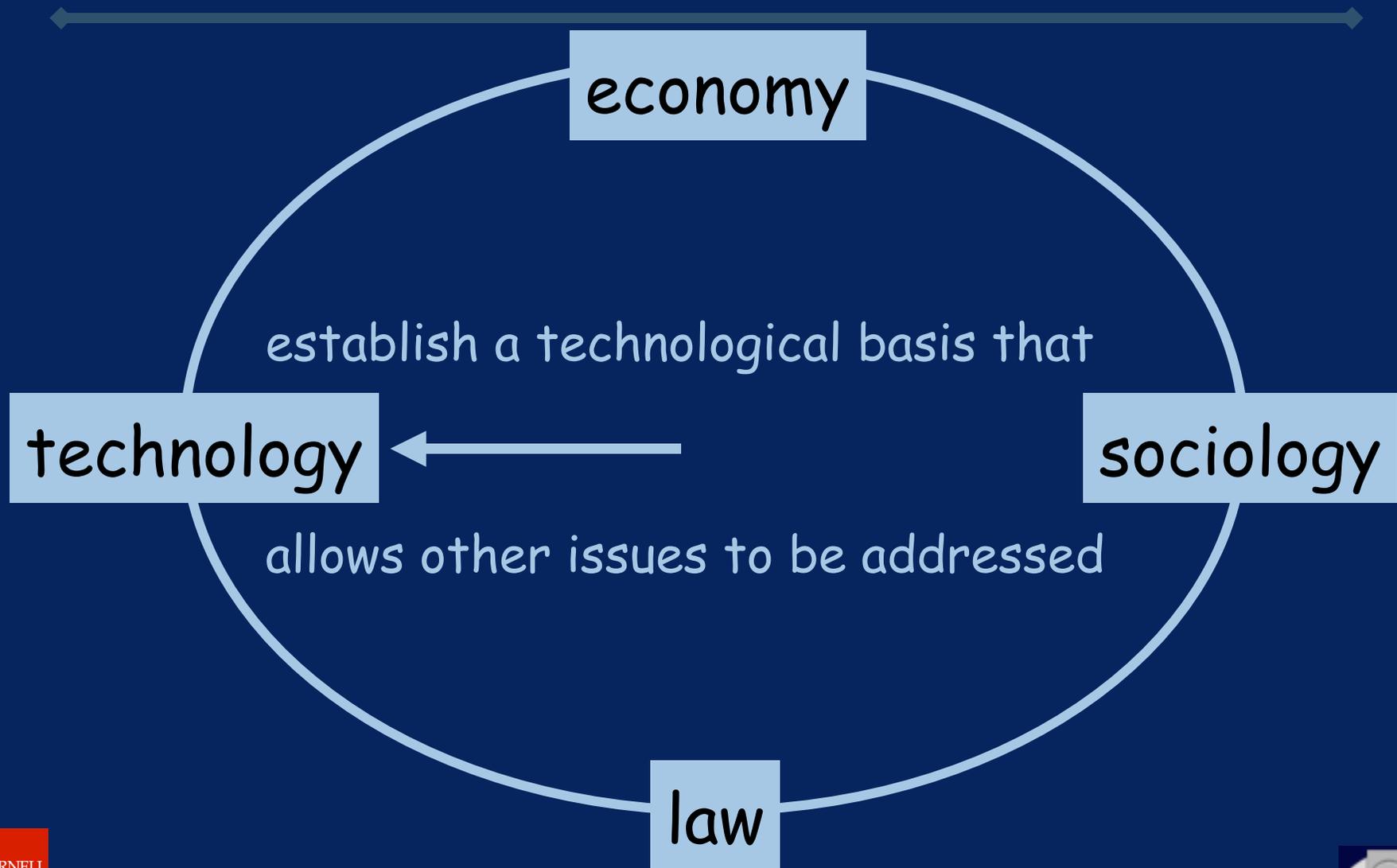
"The OAI develops and promotes interoperability standards that aim to facilitate the efficient dissemination of content." (from OAI mission statement)

- Technological framework around OAI-PMH protocol
- Application independent
- Independent of economic model for content

Also ... a community and a "brand"



OAI in context



OAI and Open Access

- There is "A" difference
 - Open **Archives** Initiative
 - Open **Access**
- The OAI is not tied to a particular political agenda - **technical focus**
- BUT... the OAI provides functionality that is essential for many Open Access proposals

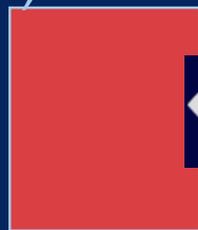


OAI Protocol for Metadata Harvesting

⇒ "OAI-PMH"

- Simple protocol. Free implementations available.
- Designed to allow harvesting of any XML metadata (schema described)

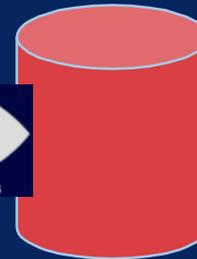
Service Provider
(Harvester)



Protocol requests



Data Provider
(Repository)



XML metadata

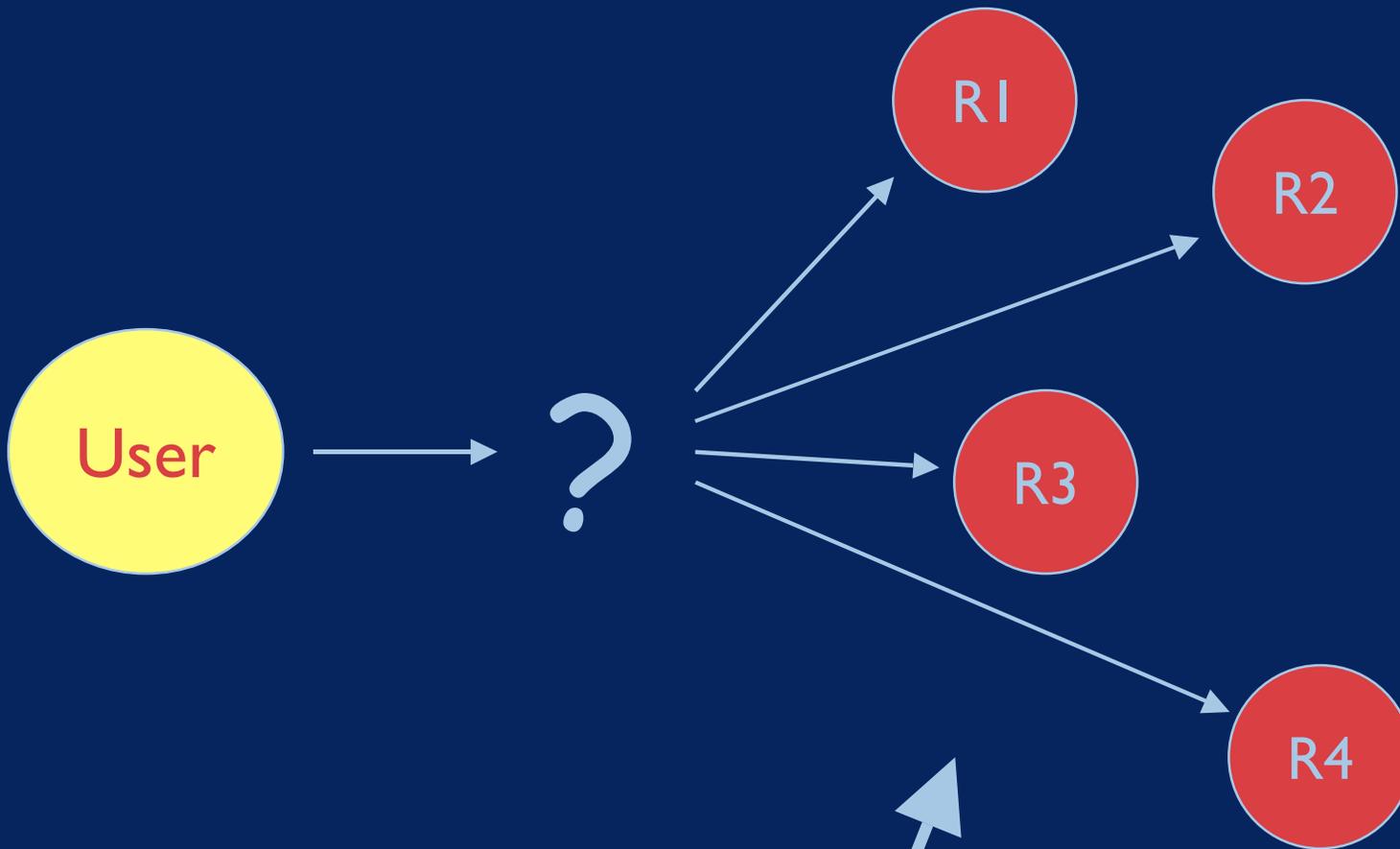


OAI-PMH - M is for Metadata

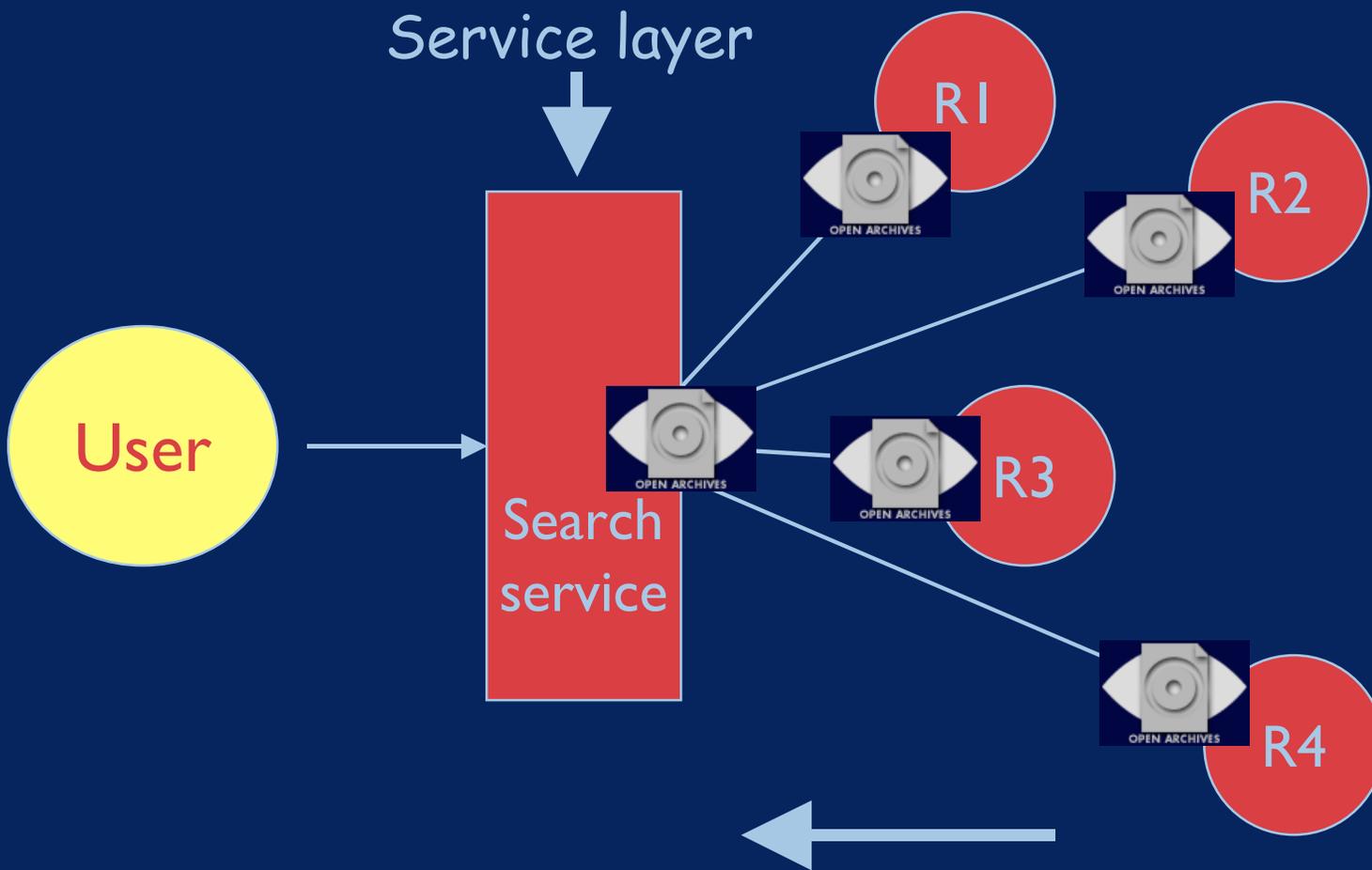
- Simple Dublin Core mandated for base level interoperability
- DC typically generated by automated cross-walk if base metadata is not DC
- Support for multiple metadata formats, e.g. expose MARC and DC for single item
- Strategy perhaps gone slightly awry, service providers complain about bad metadata and bad use of DC. Data providers should also expose original format



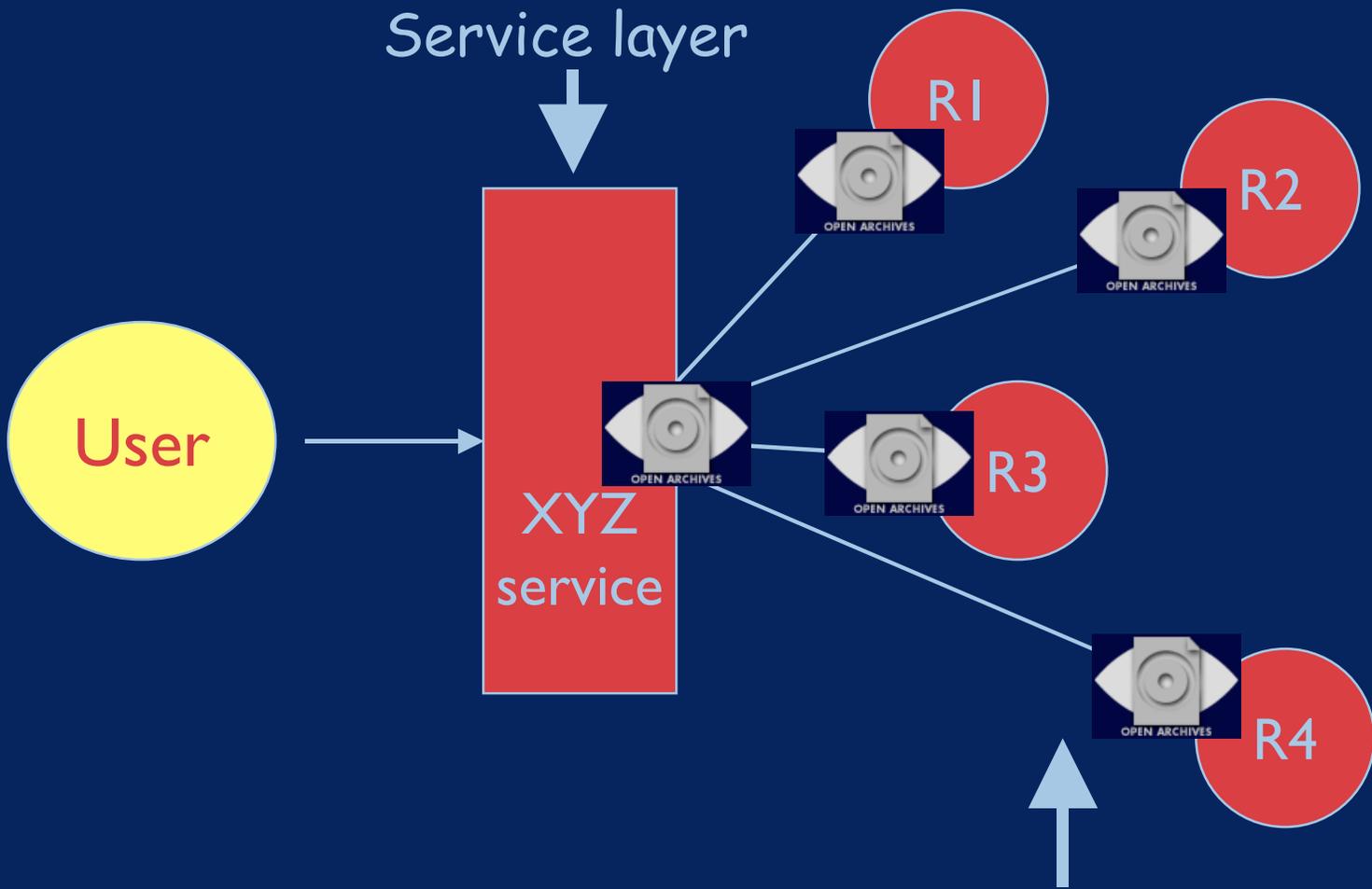
OAI for discovery



OAI for discovery



OAI for XYZ



Global network of resources exposing metadata ||



Building the network

- Services have to be able to locate repositories
 - Outside knowledge
 - Through OAI registry
 - Through <friends> data
- Still issues of selection and local collection building
- Network may contain intermediate aggregators and proxies

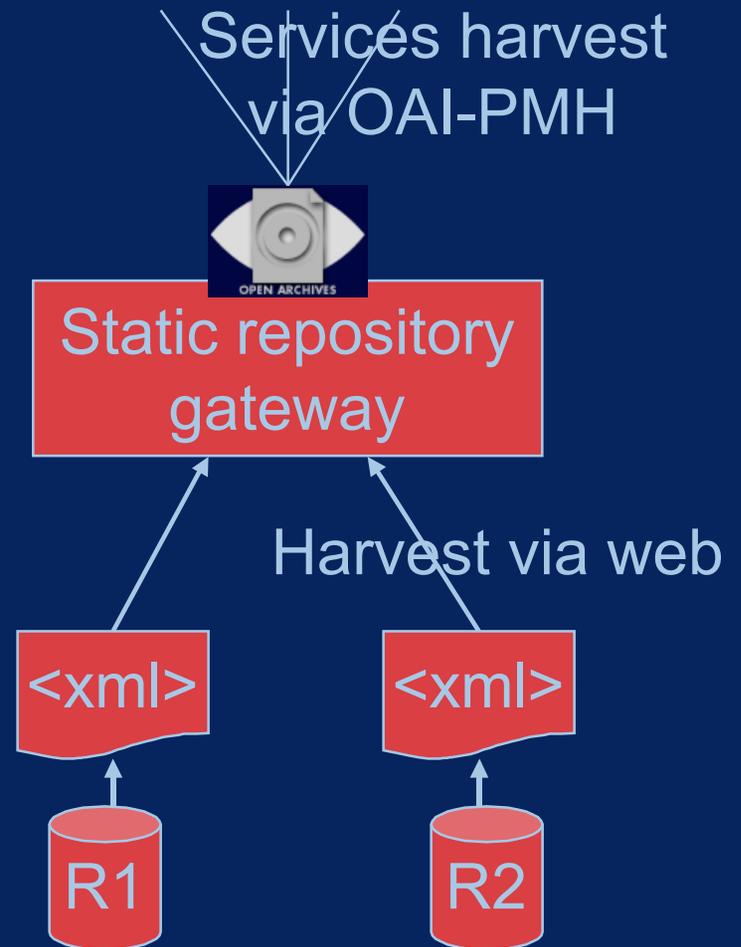


Too small to implement OAI-PMH?

e.g. a collection of
100 working papers

“Static repository”
version of OAI:

- Expose XML file on web server
- Register with gateway



Facilitating New Models of Scholarly Communication

The role of the OAI in Open Access models, institutional repositories and perhaps in disaggregated systems

CORNELL



Eprint archives

Eprint:

- Scholarly literature including journal articles, pre-prints, technical reports, books, theses and dissertations. May or may not be refereed.
- Open Access to full-content via Internet

Archives (metadata available via OAI):

- **arXiv** (aka xxx) eprint archive (260k eprints)
- **RePEc** (231k records, 6k eprints)
- **NASA NTRS** (368k records, 12k eprints)
- **NDLTD** (e.g. VTETD, 3.6k total, 2.4k eprints)
- **CERN Document Server** (41k eprints)
- **Organic eprints** (1.4k eprints)



Institutional repositories

- Institutionally defined: content generated by institutional community
- Scholarly content: preprints and working papers, published articles, enduring teaching materials, student theses, etc.
- Cumulative and perpetual: preserve ongoing access to material
- Open Access: free, online
- Interoperable?



Institutional repositories

registration	yes
certification	imprimatur of institution/department, other methods can be overlaid
awareness	via OAI interoperability framework - search and alerting services...
archiving	possible, library in control
<i>rewarding</i>	<i>new metrics? separate certification?</i>



Obstacles to implementation

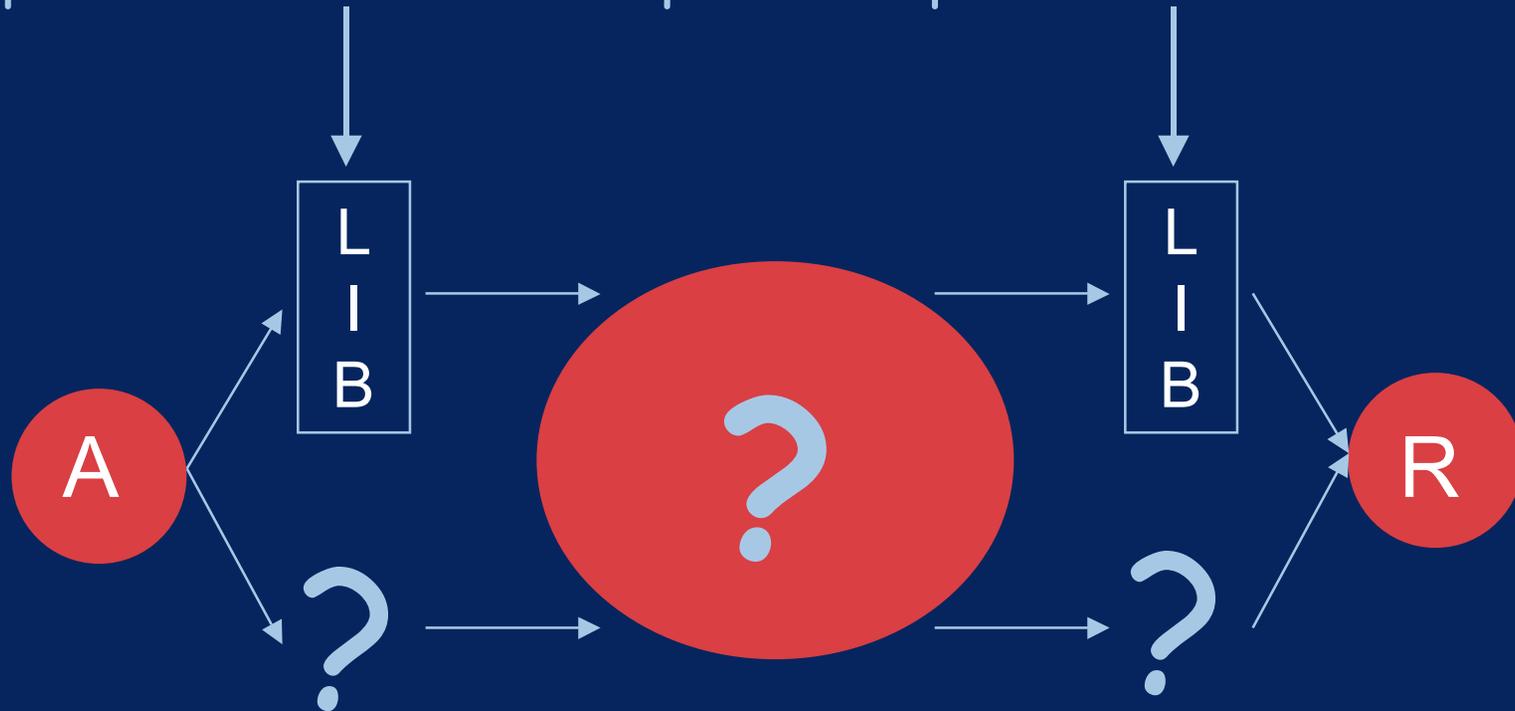
- Technical issues:
 - global level/interoperability (OAI, ...)
 - institutional level
- Unknown cost parameters (now starting to get experience)
- Dependence on current journal system role in academic advancement (rewarding)
- Systemic inertia
- Faculty participation



New library positions?

capture and share the input...

...portals and services

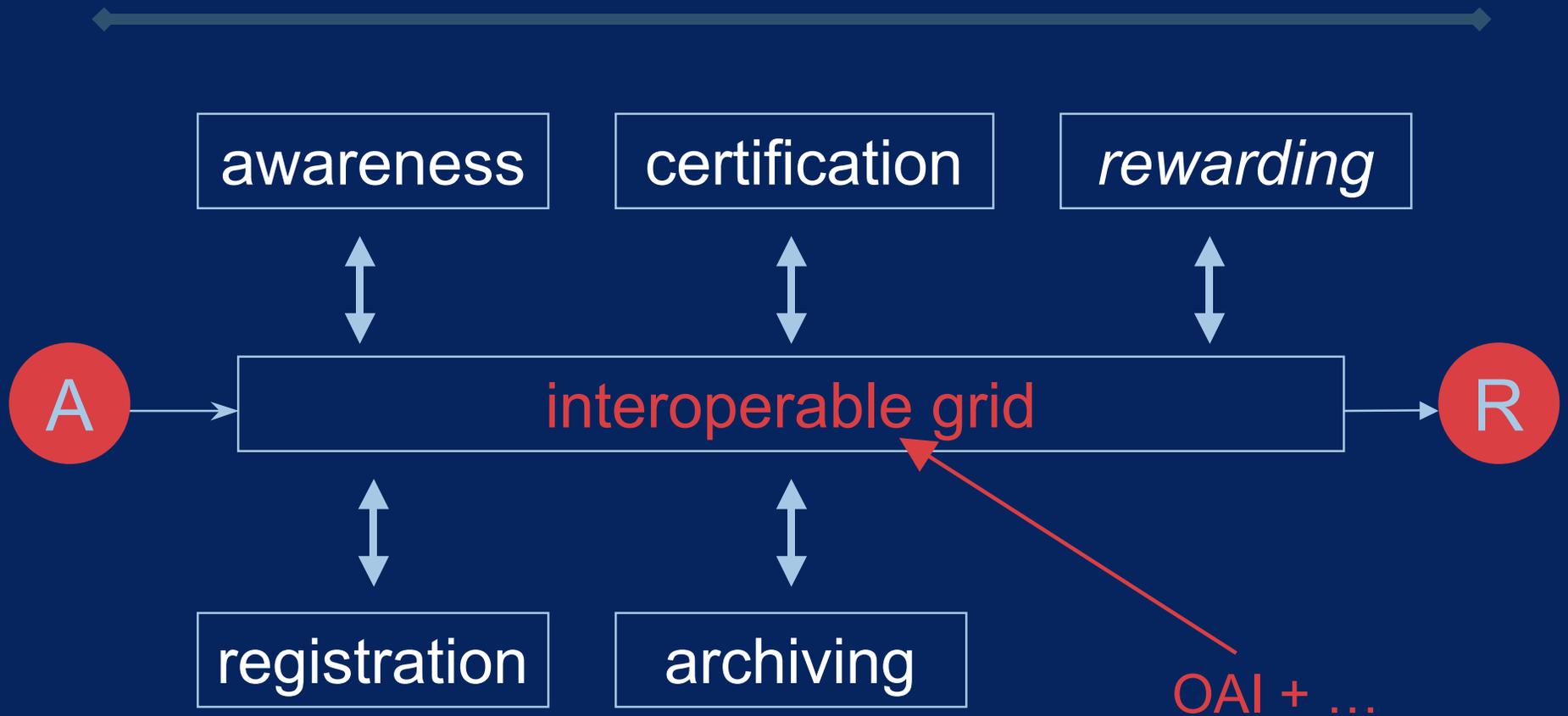


Disaggregation?

- Traditional journal publishing combines functions: registration, certification, awareness, archiving.
- How about eprints being the starting point of a new value chain in which the raw material - the non-certified eprint - is open access?
- Other functions might be fulfilled by different networked parties



A disaggregated view



OAI in a disaggregated system

Achieve interoperability by ensuring that **information about** the fulfillment of the different functions:

- can travel across the system
- can be shared by nodes of the system

metadata



The promise of the OAI

- So far: harvesting of descriptive metadata, search and browse services
- Provides necessary infrastructure for the growing number of discipline-specific and institutionally based repositories.
- Better interoperability will promote adoption of Open Access models.
- Will support new, disaggregated models of scholarly communication.





Questions?



CORNELL

