

Persistent Identifier Systems

30 September 2009

Larry Lannom

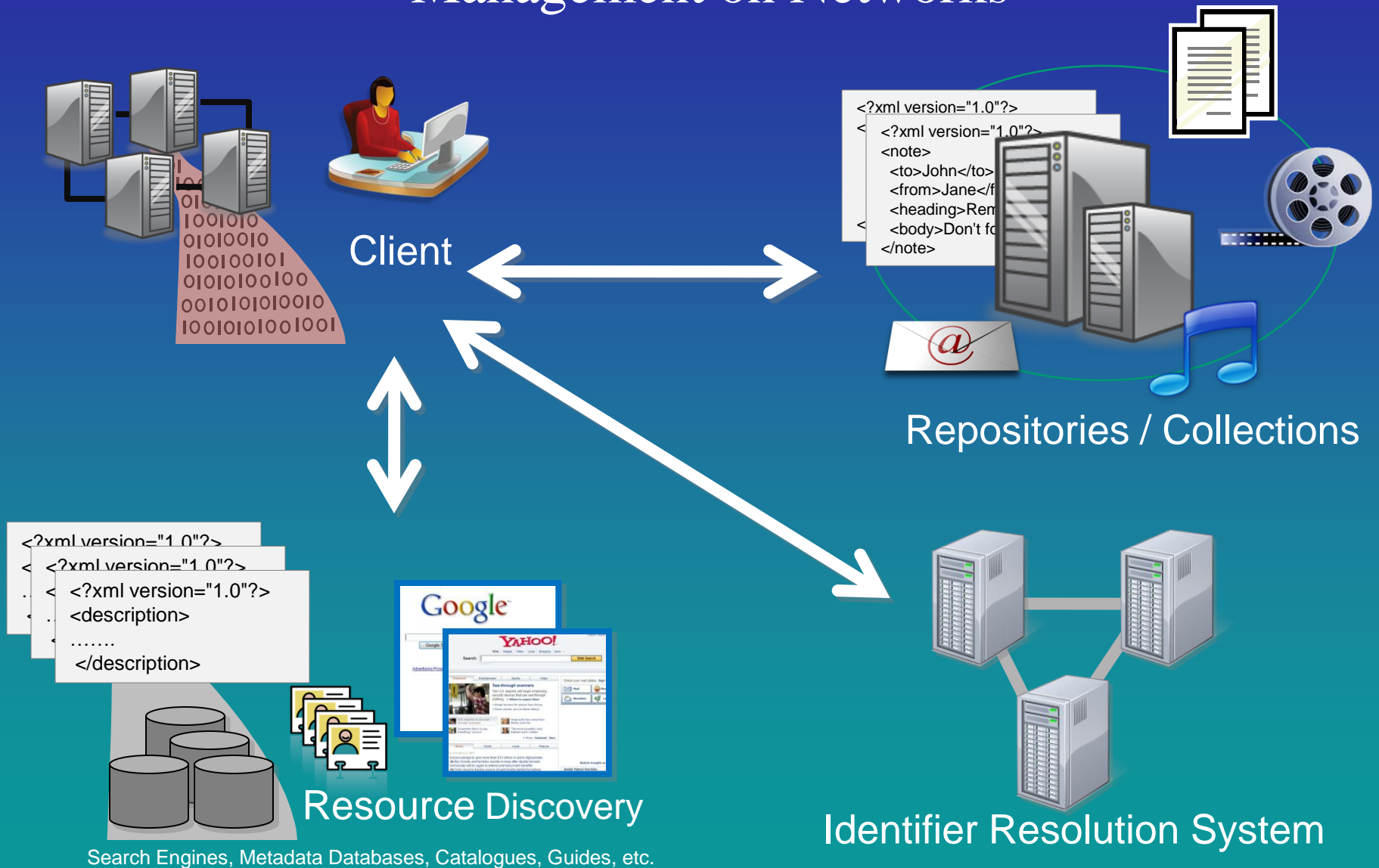
Corporation for National Research Initiatives

<http://www.cnri.reston.va.us/>

<http://www.handle.net/>



Role of Identifier Resolution Systems in Information Management on Networks



Search Engines, Metadata Databases, Catalogues, Guides, etc.

Why Worry About Identifiers?

- Managing increasing amounts of primary and secondary data on the Net over long periods of time
- Managing increasingly complex data relationships on the Net over long periods of time
- When the attributes of that data such as location(s), responsible parties, and the underlying systems may change dramatically over time
- Science builds on past work and increasingly relies on collaboration within virtual distributed communities
- All of this absolutely requires reliable, long-term persistent references to bind together the distributed data, processes, and parties involved

Handle System

- Provides basic identifier resolution system for Internet
 - Go from object name to current state data
 - Name can persist over changes in location and other attributes
- Logically a single system, but physically and organizationally distributed and highly scalable
- Enables association of one or more typed values, e.g., IP address, public key, URL, with each id
- Optimized for speed and reliability
- Secure resolution with its own PKI as an option
- Open, well-defined protocol and data model
- Provides infrastructure for application domains, e.g., digital libraries & publishing, e-research, id mgmt ...

Handle String

- <prefix> / <suffix>
- Examples
 - 10.1525/bio.2009.59.5.9
 - 4263537/5030
- Character Set: Unicode 2.0
- Encoding: UTF-8
- Prefixes
 - Currently allocating only numeric
 - Any text possible

Handles Resolve to Typed Data

HANDLE



10.123/456

DATA
TYPE



URL

URL

DLS

HS_ADMIN

XYZ

HANDLE
DATA



http://acme.com/....

http://a-books.com/....

acme/repository

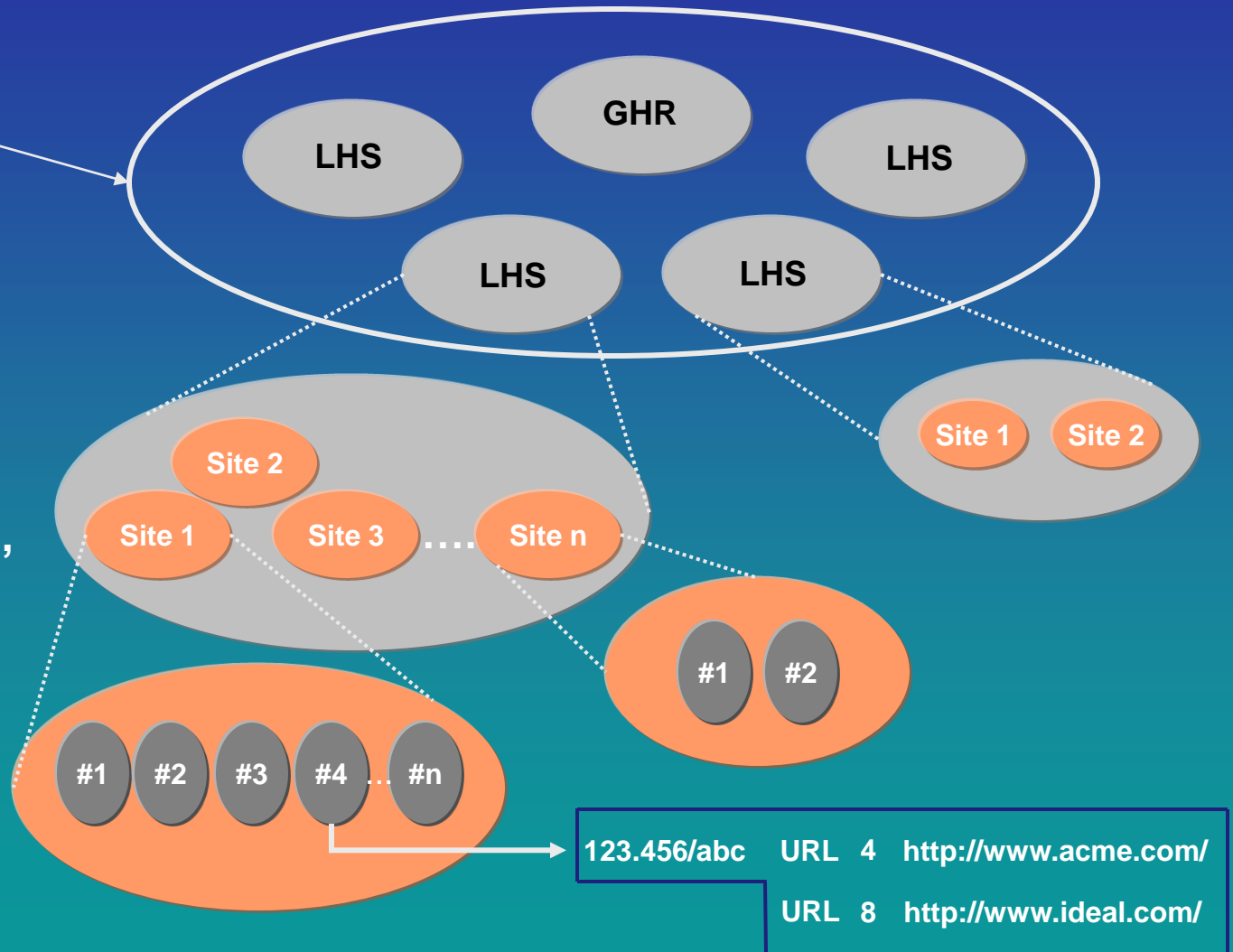
acme.admin/jsmith

1001110011110

Handle Resolution



The Handle System is a collection of handle services, each of which consists of one or more replicated sites, each of which may have one or more servers.



Handle Clients

Request to Client:
Resolve hdl:10.1000/1



Client

1. Sends request to Global to
resolve 0.NA/10.1000
(naming authority
handle for 10.1000)



Global Handle
Registry

Handle Clients

Request to Client:
Resolve hdl:10.1000/1



Client

2. Global Responds with
Service Information for 10.1000



Global Handle
Registry

xcccXV	xC	xC	xC	...
xcccXV	xC	xC	xC	..
xccX	xC	xC	xC	..
xccX	xC	xC	xC	..
xcccXV	xC	xC	xC	..
xccX	xC	xC	xC	..
xccX	xC	xC	xC	..
xcccXV	xC	xC	xC	..
xccX	xC	xC	xC	..
xccX	xC	xC	xC	..

Service Information
Acme Local Handle Service

Handle Clients

XCCCXV	XC	XC	XC	...
XCCCXV XCCX XCCX	XC XC XC	XC XC XC	XC XC XC
XCCCXV XCCX XCCX	XC XC XC	XC XC XC	XC XC XC
XCCCXV XCCX XCCX	XC XC XC	XC XC XC	XC XC XC

	IP Address	Port #	Public Key	...
Primary Site				
Server 1	123.45.67.8	2641	K03RLQ...	...
Server 2	123.52.67.9	2641	5&M#FG...	...
Secondary Site A				
Server 1	321.54.678.12	2641	F^*JLS...	...
Server 2	321.54.678.14	2641	3E\$T%...	...
Server 3	762.34.1.1	2641	A2S4D...	...
Secondary Site B				
Server 1	123.45.67.4	2641	N0L8H7...	...

Service Information - Acme Local Handle Service

Handle Clients

XCCCXV	XC	XC	XC	...
XCCCXV XCCX XCCX	XC XC XC	XC XC XC	XC XC XC
XCCCXV XCCX XCCX	XC XC XC	XC XC XC	XC XC XC
XCCCXV XCCX XCCX	XC XC XC	XC XC XC	XC XC XC

	IP Address	Port #	Public Key	...
Primary Site				
Server 1	123.45.67.8	2641	K03RLQ...	...
Server 2	123.52.67.9	2641	5&M#FG...	...
Secondary Site A				
Server 1	321.54.678.12	2641	F^*JLS...	...
Server 2	321.54.678.14	2641	3E\$T%...	...
Server 3	762.34.1.1	2641	A2S4D...	...
Secondary Site B				
Server 1	123.45.67.4	2641	N0L8H7...	...

Service Information - Acme Local Handle Service

Handle Clients

XCCCXV	XC	XC	XC	...
XCCCXV XCCX XCCX	XC XC XC	XC XC XC	XC XC XC
XCCCXV XCCX XCCX	XC XC XC	XC XC XC	XC XC XC
XCCCXV XCCX XCCX	XC XC XC	XC XC XC	XC XC XC

	IP Address	Port #	Public Key	...
Primary Site				
Server 1	123.45.67.8	2641	K03RLQ...	...
Server 2	123.52.67.9	2641	5&M#FG...	...
Secondary Site A				
Server 1	321.54.678.12	2641	F^*JLS...	...
Server 2	321.54.678.14	2641	3E\$T%...	...
Server 3	762.34.1.1	2641	A2S4D...	...
Secondary Site B				
Server 1	123.45.67.4	2641	N0L8H7...	...

Service Information - Acme Local Handle Service

Handle Clients

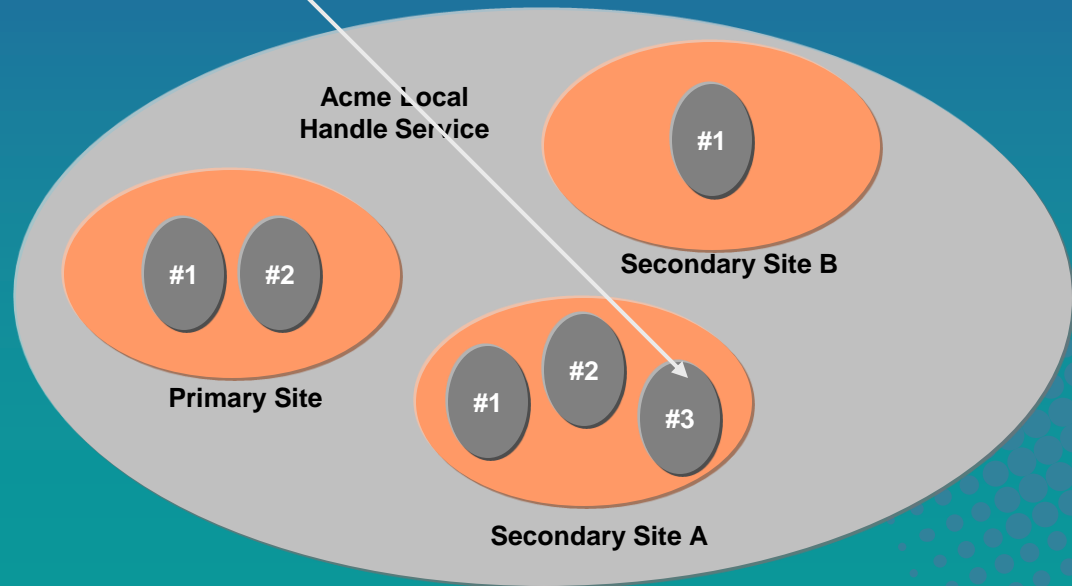
Request to Client:
Resolve hdl:10.1000/1



Client

3. Client queries Server 3
in Secondary Site A
for 10.1000/1

Global Handle
Registry



Handle Clients

Request to Client:
Resolve hdl:10.1000/1

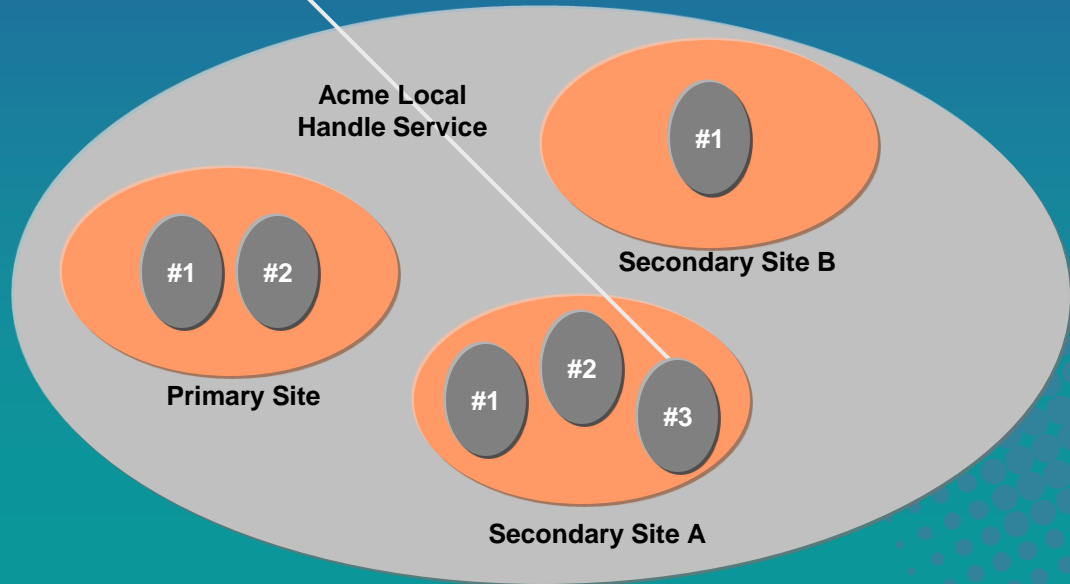


Client

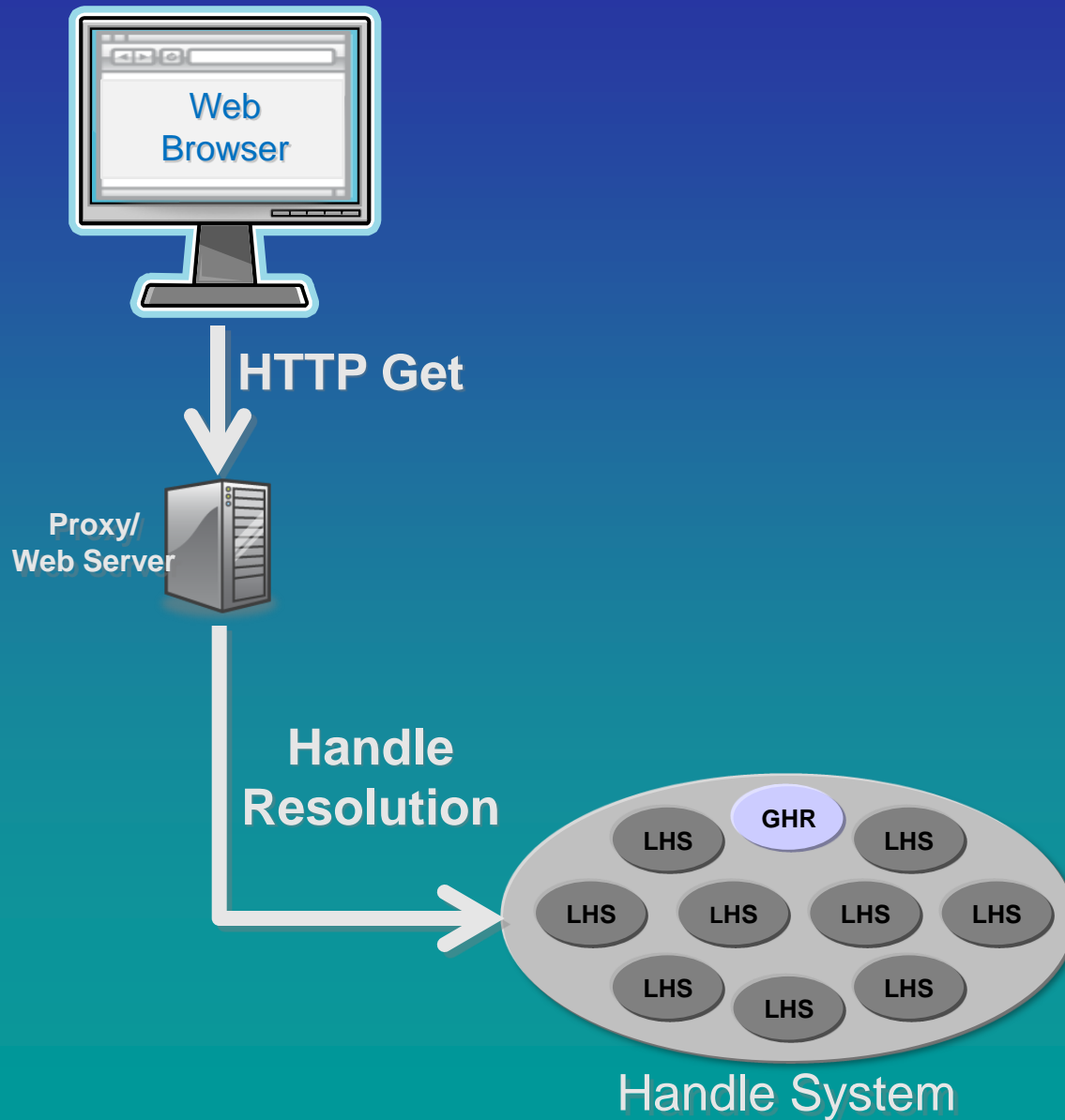


Global Handle Registry

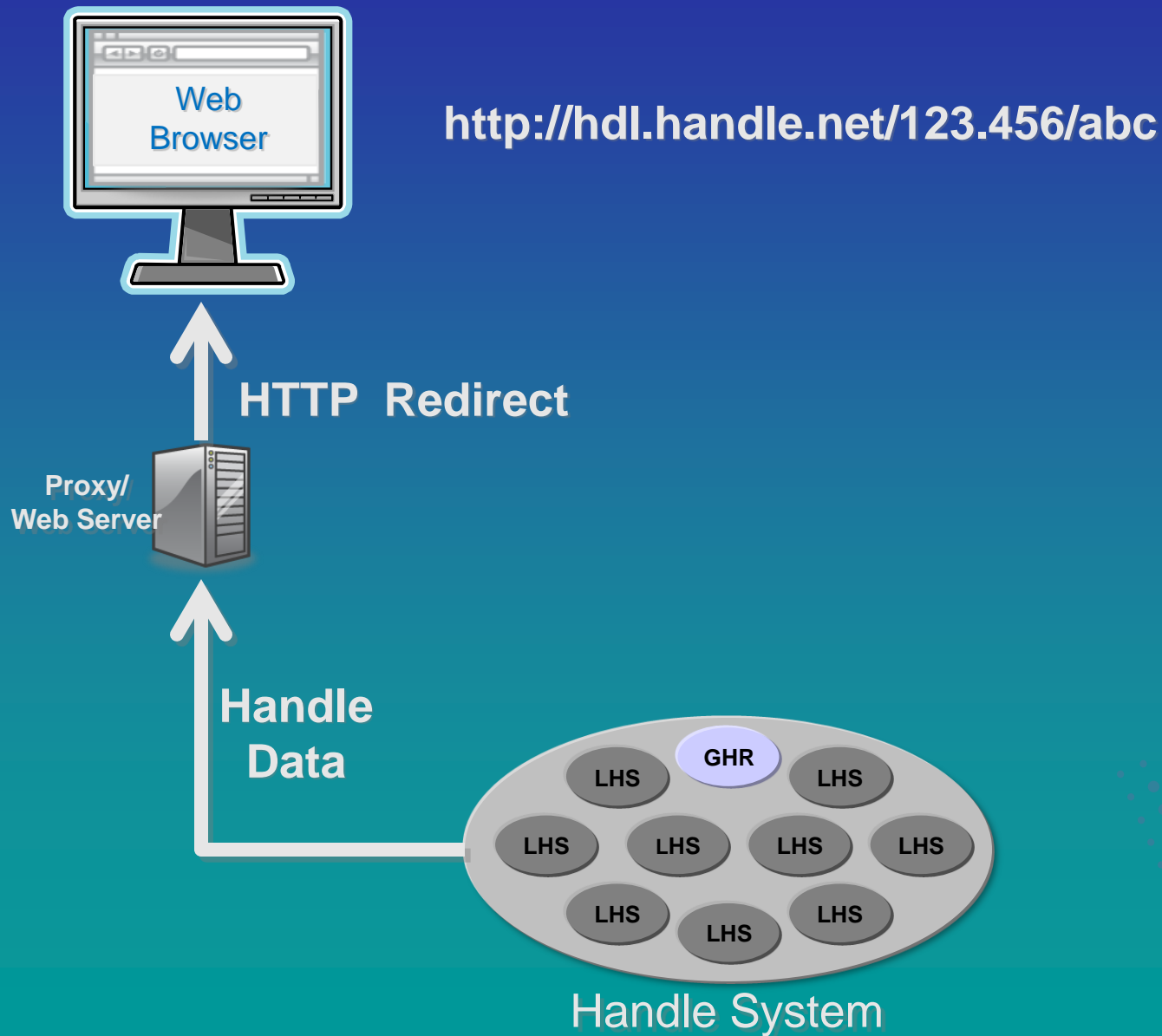
4. Server responds with
handle data



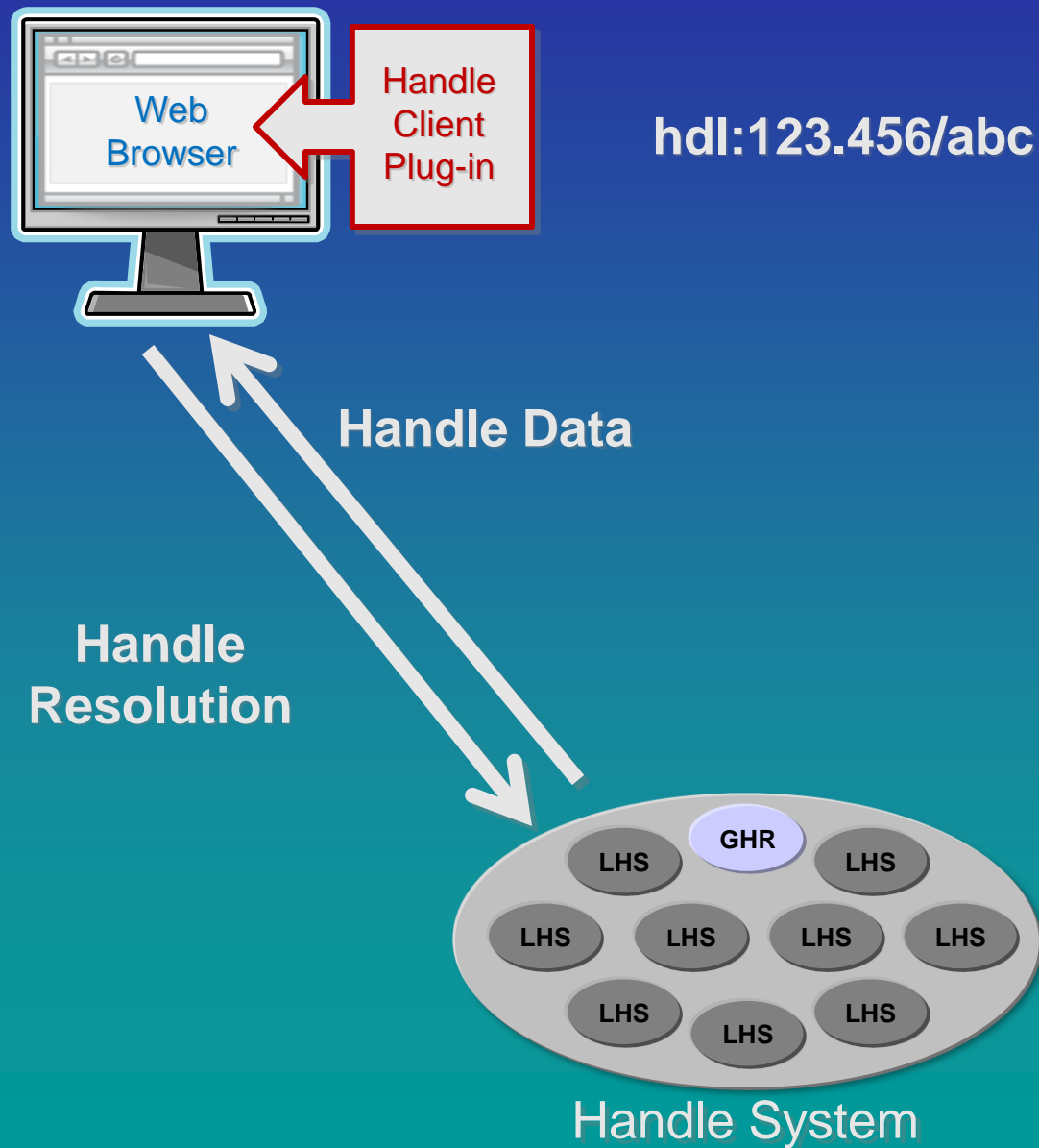
Handle Clients



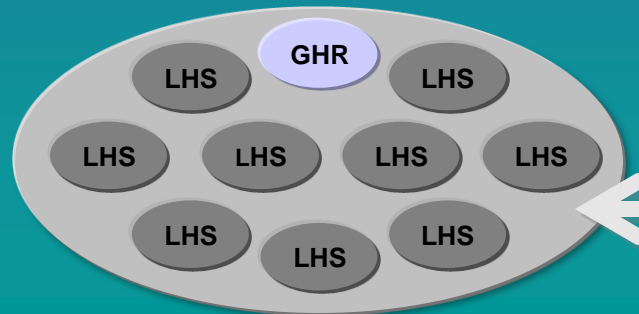
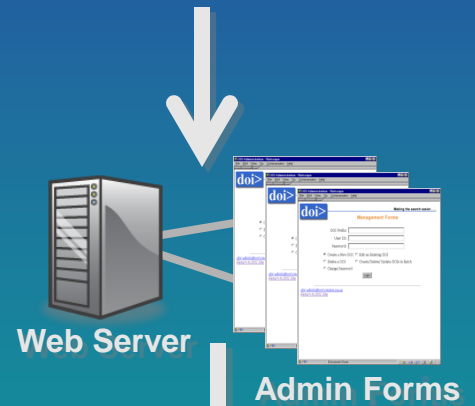
Handle Clients



Handle Clients



Handle Clients

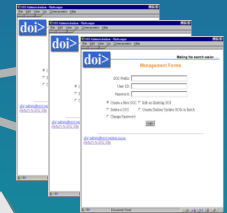


Handle System

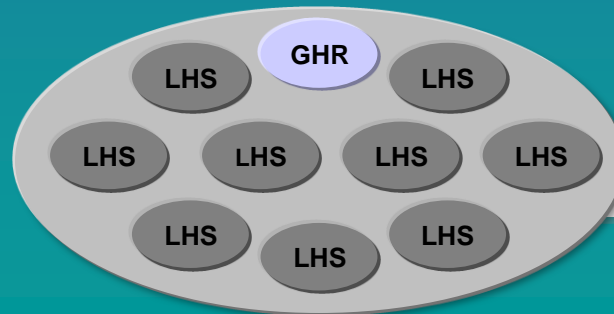
Handle Clients



HTTP



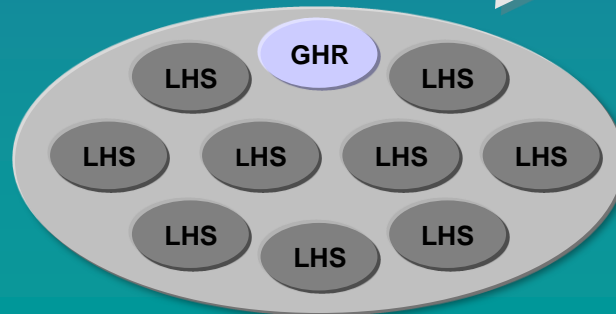
Admin Forms



Handle System

Handle Admin

Handle Clients



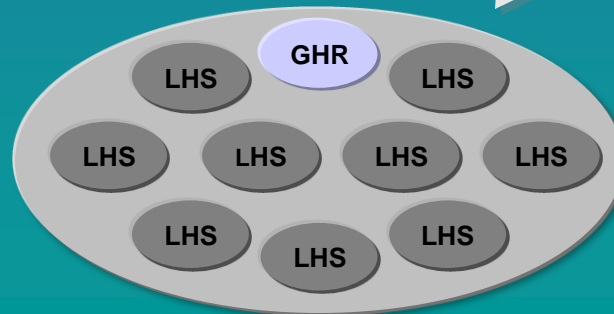
Handle System



Handle Clients

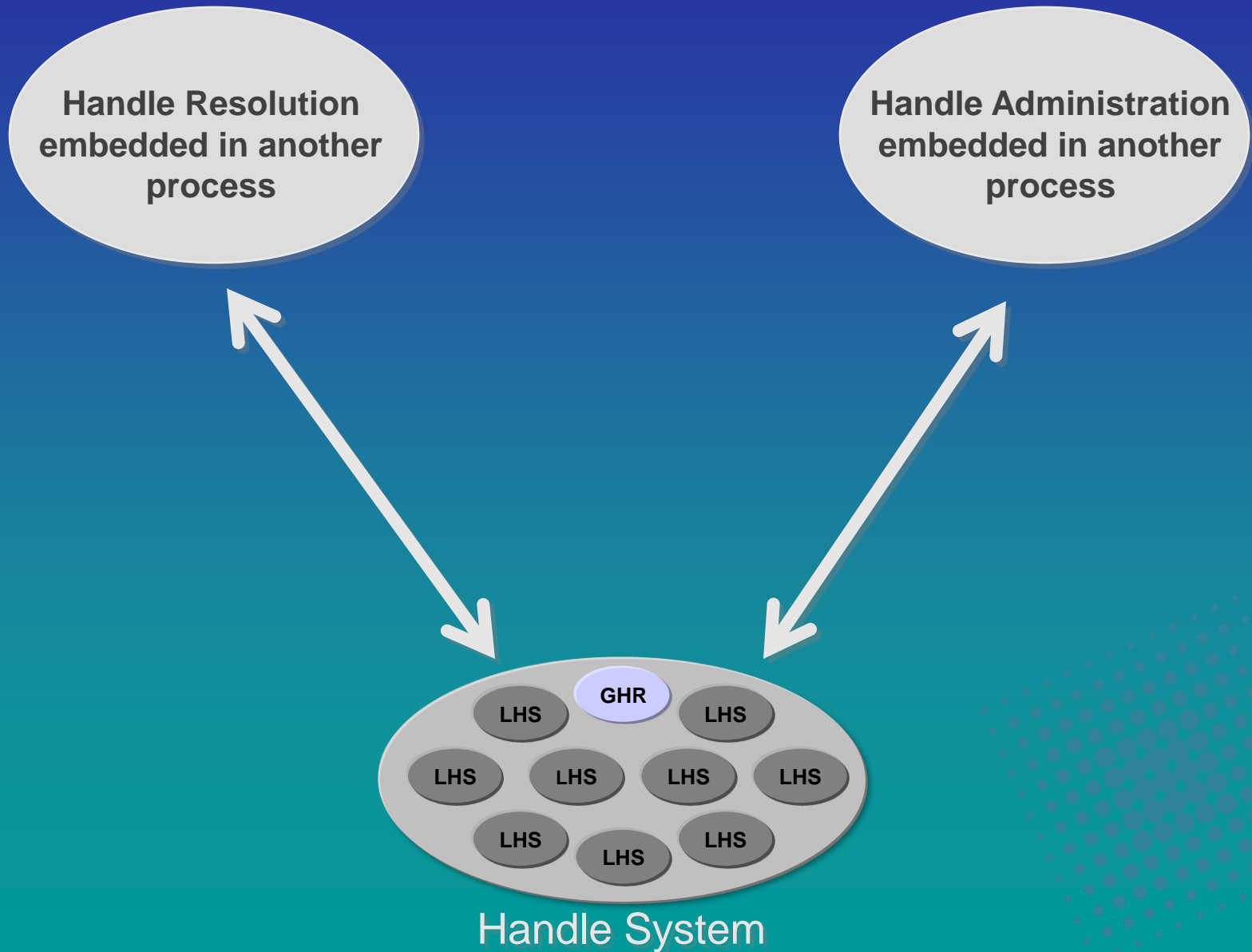


Handle Administration
embedded in another
process



Handle System

Handle Clients



Handle System Software

- Server (v6.2)
 - Java 1.4.2 and higher
- Client Library
 - Java & C versions available
- Proxy servlet
 - Java servlet, typically runs under Apache Tomcat
 - Build your own or use hdl.handle.net
- Misc. CNRI software (admin tools, browser plug-ins, etc.)
- Misc. community software (alternate clients, database modules, etc.)
- All available at www.handle.net
- Alternate complete implementations
 - Two known to CNRI, neither public
 - Both developed from spec, but they talked to us

Using a Resolution System with Existing Identifiers

- No lack of identifiers in the world
- Actionable ISBN scheme
 - Example: 10.97812345/99990
 - The syntax specification, reading from left to right, is:
 - Handle System DOI name prefix = "10."
 - ISBN (GS1) Bookland prefix = "978." or "979."
 - ISBN Publisher prefix = variable length numeric string of 2 to 8 digits
 - Prefix/suffix divider = "/"
 - ISBN Title enumerator and checkdigit = variable length numeric string of 8 to 2 digits

Template Handles

- An unlimited number of handles are computed on the fly from a single registered template
- Re-write rules and delimiter can be defined at the prefix level, e.g., use ‘-’ as delimiter and re-write any URL values, e.g., for any handle under the prefix 123
- Any handle under that prefix can be divided into base and extension, e.g., 123/456-abc has a base of 123/456 and an extension of abc. The base is registered.
- The data at 123/456 will then be combined with the extension string (abc) using the re-write rule
- Resolve “123/456-abc” and get back <http://repository.com/getobject?id=123/456&part=abc>
- Resolve “123/456-def” and get back <http://repository.com/getobject?id=123/456&part=def>
- Implemented and available in next handle release

Template Handles

- Directly results from modularity of the current implementation
 - Backend handle storage is pluggable
 - A new storage module allows handles to be computed
 - The rest of the handle resolution mechanisms are unchanged, only the storage module was enhanced
- Any exception handles can be individually registered to over-ride the template
- Re-write rules at the base level will over-ride the prefix level rules
- Re-write rules use Java regular expression language
- Templates allow handle strings to remain static in reference form while millions of resolution values can be changed at a single stroke

Multiple Resolution: “Chooseby”

- Structured alternatives, e.g., multiple locations, in a single handle value
- Include selection criteria in that same value
- Handle client application, e.g., proxy server, performs evaluation
- Type = 10320/loc; value =

```
<locations chooseby="locatt, country, weight">  
  <location id=0 href="http://abc... Country="gb" weight=0>  
  <location id=1 href="http://def... weight=1>  
  <location id=2 href="http://xyz... weight=1>  
</locations/>
```
- If the user is in the UK they are redirected to <http://abc...>, if not then either <http://def...> or <http://xyz...> at random, 50/50
- Currently deployed in CNRI-run proxies and will be included in the next release
- Approach extensible for future selection methods, e.g., chooseby language or other value known to the proxy

Multiple Resolution: “Chooseby”

10.1525/bio.2009.59.5.9	URL	http://caliber.ucpress.net/doi/abs/10.1525/bio.2009.59.5.9
	HS_ADMIN	handle=0.na/10.1525; index=200; [delete hdl,add val,read val,modify val,del admin,add admin,list]
	10320/loc	<locations chooseby="locatt, country, weighted"> <location id="1" cr_type="MR-LIST" href="http://mr.crossref.org/iPage?doi=10.1525%2Fbio.2009.59.5.9" weight="1" /> <location id="2" cr_src="unca" label="SECONDARY_BIOONE" cr_type="MR-LIST" href="http://www.bioone.org/doi/full/10.1525/bio.2009.59.5.9" weight="0" /> </locations>

The evaluation falls through the first two criteria and the proxy uses 'weighted' as the selection criteria. The first location (<http://mr.crossref.org>) wins with a weight of 1. That location goes to a script on the CrossRef site that builds the page a user sees when resolving the DOI name as <http://dx.doi.org/10.1525/bio.2009.59.5.9>. The page is built to include the original URL value plus the 10320/loc data plus some additional information held by CrossRef.

Multiple Resolution: “Chooseby”

The page displayed includes both the original URL and the added BioOne link:

Current Links for BioScience doi:10.1525/bio.2009.59.5.9 - Internet Explorer provided by Dell

http://mr.crossref.org/iPage/?doi=10.1525%2Fbio.2009.59.5.9

File Edit View Favorites Tools Help

Convert Select

Favorites Current Links for BioScience doi:10.1525/bio.2009...

BioScience

Motivating Online Publication of Data
Mark J. Costello
BioScience (2009),59(5):418
doi:10.1525/bio.2009.59.5.9

Available at the following locations:

CALIBER
journals of the university of california press

BioOne™ RESEARCH EVOLVED

American Institute of Biological Sciences

Since 1964, *BioScience* has presented readers with timely and authoritative overviews of current research in biology, accompanied by essays and discussion sections on education, public policy, history, and the conceptual underpinnings of the biological sciences.

www.aibs.org

[Caliber](http://caliber.ucpress.net/doi/abs/10.1525/bio.2009.59.5.9)

TYPE = URL

VALUE = <http://caliber.ucpress.net/doi/abs/10.1525/bio.2009.59.5.9>

[BioOne](http://www.bioone.org/doi/full/10.1525/bio.2009.59.5.9)

TYPE = 10320/loc

VALUE = <http://www.bioone.org/doi/full/10.1525/bio.2009.59.5.9>