

# Distribution Alternatives for Superimposed Information Services in Digital Libraries

Sudarshan Murthy, David Maier, Lois Delcambre

Department of CSE, OGI School of Science & Engineering at OHSU

<http://cse.ogi.edu/sparce>

<mailto:smurthy@cse.ogi.edu>

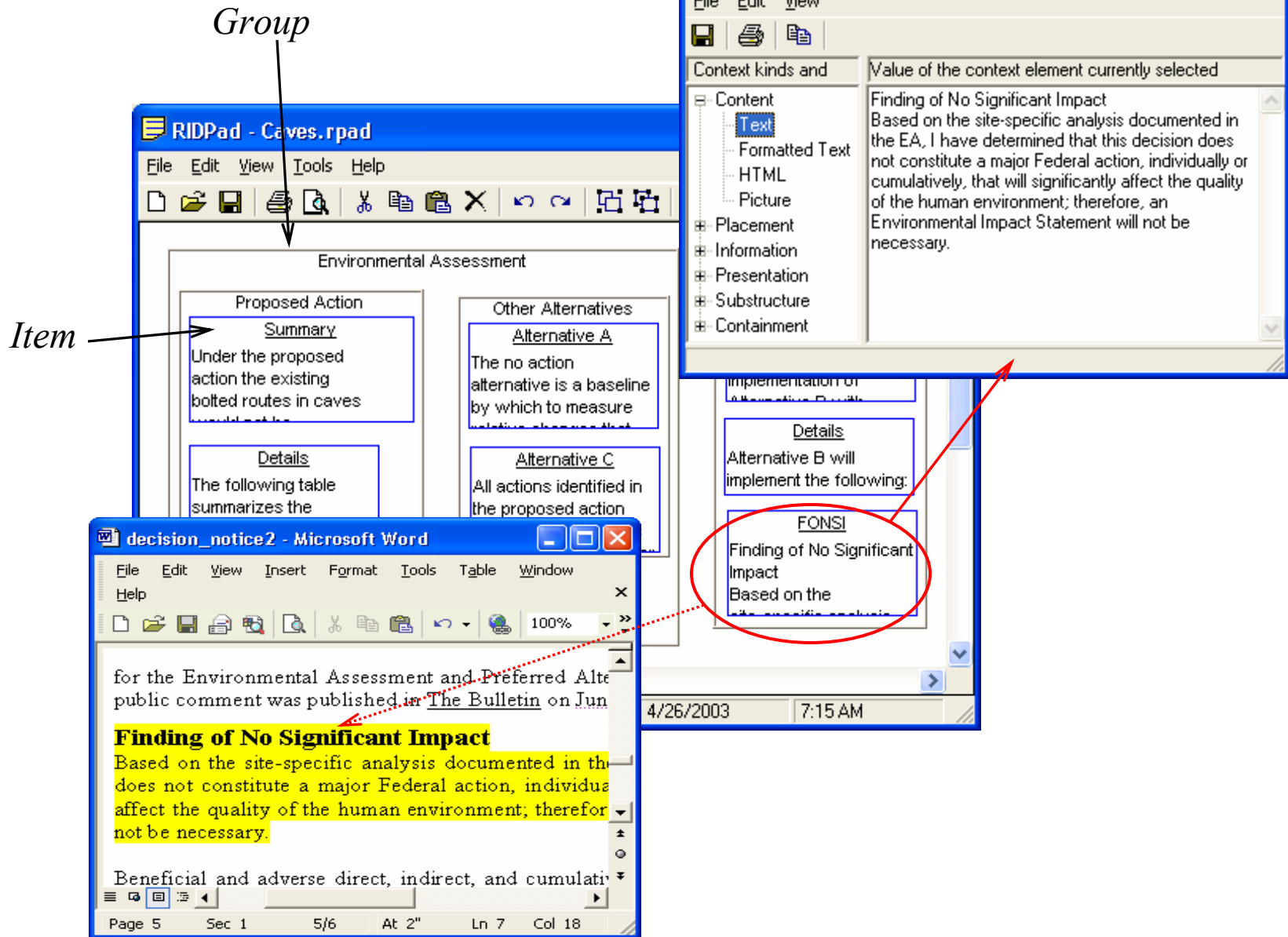
Or

**Digital Libraries: Where to Put the Parts**

# Superimposed Information\*



- People often *superimpose* new information onto existing information
  - Annotations, summaries, ...
- They use many means
  - Mark up paper
  - Place sticky notes on the paper
- They combine existing information and their interpretations to get “their” view



# Superimposed Information Services in DLs

- Related DL patron activities
  - Create, view, and share annotations and summaries
- Superimposed information services can support and potentially extend these activities

# Component-based Architectures

- We use a component-based middleware architecture, *SPARCE*, to support superimposed information management
- Component-based architectures have been used to support annotation and metadata over DL information
  - OAI-PMH, ODL
- Flexibility of deployment (distribution) is a benefit of using component-based architectures

# Goals

- Propose distribution alternatives for architectural components to add superimposed information services to a digital library
- Propose metrics to compare performance of the alternatives

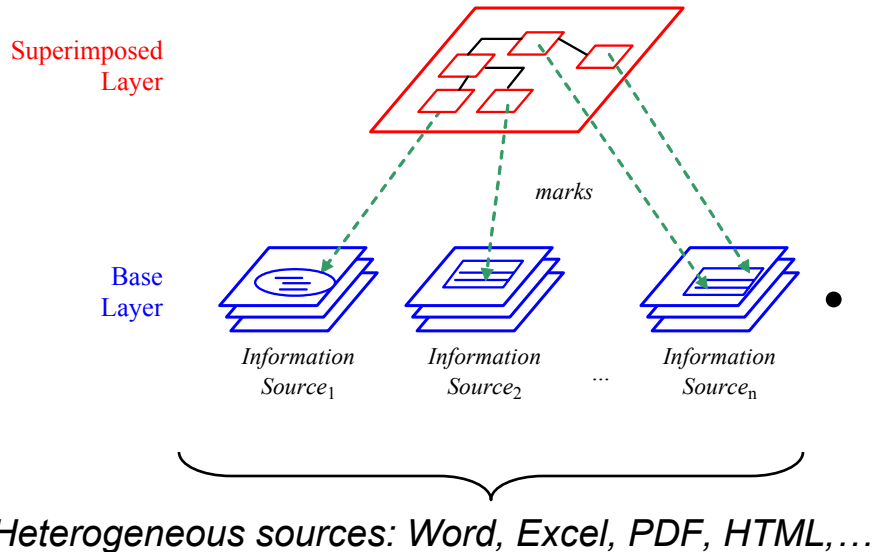
We use SPARCE to illustrate the alternatives

# Outline

- Motivation
- Background
  - Superimposed information management, SPARCE
- Distribution Alternatives
  - Metrics
  - Example alternatives
- Discussion
- Conclusion

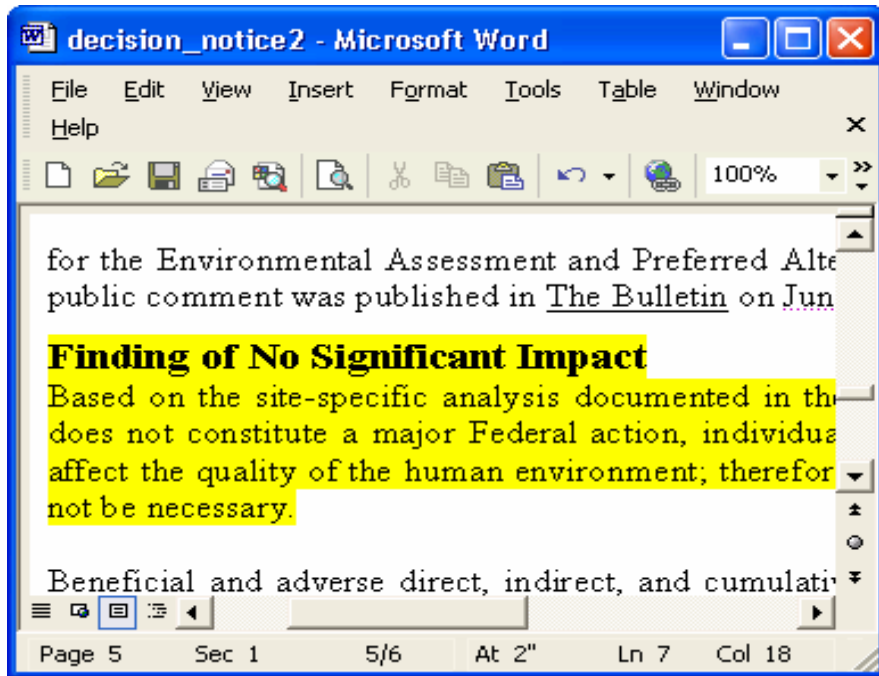


# Superimposing Information



- Overlaying new information on top of selected, existing information
  - Add new data
  - Impose new schema or model
- **Mark** is a reference to base element
  - Many implementations, ~ one per base type
  - Addressing scheme depends on base type

# Excerpts and Contexts



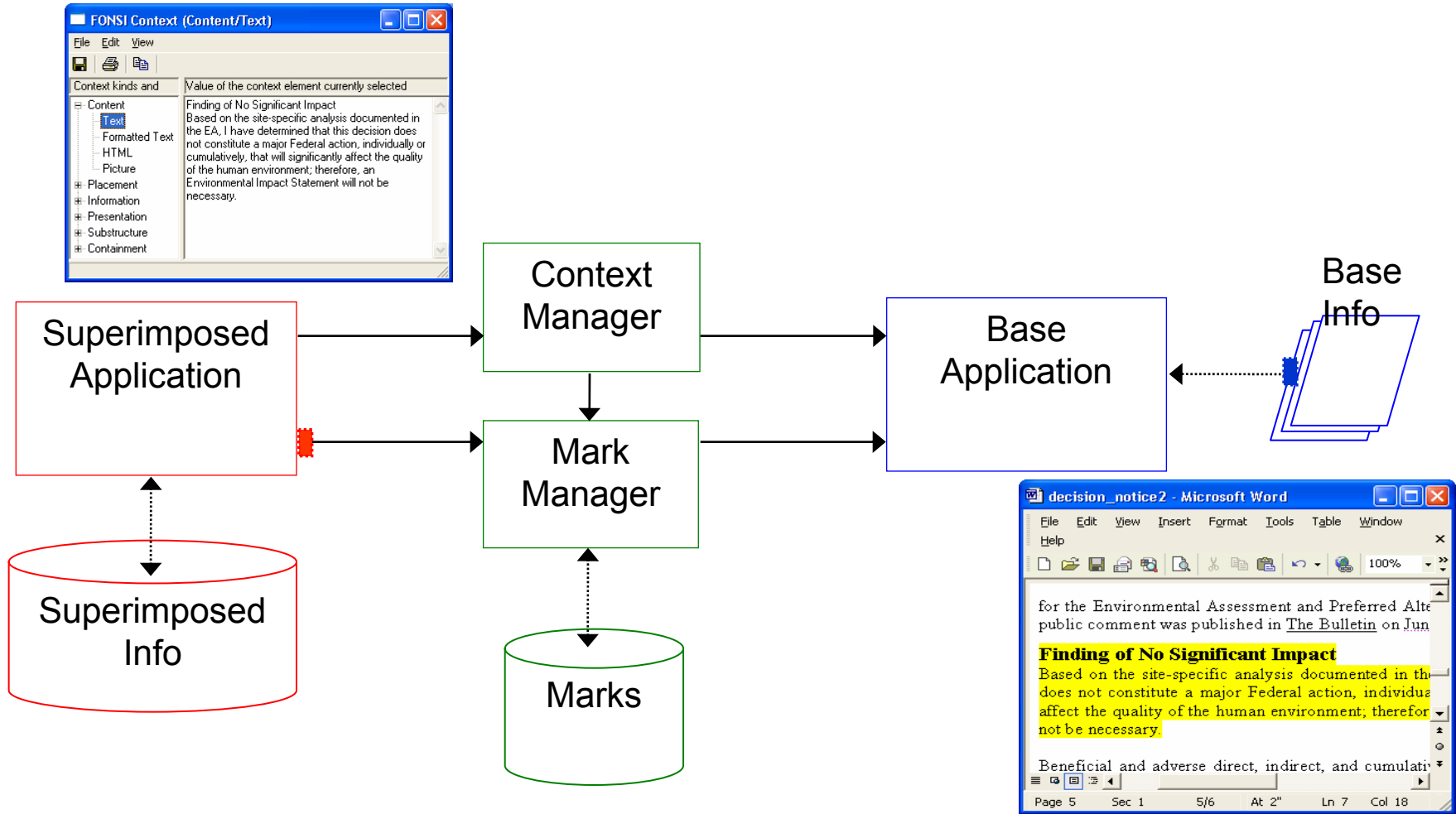
Name	Value
Excerpt	Finding of No Significant Impact ... not be necessary
Font name	Times New Roman
HTML	<pre>&lt;p&gt;&lt;b&gt;&lt;span style='font-size:14'&gt;Finding of No Significant Impact&lt;/span&gt;&lt;/b&gt;&lt;/p&gt; &lt;p&gt;&lt;span style='font-size:12'&gt;Based on ...&lt;/span&gt;&lt;/p&gt;</pre>

- *Excerpt* is the content of a marked region
- *Context* is information related to a mark
- What constitutes a context varies
- A mediator called *context agent* retrieves context of a mark

# SPARCE

- The *Superimposed Pluggable Architecture for Contexts and Excerpts*
  - Middleware for superimposed information management
- Provides mark and context management services
  - Use the same programmatic interface to work with any base type

# SPARCE Reference Model



# Distribution Alternatives

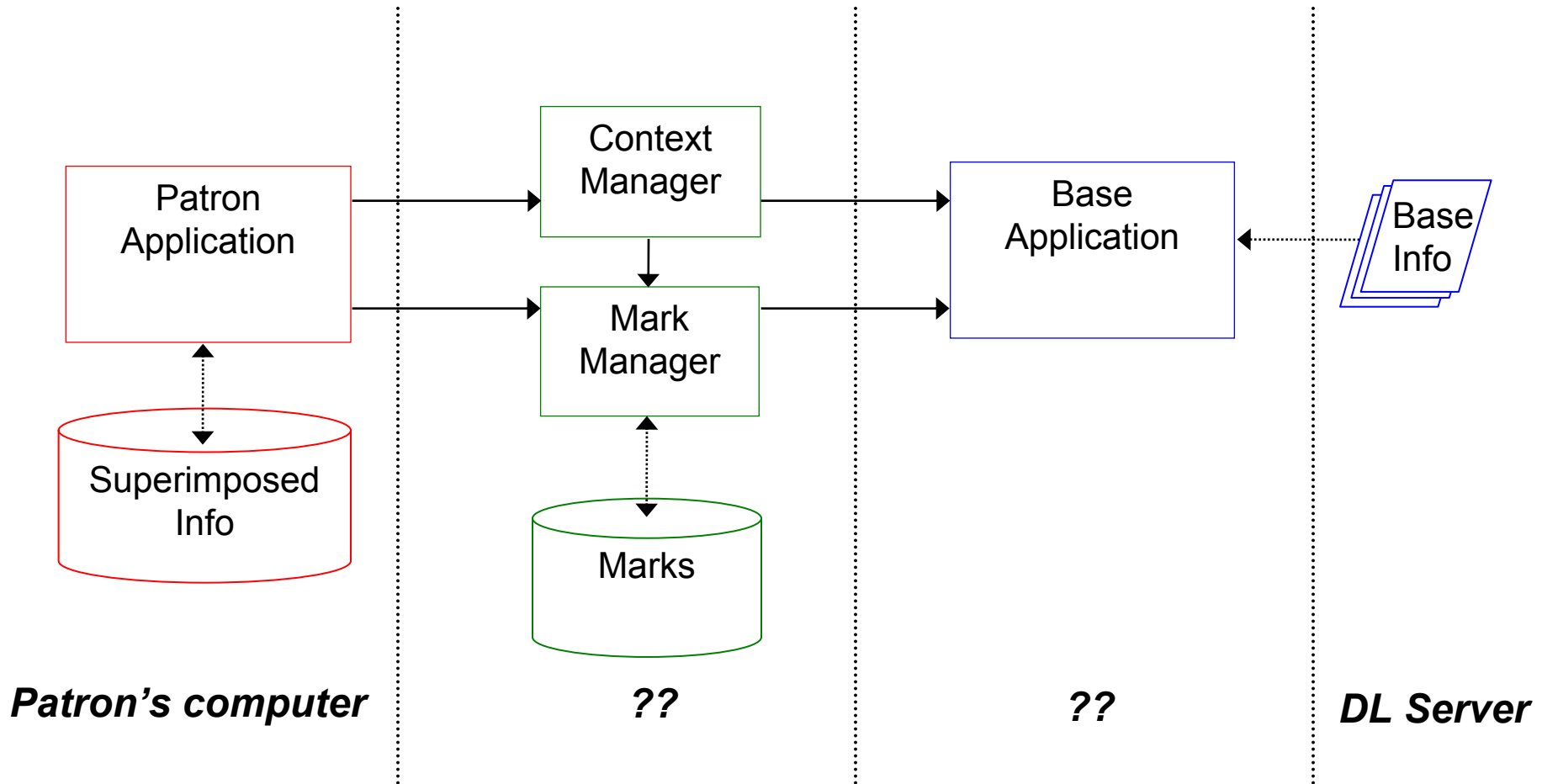
# Distribution of Components

- Distribution affects latency, feature set, maintenance cost, load, security, ...
- Latency example: SOAP-based web service with 400-byte array input and output
  - Local: 7.87 ms
  - LAN (1 hop): 10.53 ms
  - WAN (> 18 hops): **689.39 ms**

# Metrics

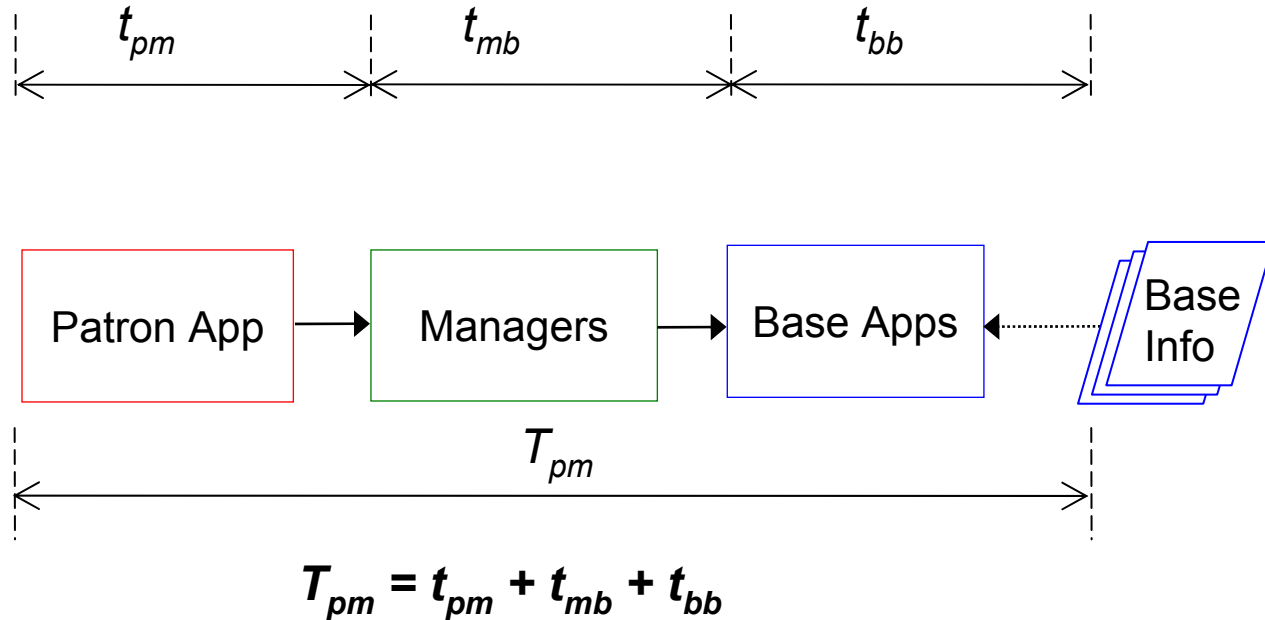
- Latency
  - e.g., Time to serve a patron's request
  - Reduces as *distance* reduces
- Load on server
  - e.g., Number of active processes
  - Increases with the number of components
- Maintenance cost
  - e.g., Effort to update a component
  - Thin clients are less expensive to maintain

# Distributing SPARCE's Components



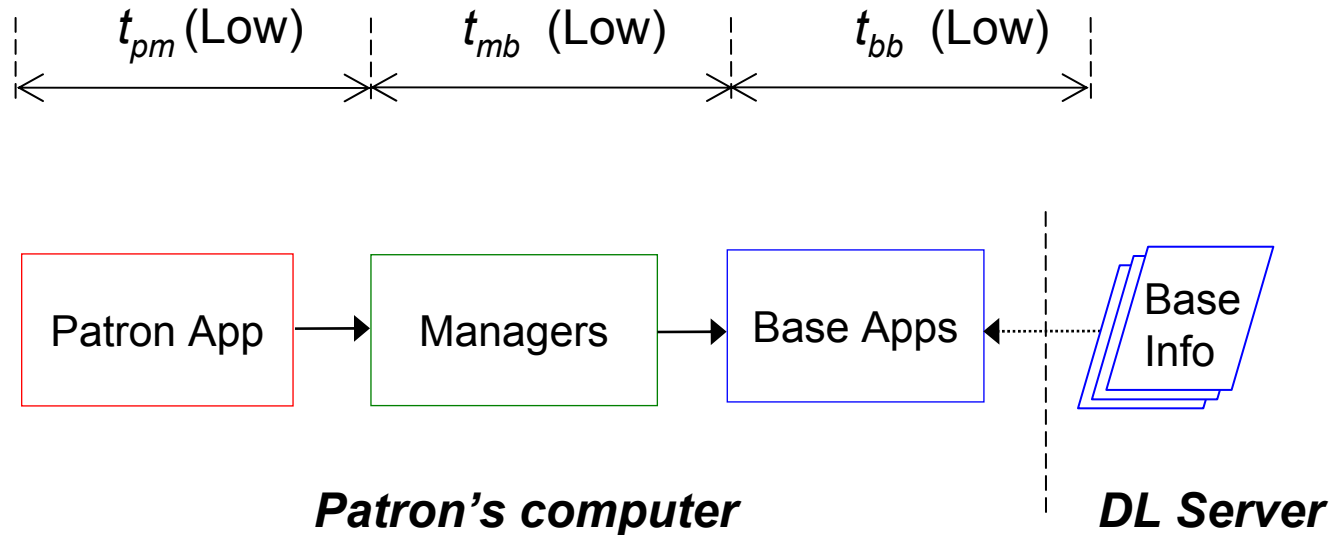


# Latency



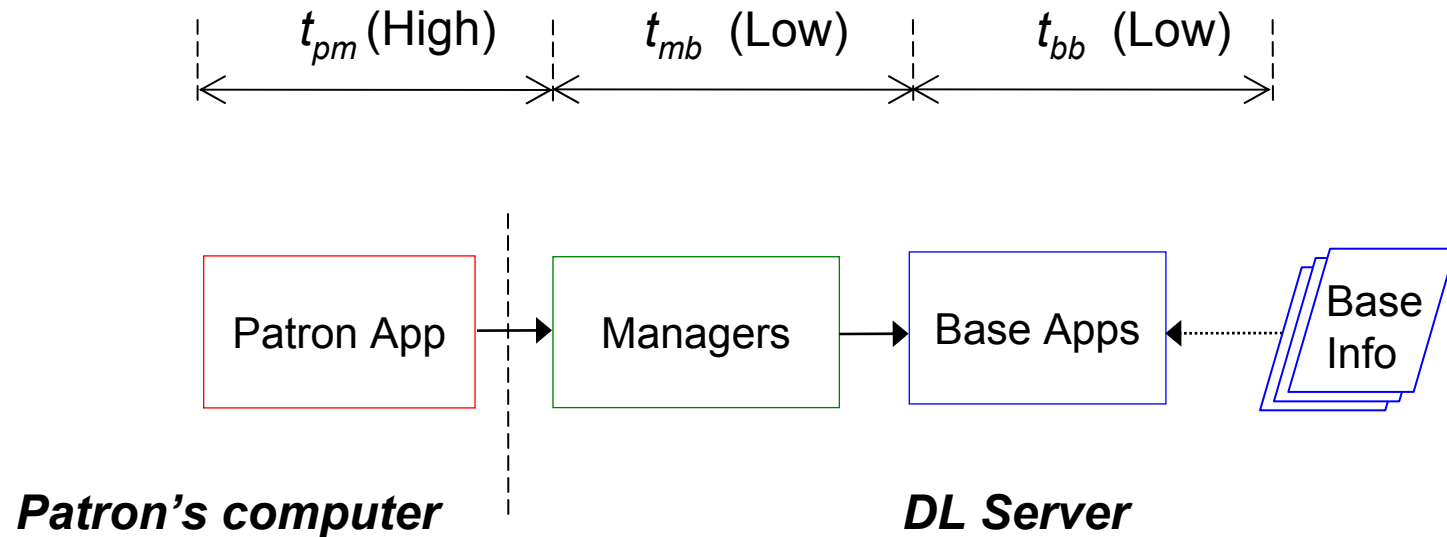
- The overall latency  $T_{pm}$  may be reduced by reducing  $t_{pm}$ ,  $t_{mb}$ , and  $t_{bb}$

# Alternative A



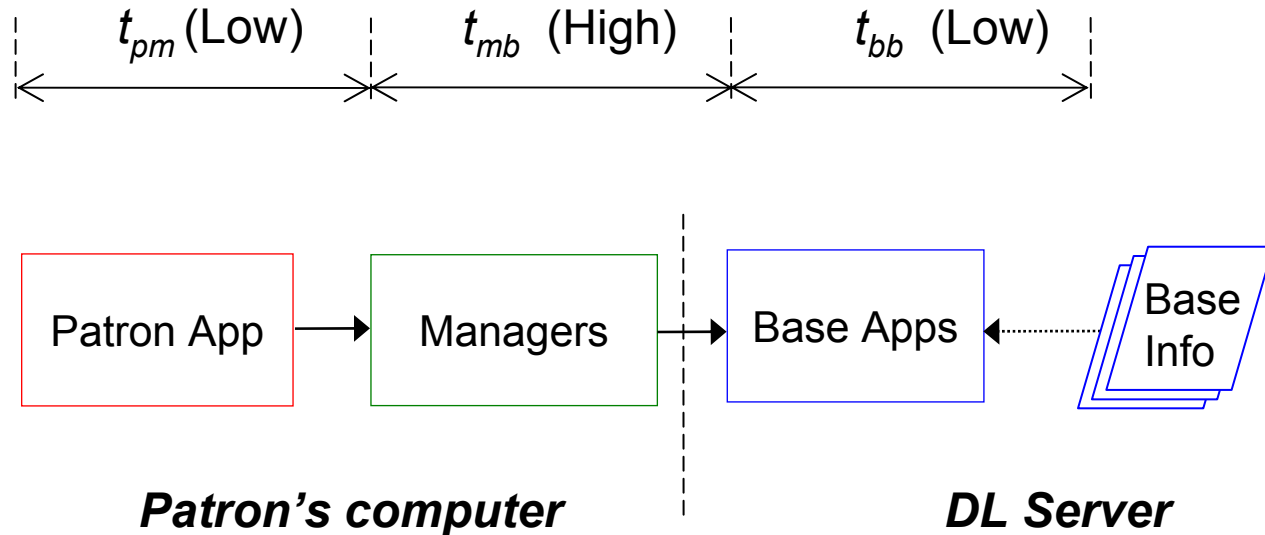
- Base documents are usually downloaded once (per session), so  $t_{bb}$  is likely to be low
- Patron must have all base applications available locally

# Alternative B



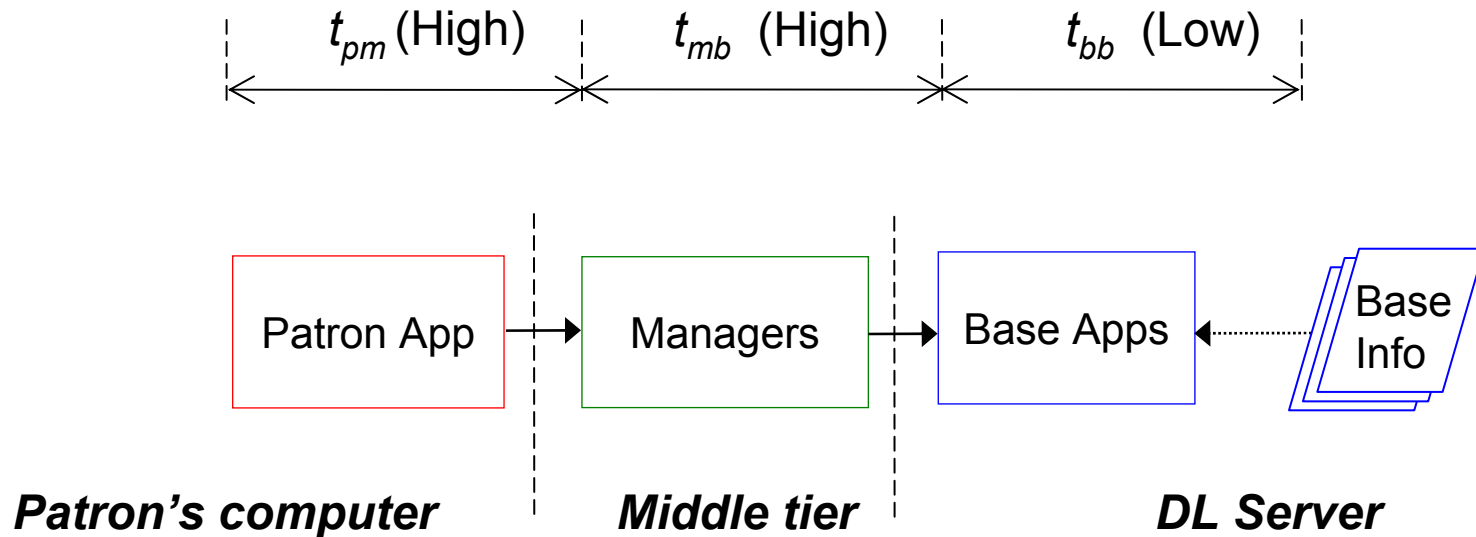
- Minimize number of round trips to managers
- Mark activation would be *unable* to exploit base apps on patron's computer (“call back” may be possible)

# Alternative C



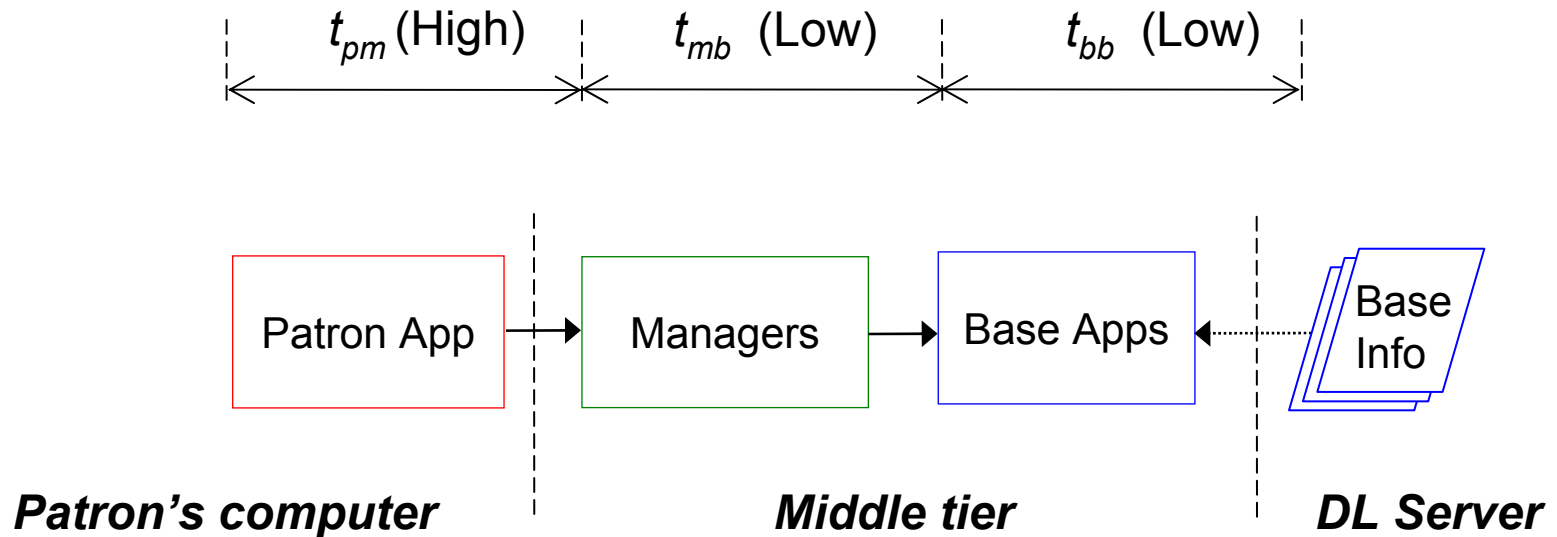
- Minimize number of round trips to base apps
- Mark activation would be *able* to exploit patron's base apps

# Alternative D



- Managers could be shared by more than one DL (server)
- Mark activation would be *unable* to exploit patron's base apps

# Alternative E



- Managers *and* base apps could be shared
- Mark activation would be *unable* to exploit patron's base apps

# Summary of Alternatives

Alternative	Maintenance Cost (Location)			DL Server Load	Round-trip time	
	Patron apps	Managers	Base apps		$t_{pm}$	$t_{mb}$
A	High (P)	High (P)	High (P)	Low	Low	Low
B	Low (P)	Low (D)	Low (D)	High	High	Low
C	High (P)	High (P)	Low (D)	Medium	Low	High
D	Low (P)	Low (M)	Low (D)	Medium	High	High
<b>E</b>	<b>Low (P)</b>	<b>Low (M)</b>	<b>Low (M)</b>	<b>Low</b>	<b>High</b>	<b>Low</b>

Locations: P=Patron's computer, D=DL Server, M=Middle tier

- Prioritize among latency, server load, maintenance cost
- Prioritize the latency terms  $t_{pm}$  and  $t_{mb}$  to minimize

# Discussion

- Heterogeneous distribution alternatives
  - Patron apps of different capabilities; a patron may use DLs that employ different alternatives
- Sharing annotations and marks
  - Marks may be shared or replicated
- Mark manager could be useful for other services such as indexing
- Security
  - Minimize #interface points; narrow functionality



# Improving Performance

- Minimize number of round trips
  - Batch requests where possible
  - #round trips may not be an issue for high bandwidth. e.g., bandwidth between middle tier and DL server likely higher than between patron and DL
- Replication
  - Replicate DL server to handle large loads
  - Instantiate manager modules on patron's computer for local base apps; use DL server at other times
- Cache context

# Conclusion

- Enhancing digital library services with superimposed information services can benefit patrons
- Several distribution alternatives exist, each have trade-offs that DL architects must consider

# Questions?

*Visit*

<http://www.cse.ogi.edu/sparce>

# Related Work

- InfoBus: Roscheisen and others, 1998
- UC Berkeley DL Project: Wilensky, 2000
- ODL: Suleman and Fox, 2001
- PMH: OAI, 2002
- XOAI-PMH: Suleman and Fox, 2002
- FEDORA: Staples and others, 2003
- COM, CORBA, SOAP