Computation and (Today's) Internet Services

Leslie L. Daigle

Vice President Research

Bunyip Information Systems Inc

Montréal, Canada

Patrik Fältström

Senior Researcher

Tele2

Stockholm, Sweden

November 5, 1996

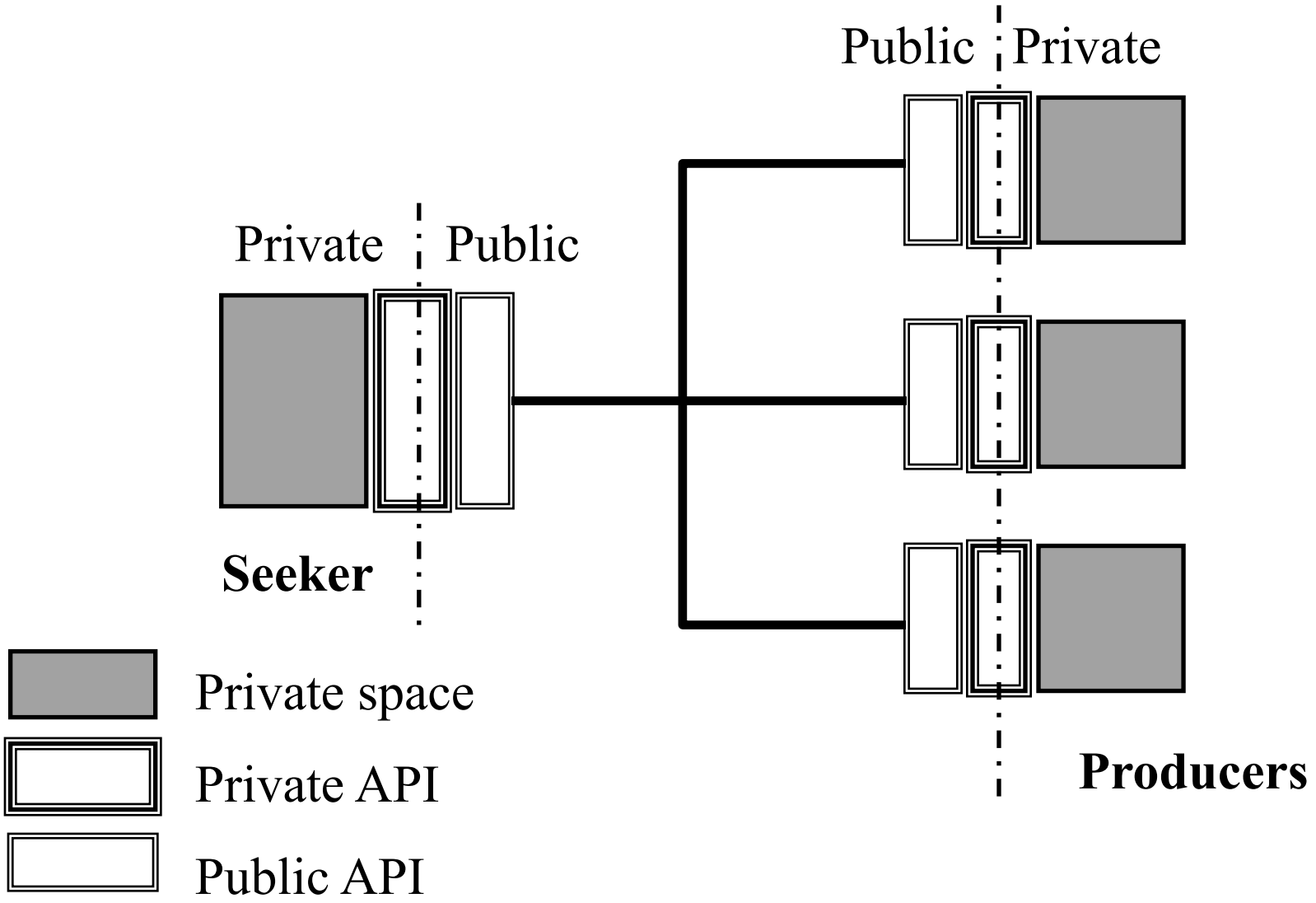
Remember that we have...

- ...different **owners** of
  - Source
  - Metadata
  - Index
  - Query
  - Client's knowledge

Who are the owners?

- Producer
  - Owner of information
- Indexer
  - Mediator or “middle man”
- Consumer
  - Seeker of information

Privacy



The Consumer's Eye-View

- How do I find the stuff I want?
- How do I know if I've found all available resources?
- How do I limit the number of "hits" I get?
- How can I express my needs in the language of the search tools?

The Producer's Perspective

- Once I have produced electronic information, how do I advertise it?
- How do I ensure that the right people find it?
- How can I limit the range of people who find it (security)?

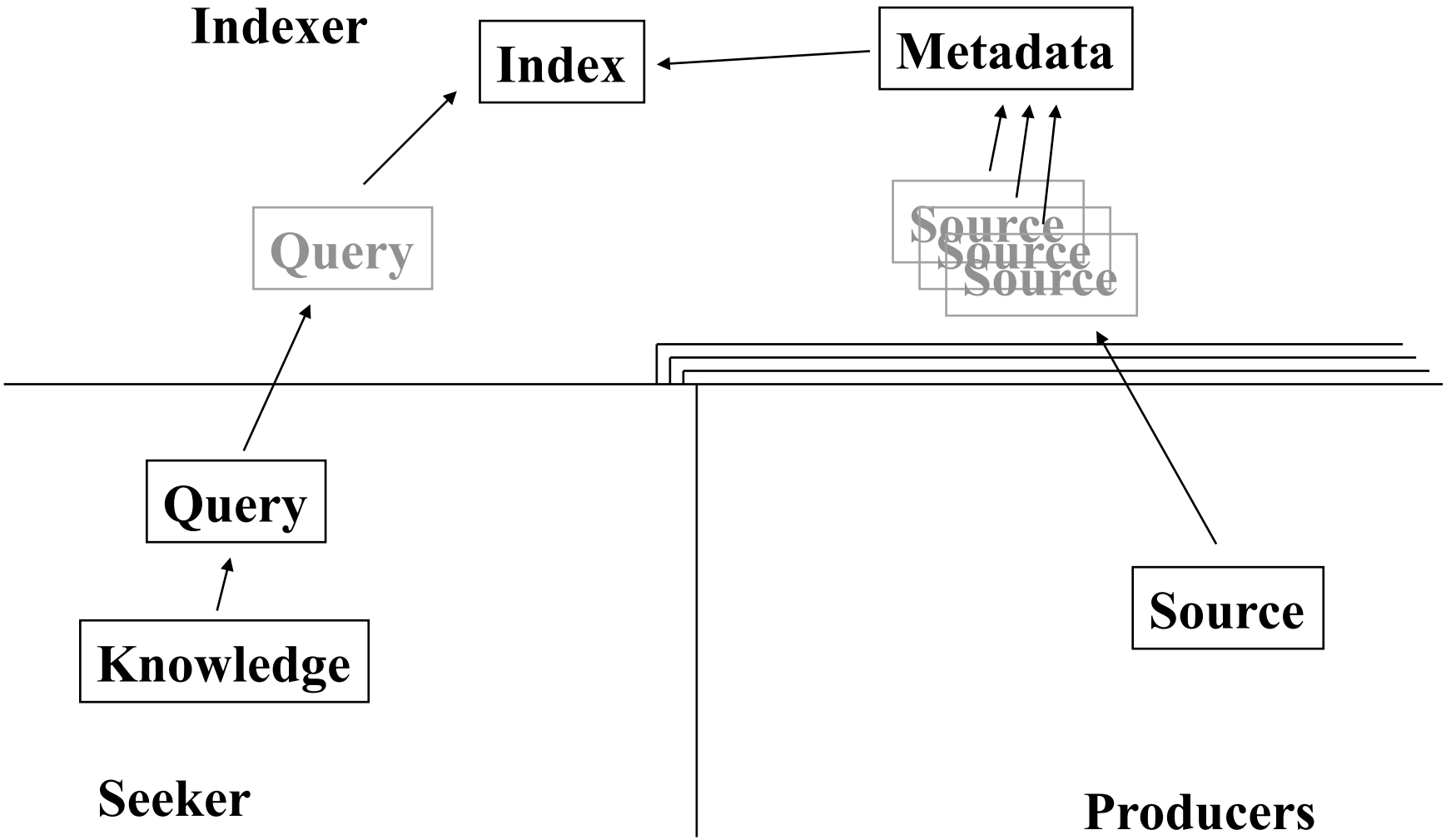
Models of Distributed Internet Computing - 1996

- Common
  - Full-sweep “pull” indexing
  - Cooperative (distributed) “pull” indexing
- New
  - Forward knowledge “push”
- On the horizon...
  - New paradigms (URAs, URNs)

Full-Sweep “pull”

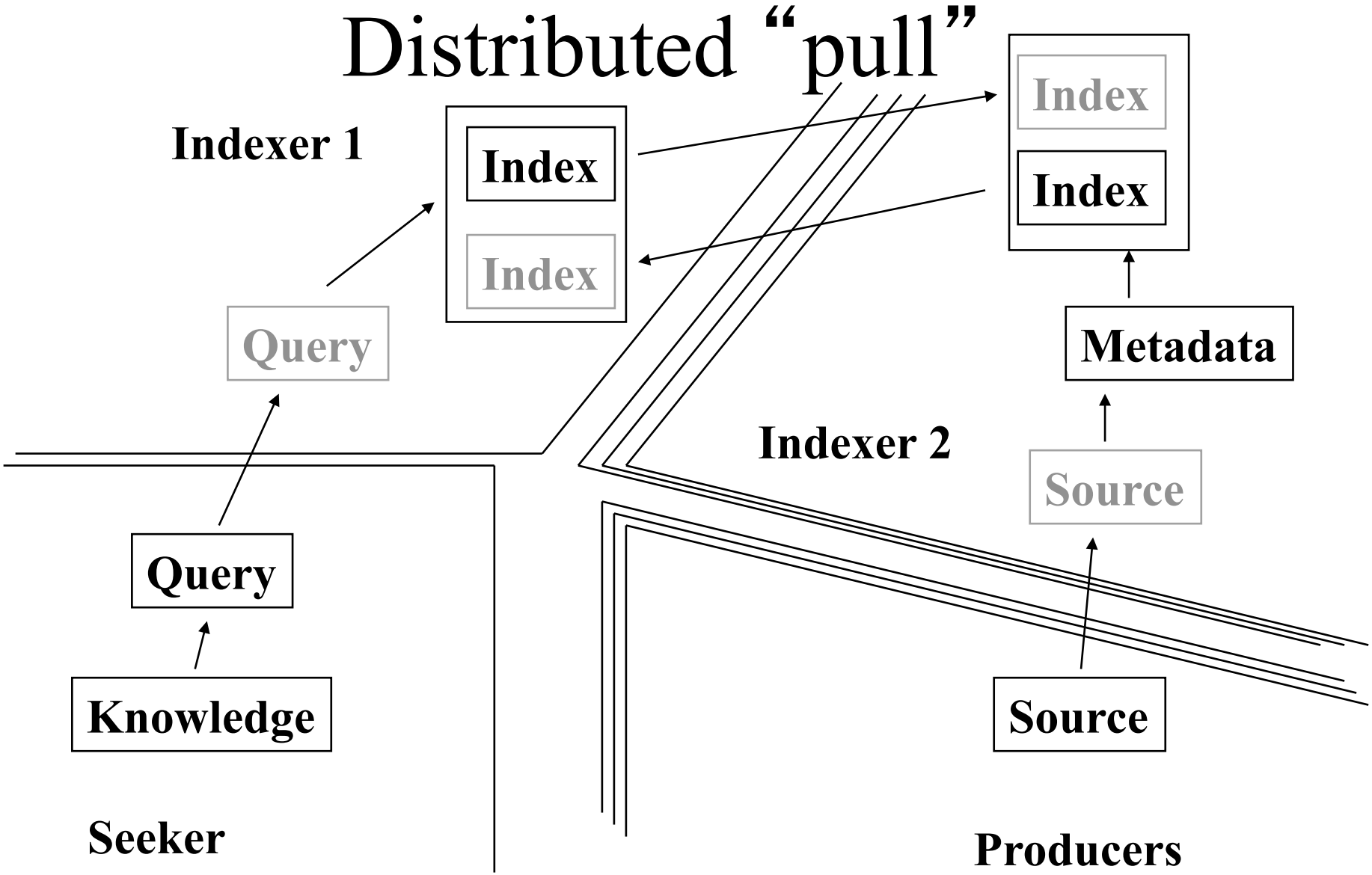
- Lycos, Alta-Vista et.al.
  - *Pro* - proprietary indexing format
  - *Pro* - uniformity and consistency of index data
  - *Con* - huge undertaking to fully map the space

Full-Sweep “pull”



Distributed “pull”

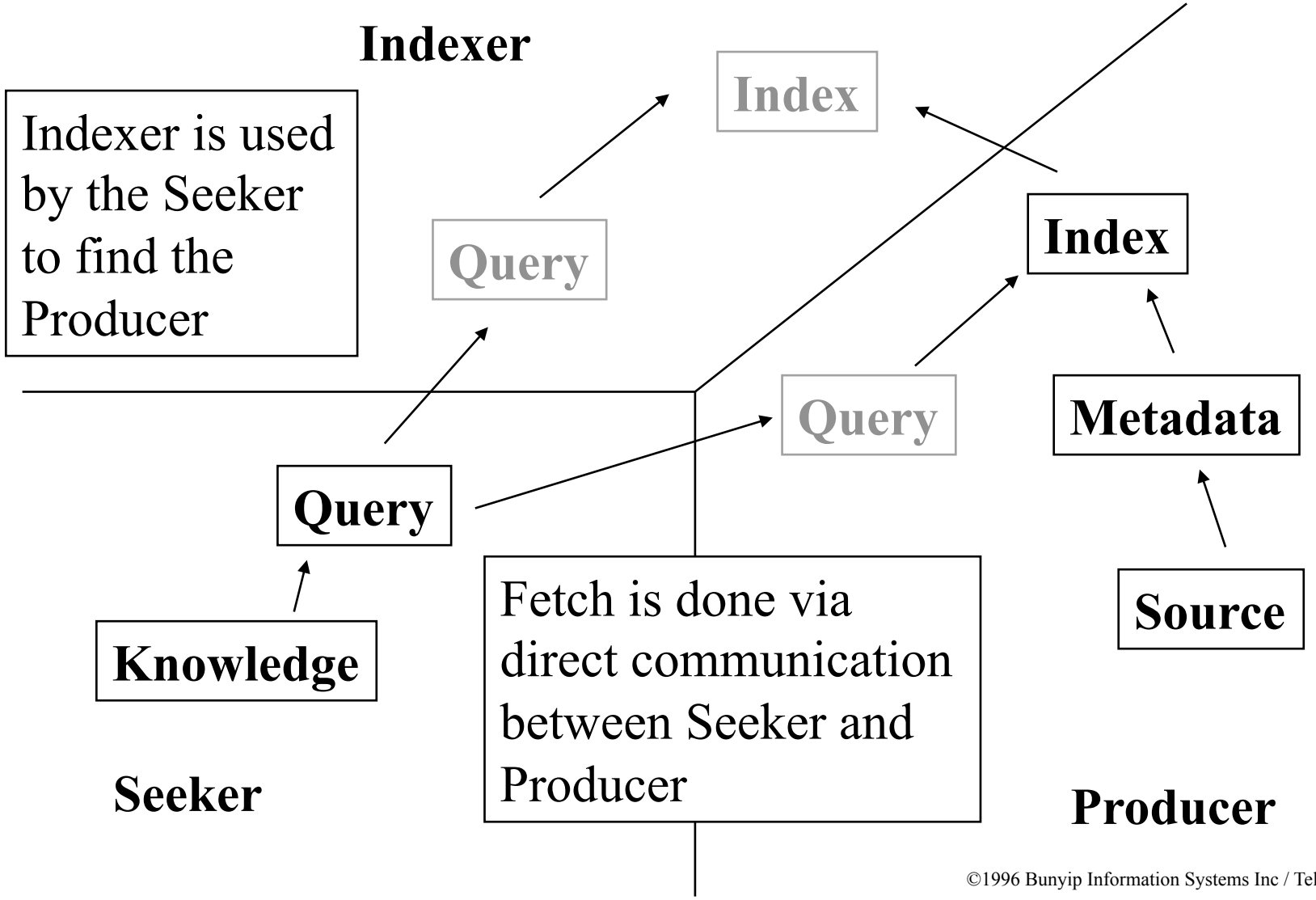
- Archie et.al.
  - *Pro* - maintain proprietary data format
  - *Pro* - maintain uniformity and consistency of index data
  - *Pro* - multiple parts of the space can be mapped independently
  - *Con* - integrating and coordinating results from different participating sources



Forward Knowledge “push”

- Whois++, Common Indexing Protocol et.al.
  - *Pro* - update is in the hands of the information producer
  - *Pro* - searches are done close to the source
  - *Pro* - specialized index hierarchies can be built
  - *Con* - update is in the hands of the information producer

Forward Knowledge “push”



Fetch and Search

- Fetch
  - You know what you are looking for
  - You have the path to the file in a filesystem
  - You have the domainname to lookup in DNS
  - You have the Distinguished Name in X.500
- Search
  - You do not know what you are looking for
  - You know some words in the file
  - You have the name of the host (not domain) in DNS
  - You have the name of a person to find in X.500

Fetch and Search

- Whats the point?
  - Fetching things is simple
  - Searching locally in a controlled environment is easy
  - Searching globally, in a heterogenous environments is very hard
- Indexing is needed!

Distributing knowledge

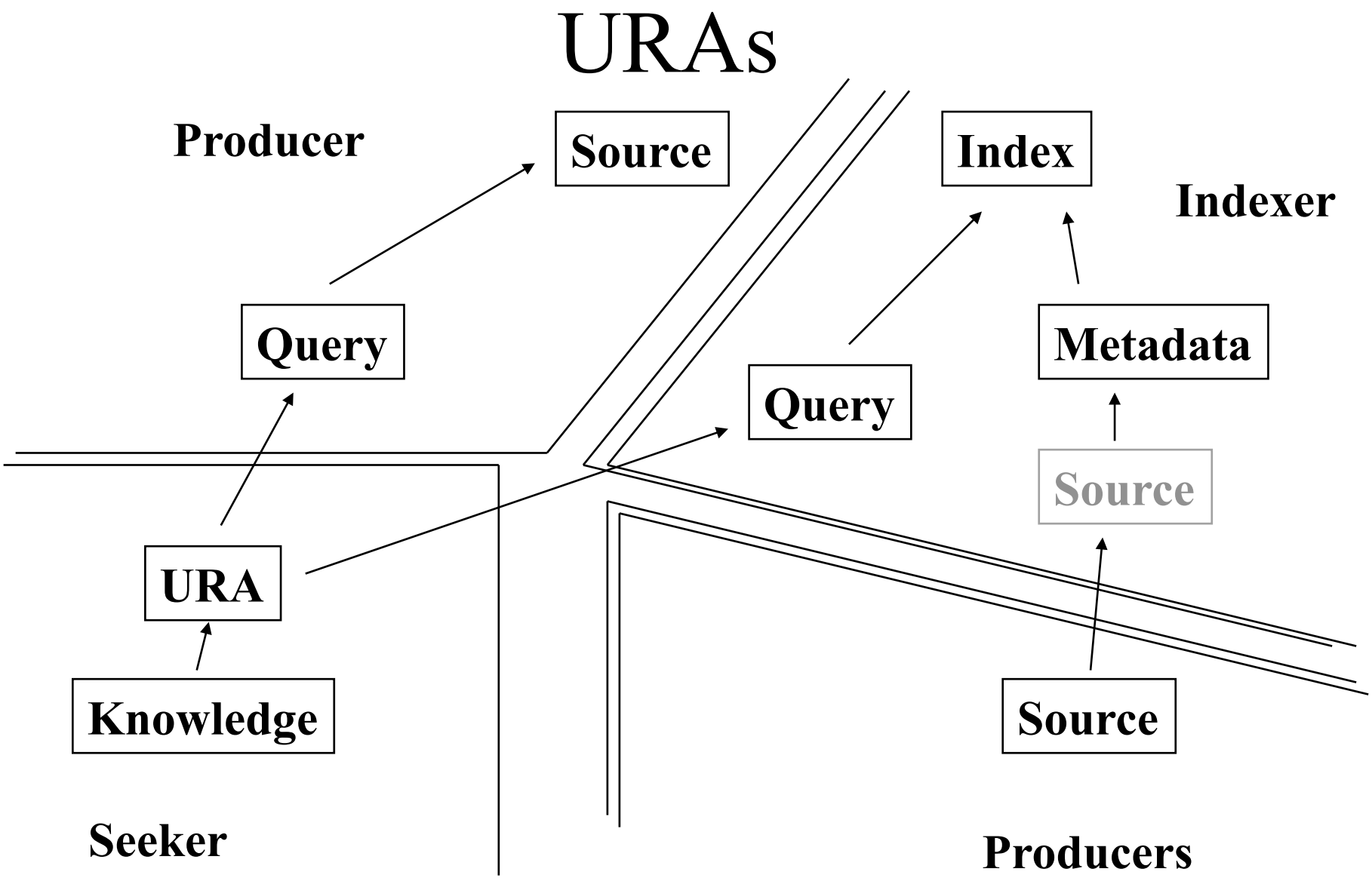
- Passing (pushing) index information is a way of announcing the knowledge you have
- A Seeker can use that information to find a Producer
- This is called Query-Routing
  - See Common Indexing Protocol, the FIND working group in the IETF and Whois++

Future Paradigms

- More client-side computation
- Longevity through indirection

Uniform Resource Agents

- Client-side information representation
- Provide a formalized structure for encapsulating composite Internet activities
  - URA header
  - Activation data
  - Targets
  - Experience information
  - Activity
  - Response filter



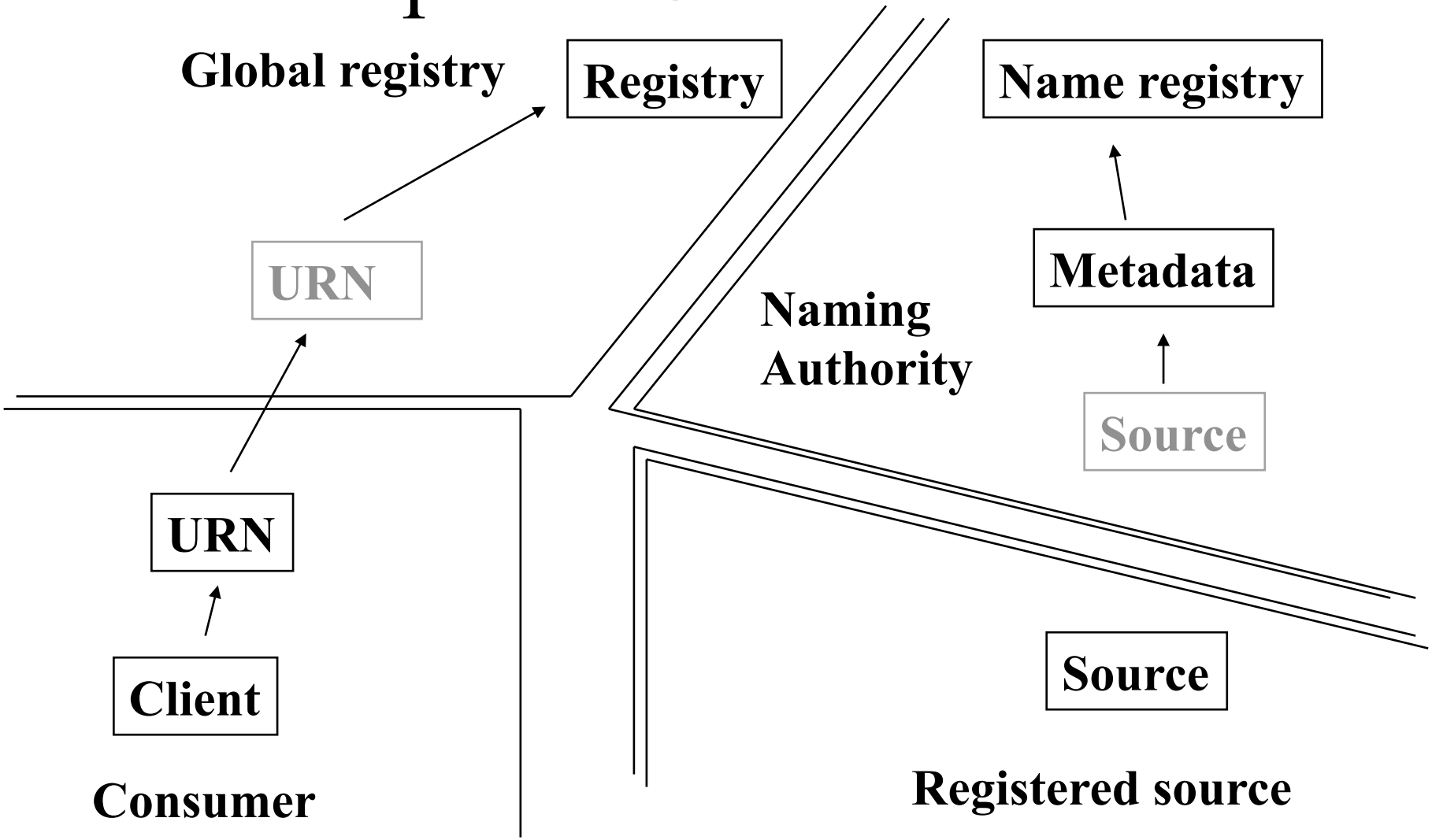
Uniform Resource Names

- A rose, by any other name... would be hard to find on the Internet
- URL - resource location
  - 3rd item from the left, 4th shelf down
- URN - resource name
  - Moby Dick

UR{L|N} Computational Distinctions

- URL
  - specify location & protocol to access desired resource
  - resource may move, disappear, change
- URN
  - mapping from identifier to content must persist across changes in ownership, etc
  - therefore, require computation outside the space of the resource owner

The Proposed URN Framework



NAPTR - An Expermental URN Implementation

- A longlived name of the content of a resource
  - `urn:inet:acme.se:information`
- Lookup rules for “inet” (NAPTR for inet.urn.net)
- Lookup rules for acme.se (NAPTR for acme.se)
- Lookup rules for http in the domain acme.se (SRV for http.tcp.acme.se)
- Connect to www.acme.se port 80 and fetch
  - `http://www.acme.se:80/N2R/urn:inet:acme.se:information`

SRV Resource Record

- Pointing out an access point for a service in a domain
- Arguments are
  - –Order number
  - –Priority
  - –Portnumber
  - –Name of computer
- Example for http/tcp

```
http.tcp.acme.se.       IN SRV 10 10 80   server.acme.se.
http.tcp.sales.acme.se. IN SRV 10 10 1024 server.acme.se.
```

NAPTR Resource Record

- Naming authority pointer record
- Arguments are
  - Order number
  - Priority
  - Flags
  - Service
  - Regular expression
  - Replacement string

```
inet.urn.net. IN NAPTR 10 10 "" "" "/^[^:]*:inet:([^:]*):.*$/\1/i" ""
acme.se. IN NAPTR 10 10 "http+n2r" "s" "" http.tcp.acme.se.
```

Conclusion

- The Internet information space is characterized by a distribution of knowledge necessary for carrying out information (retrieval) tasks
- Different computational models exist and are being explored to deal with/make use of that distribution