

The Promise of the Open Archives Initiative

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

Thanks to Herbert van de Sompel (LANL)

CORNELL

CARL conference “Research Innovation and Scholarship:
The Role of Open Access Publishing”
21 November 2002, Ottawa, Ontario, Canada



Scholarly communication in flux

- Technology: paper → digital → networked
 - changes access and economic considerations
- Problems with existing model of scholarly communication
- Open Access model addresses some problems, creates new challenges
- The OAI provides necessary functions for Open Access scholarly communication



The journal system

- **Serials crisis:** increasing journal prices limit, rather than broaden, access to scholarly output
- **Copyright loss:** faculty sign away copyright
- **Publication delay:** journal system is slow, cannot cope with increasing volume of scholarly output
- **Criticism of peer-review:** suppresses ideas, outcome criticized, expensive
- **Inertia:** system is self-stabilizing, existing players tend to maintain status quo



Functions of scholarly communication



registration	establishing intellectual priority
certification	certifying quality/validity of research
awareness	ensuring accessibility of research
archiving	preserving research for future use

(Roosendaal & Geurts)



Functions of scholarly communication

registration	establishing intellectual priority
certification	certifying quality/validity of research
awareness	ensuring accessibility of research
archiving	preserving research for future use
<i>rewarding</i>	<i>evaluating & rewarding performance</i>

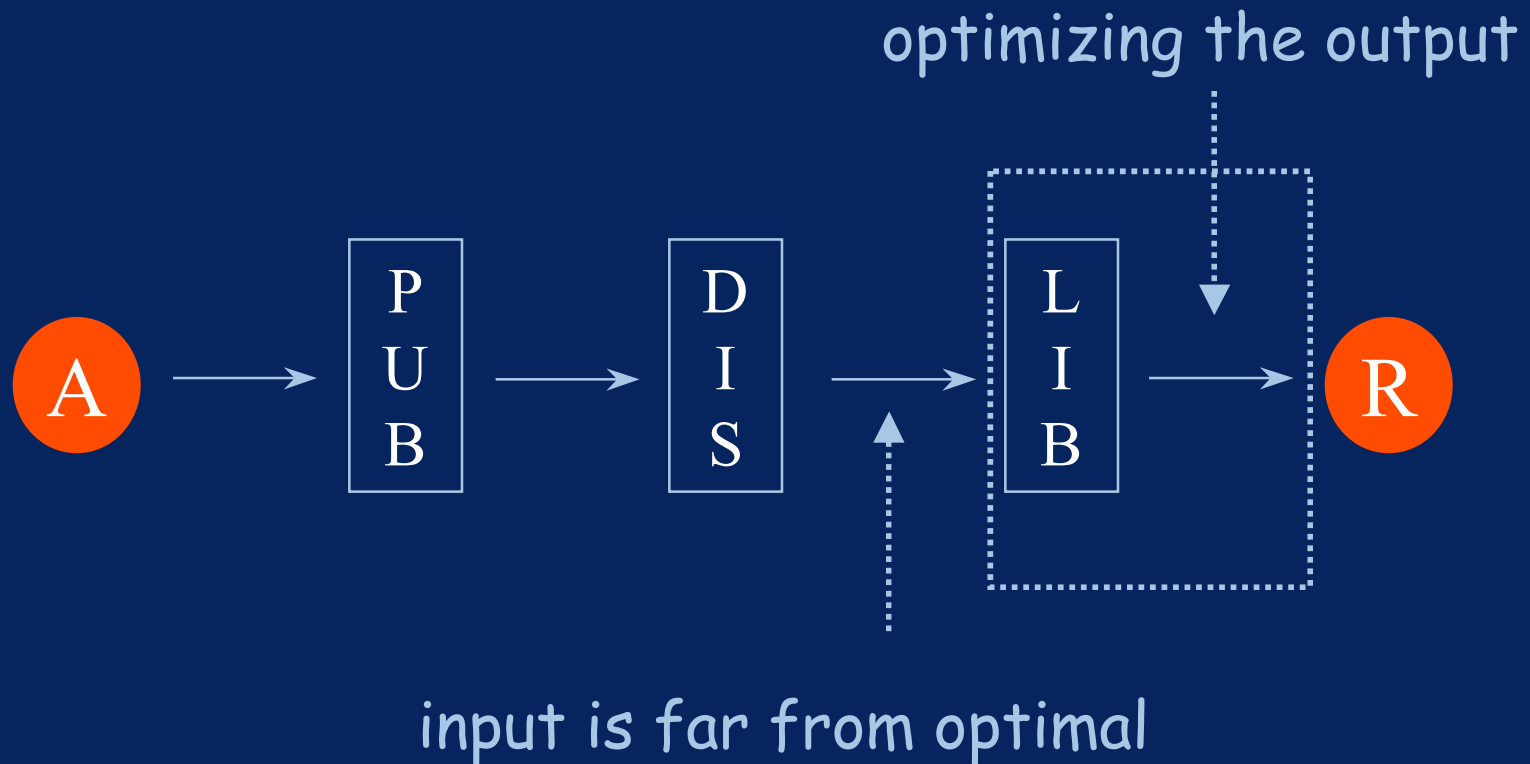


Journal system integrates functions

- Provides certification (usually via peer review)
- Accepted status of journals provides for rewarding
- Libraries provide archiving (and shoulder additional cost)



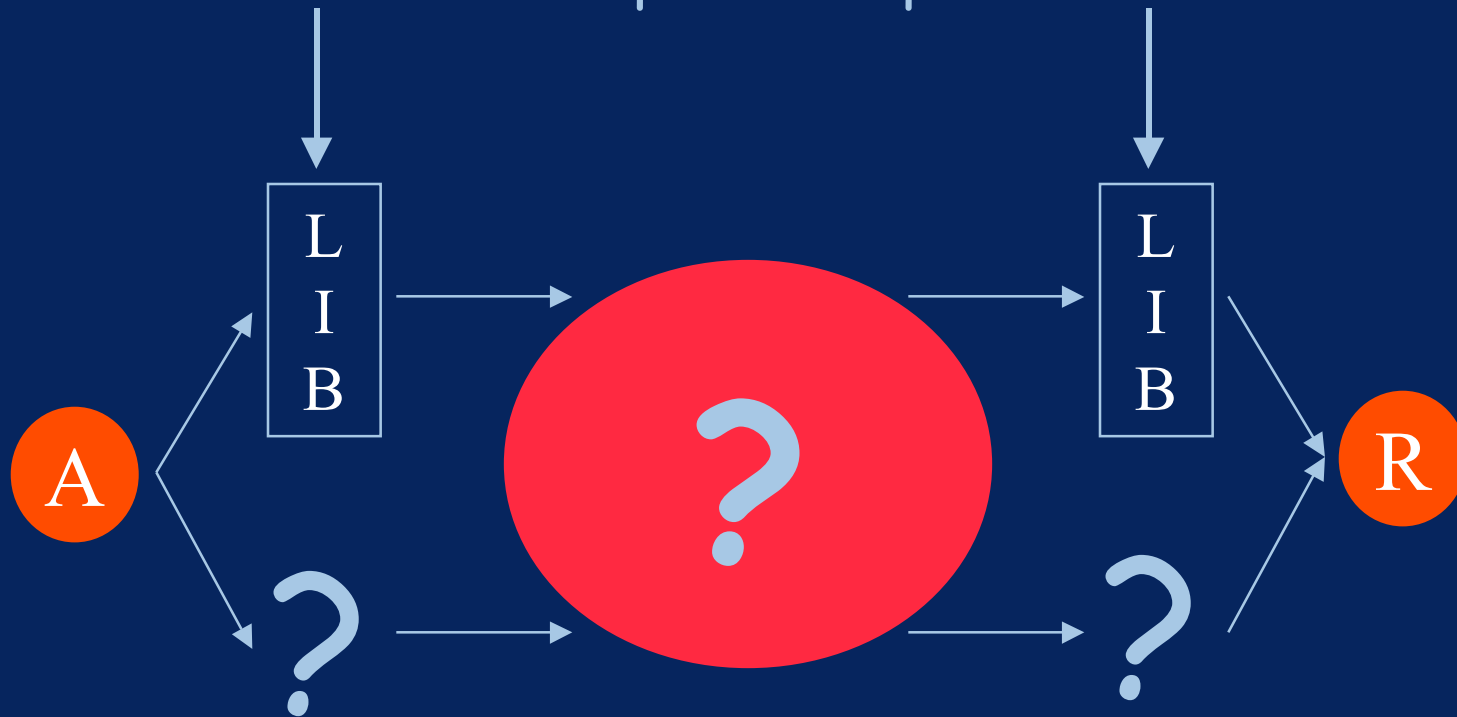
Traditional library position



Alternative library positions?

capture and share the input...

...portals and services



Pre-print and eprint systems

- **arXiv** (aka xxx) eprint archive (220k eprints)
(Physics - since 1991 - Los Alamos - Ginsparg)
- **RePEc** (231k records, 6k eprints)
(Economics - Surrey U - Krichel)
- **NCSTRL** (20k eprints)
(Computer Science - Cornell U - Lagoze)
- **NDLTD** (e.g. VTETD, 3.6k total, 2.4k eprints)
(Theses - Virginia Tech - Fox)
- **CogPrints** (1.5k eprints)
(Cognitive Sciences - Southampton U - Harnad)



Eprints

- Term used with many meanings
- Here:
 - Scholarly literature including journal articles, pre-prints, technical reports, books , theses and dissertations
 - Open Access to full-content via Internet
 - May or may not be refereed



Eprint systems

registration	yes
certification	no
awareness	yes (if critical mass)
archiving	no (generally)
<i>rewarding</i>	<i>no</i>



arXiv eprint archive

- Certification
 - Most users rely on conventional journals for certification
 - Some peer-reviewed overlay journals
- Archiving
 - Not adequately addressed
- Rewarding
 - Relies on conventional certification outside immediate community





Institutional repositories



CORNELL



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?



Rationale for institutional repositories

- Local and immediate
 - Increases institutional visibility and prestige by clarifying institutional sources of research
 - Demonstrates institution's value to public and private funding sources
 - Archives institutional production
- Global and long-term
 - Component in evolving disaggregated scholarly publishing model
- Complements existing scholarly publishing model



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
- Dependence on current journal system role in academic advancement (rewarding)
- Systemic inertia
- Faculty participation





The Open Archives Initiative



CORNELL



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?

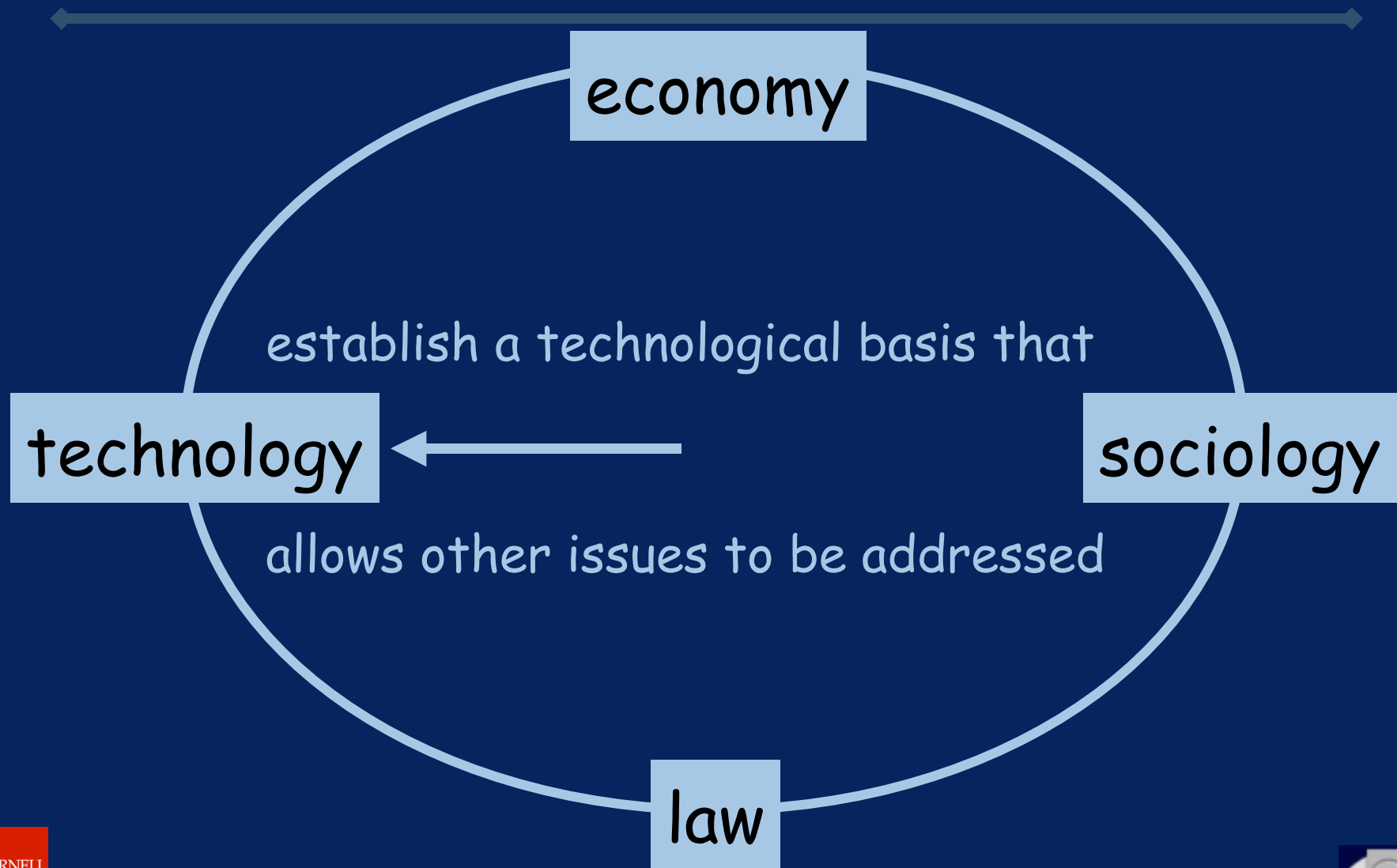
"The OAI develops and promotes interoperability standards that aim to facilitate the efficient dissemination of content." (from OAI mission statement)

- Technological framework
- Application independent
- Independent of economic model for content

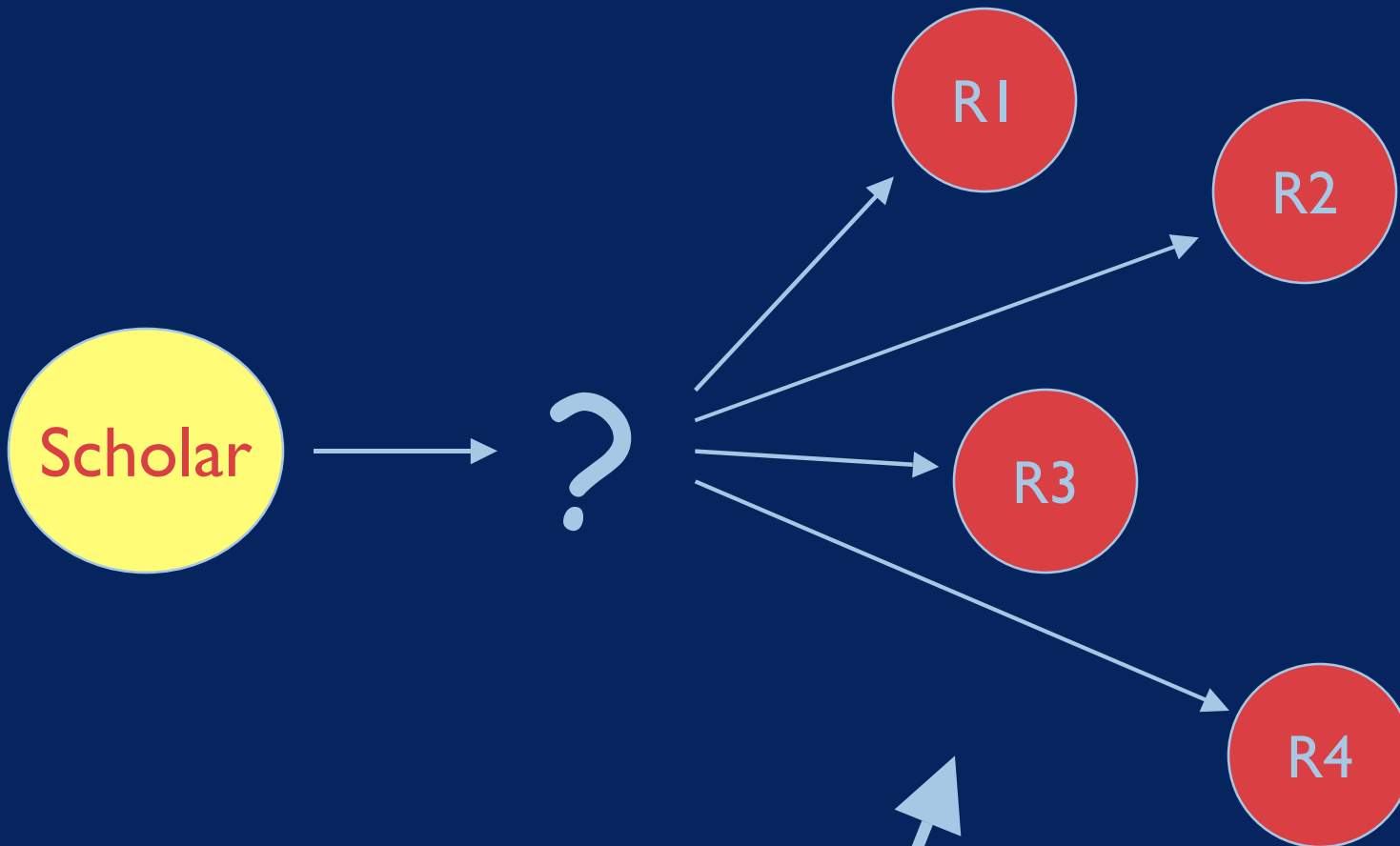
Also ... a community and a "brand"



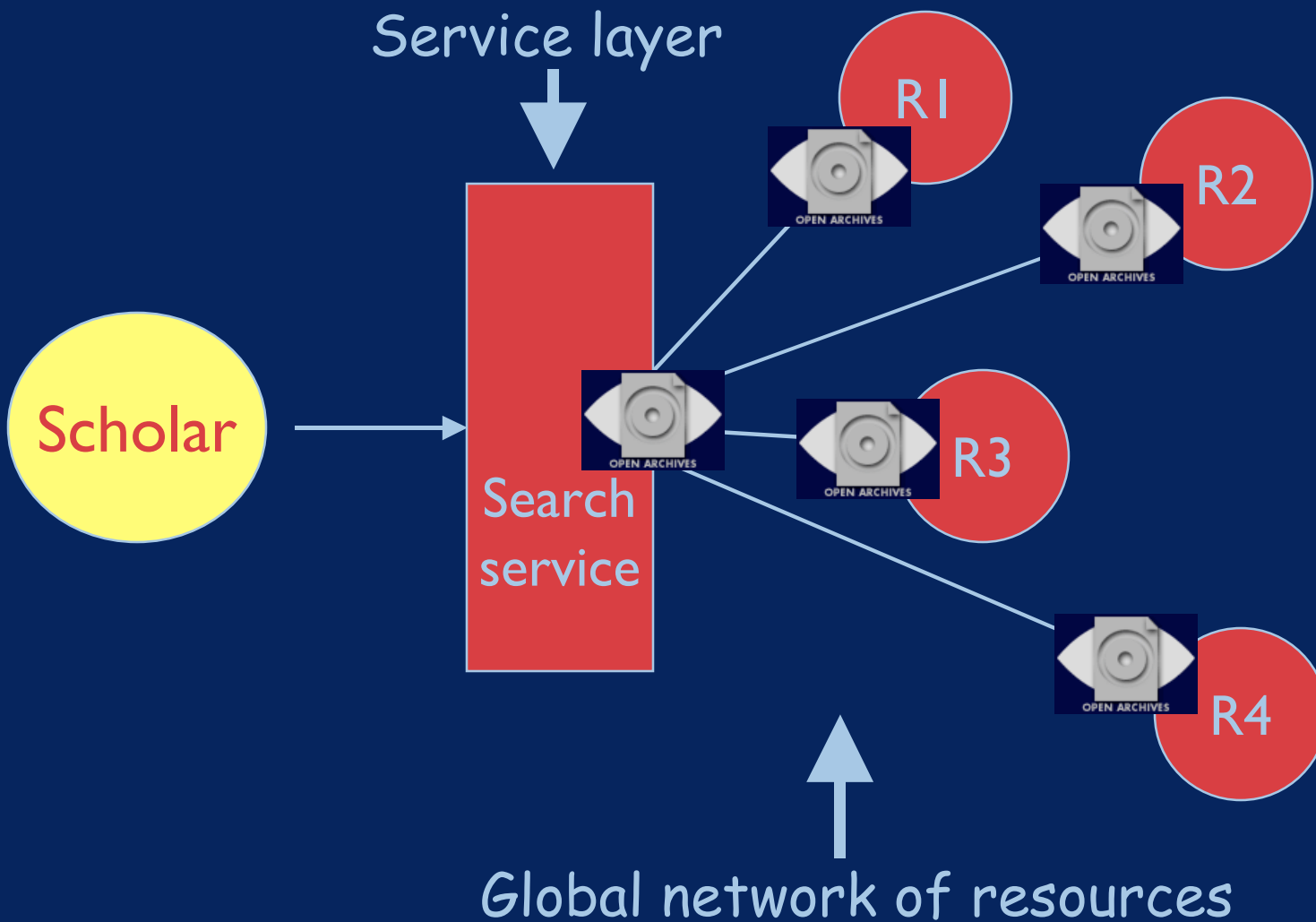
OAI in context



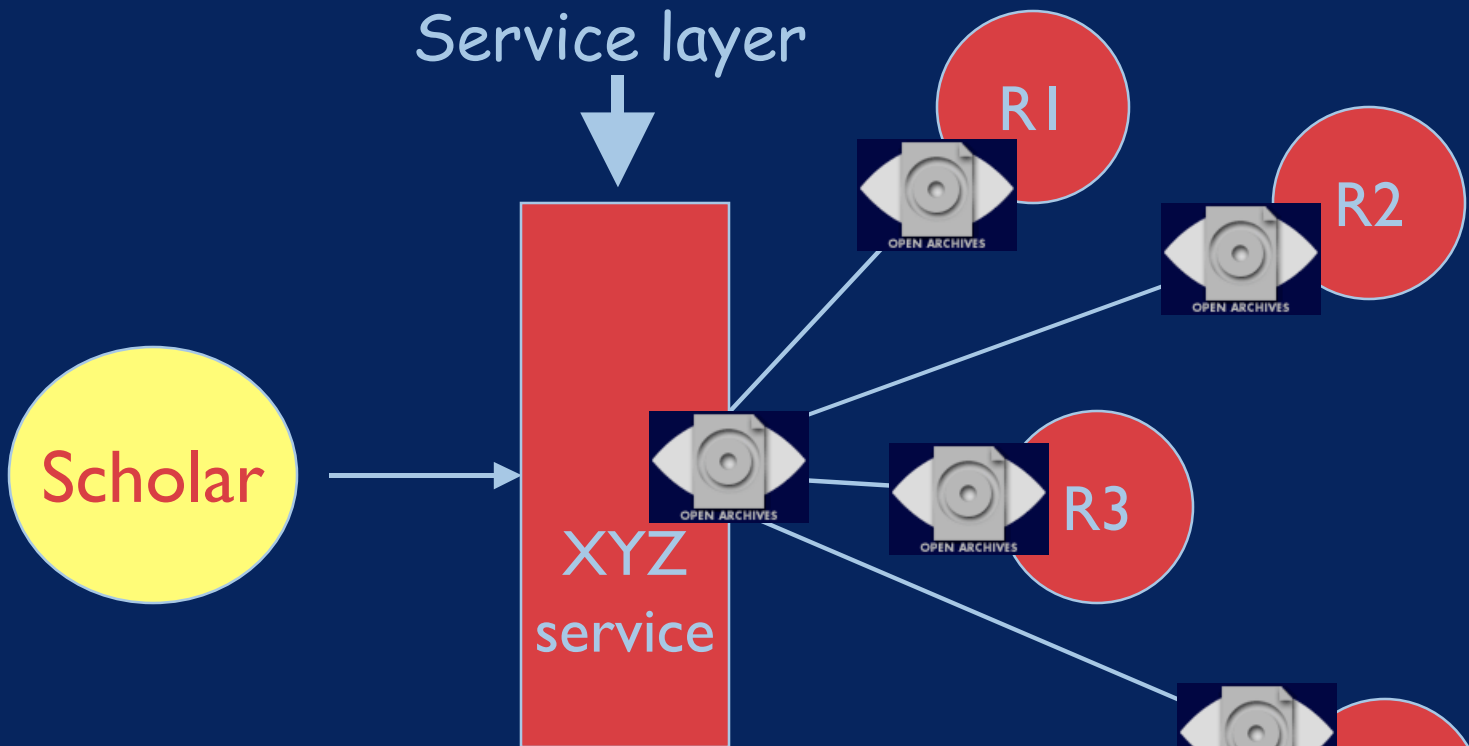
OAI for discovery



OAI for discovery



OAI for XYZ



OAI vs 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 for Open Access

- OAI will focus on support for eprints in the next year
 - Use existing, application-neutral, OAI protocol for metadata harvesting
 - Share eprint specific metadata, promote specific uses
- further Open Access agenda





Disaggregated systems



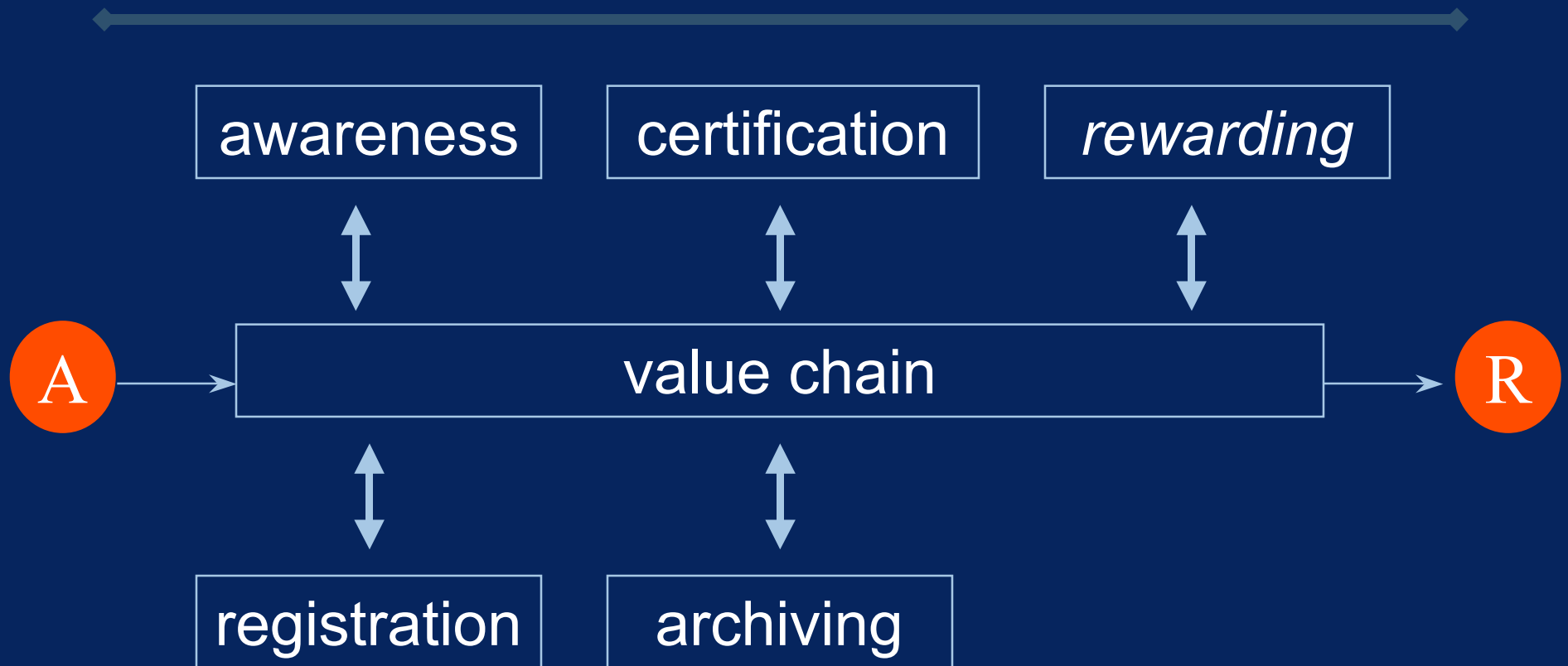
New models of scholarly communication
based on Open Access to content with
functions unbundled.



CORNELL



Systems for scholarly communication



Experience with eprints suggests

- The feasibility of a disaggregated system for scholarly communication
- The possibility of 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



Disaggregated system



- Increase cost efficiency:
 - decoupling value chain forces market efficiency of individual links
 - possible competition throughout chain
- Reveal that academy contributes most of the value
 - academic labor and institutional investment drives content, certification and archiving



Disaggregated system: How?

- Content providers: discipline-specific repositories, institutional repositories, peer-to-peer research repositories, ...
- Service providers:
 - value-added services that provide certification, awareness, archiving and rewarding functions
 - current agents of these functions (e.g. societies) can operate in disaggregated model
 - new entrants in the system possible
 - various business models possible



The OAI's role

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

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

metadata

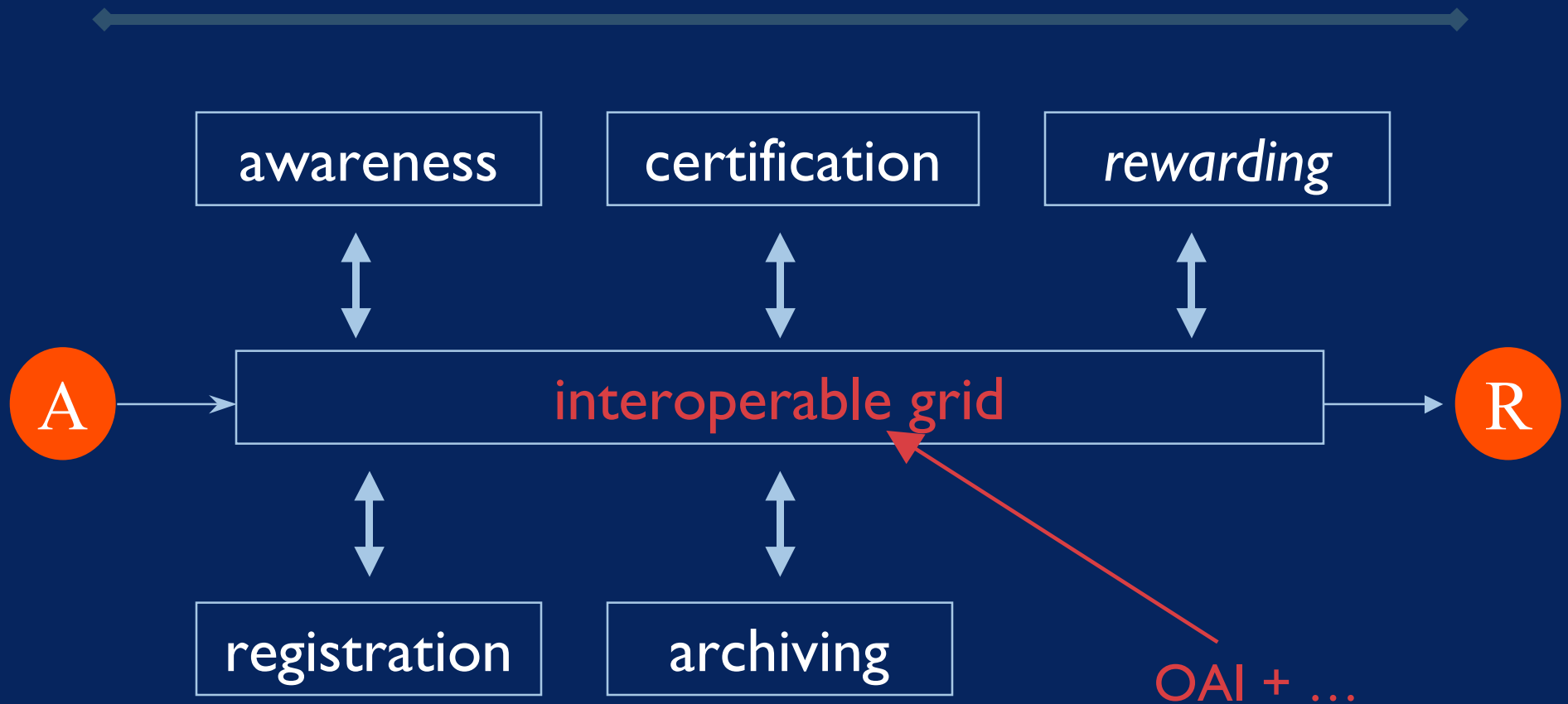


The OAI's role

- So far: harvesting of descriptive metadata, search and browse services
- Soon also harvesting of:
 - references
 - usage logs
 - certification metadata
 - rights metadata
 - ...



OAI as infrastructure



What's in it for the stakeholders?

Consider a scenario where there is Open Access to content and an OAI infrastructure supporting discovery and other services.

Scholars



As authors:

- Widest possible dissemination
- Keep rights to output

As readers:

- Unified discovery tools
- Toll free access to full content
- New services possible (e.g. find 'similar to')



Institutions

- Lowers costs to access literature
- Better control of own output (via less restrictive copyright transfer agreements/licenses; even more for institutional repositories)
- Local collections record institutional activity



Journals

- Need appropriate economic model so that open access is possible, then:
- Easy integration with existing (global) discovery facilities. Lowers barrier for new startup
- Avoid overhead of managing/controlling access to content
- Community good will



Funding bodies

- Widest possible dissemination of results
- Avoid double spending to `buy back' output of funded projects
- Assemble virtual collections or archives of output from funded work



The world?

- Open Access model does not reinforce existing economic and information inequalities, instead levels the playing field somewhat.



The promise of the OAI

- 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

