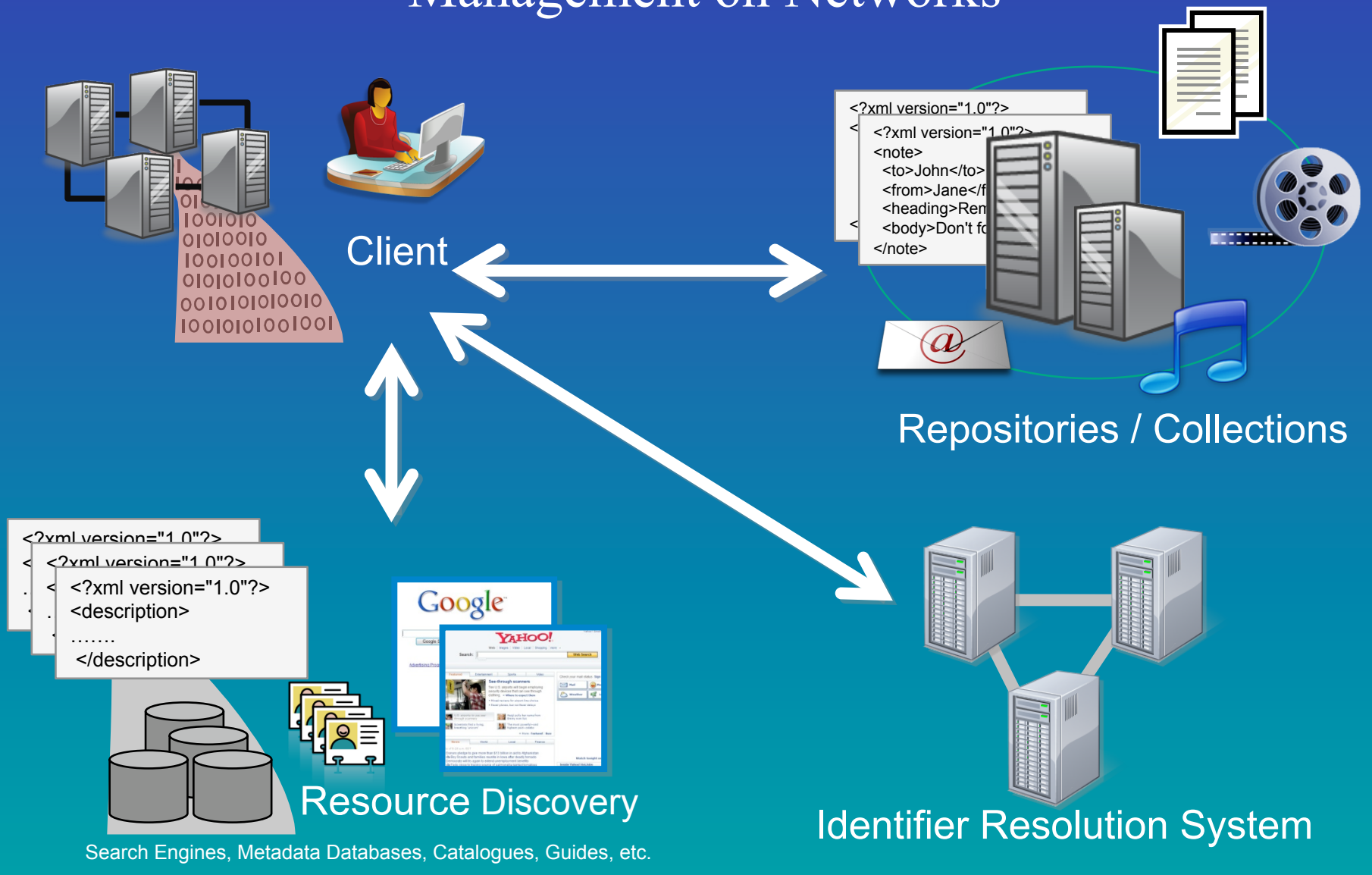
Identifier Systems in Network Architecture

11 June 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



Requirements: Identifier String

- Not based on any changeable attributes of the entity
  - Location
  - Ownership
  - Any other attribute that may change w/o changing identity
- Opaque, preferably a ‘dumb number’
  - A well known pattern invites assumptions that may be misleading
  - Meaningful semantics invite IP wars, language problems
- Unique
  - Avoid collisions, referential uncertainty
- Nice to have
  - Human-readable
  - Cut-able, paste-able, embeddable
  - Fits common systems, e.g., URI specification
- All of the above contributes to persistence

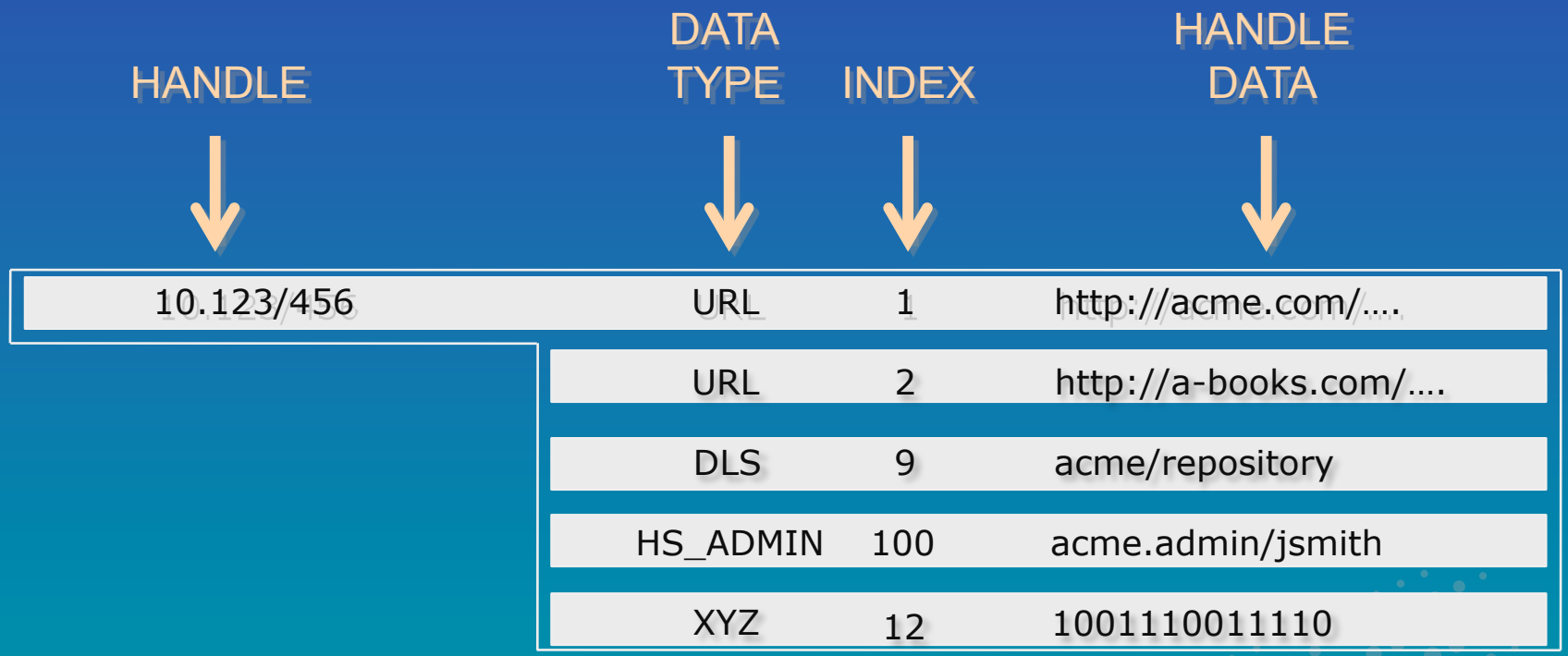
Requirements: Identifier Resolution System

- Reliable
  - Redundant, no single points of failure
  - Fast enough to not appear broken
- Scalable
  - Higher loads managed with more computers
- Flexible
  - Adapt to changing computing environments
  - Useful to new applications
- Trusted
  - Resolution/Administration must be trusted
  - Organization must be committed to the long term
- Open Architecture
  - Leverage efforts of a community in building apps on your infrastructure
- Transparent
  - Users knowing the id/infrastructure NOT a good feature
- Persistence, again

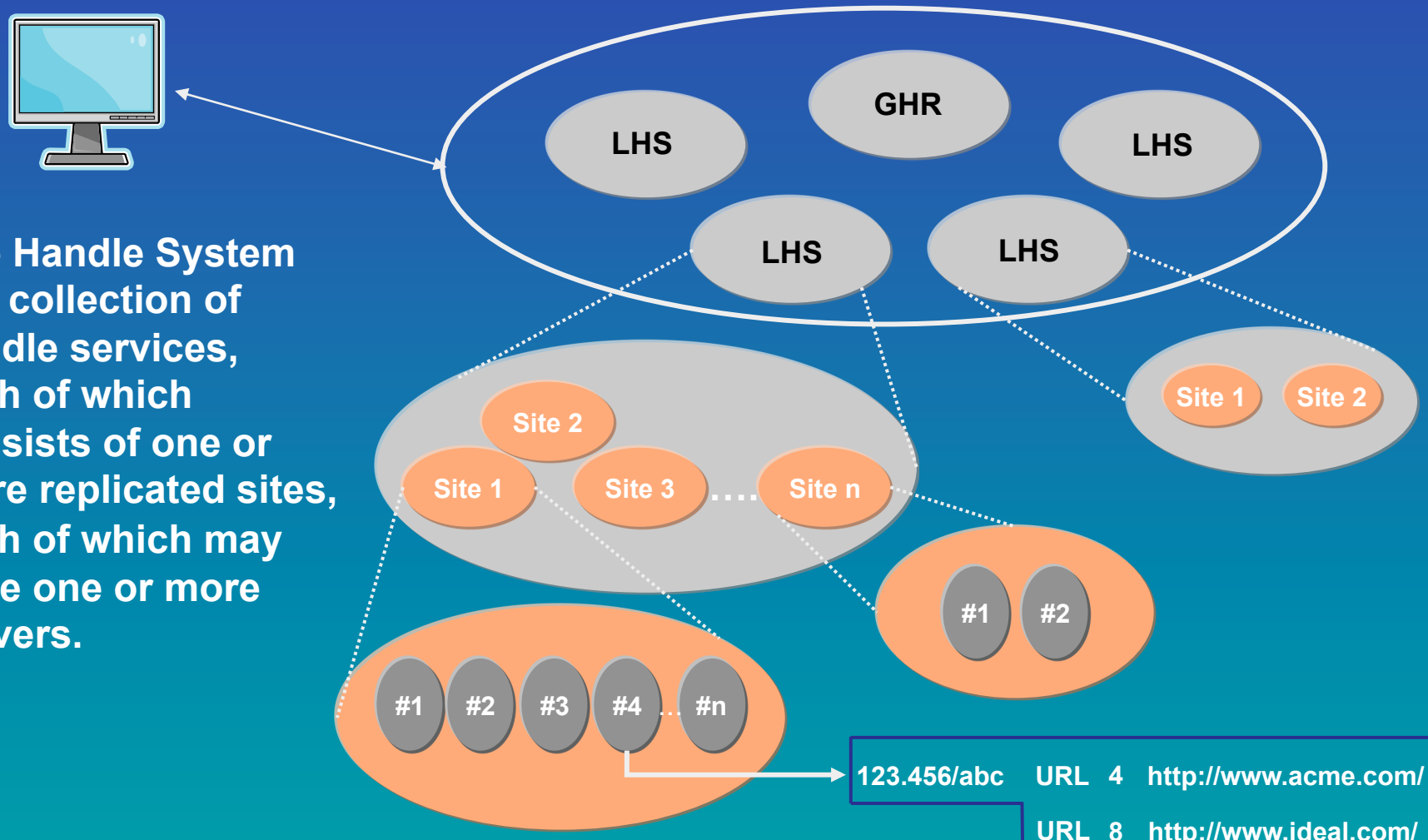
Handle System

- 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 centralized, 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, network mgmt, id mgmt ...

Handles Resolve to Typed Data



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

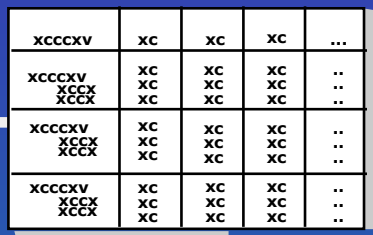
Client

2. Global Responds with
Service Information for 10.1000

Global Handle Registry

xcccxv	xc	xc	xc	...
xcccxv 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	.. .. ..

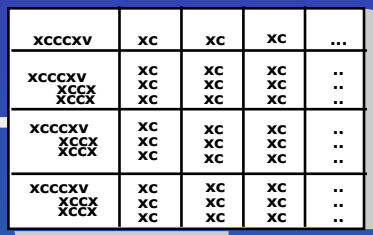
Service Information
Acme Local Handle Service



Handle Clients

	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…**	**…**

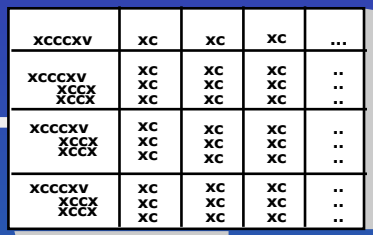
Service Information - Acme Local Handle Service



Handle Clients

	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



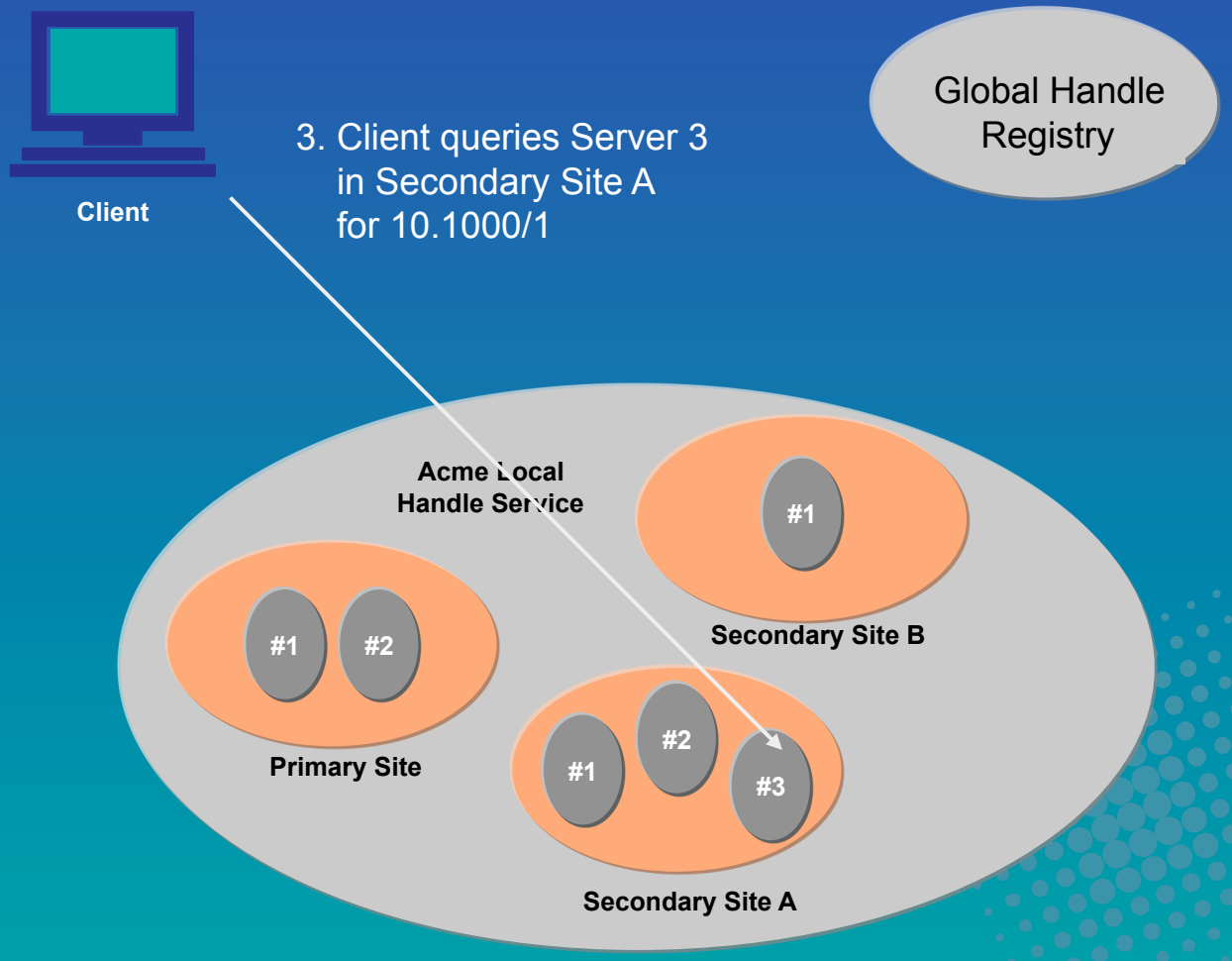
Handle Clients

	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



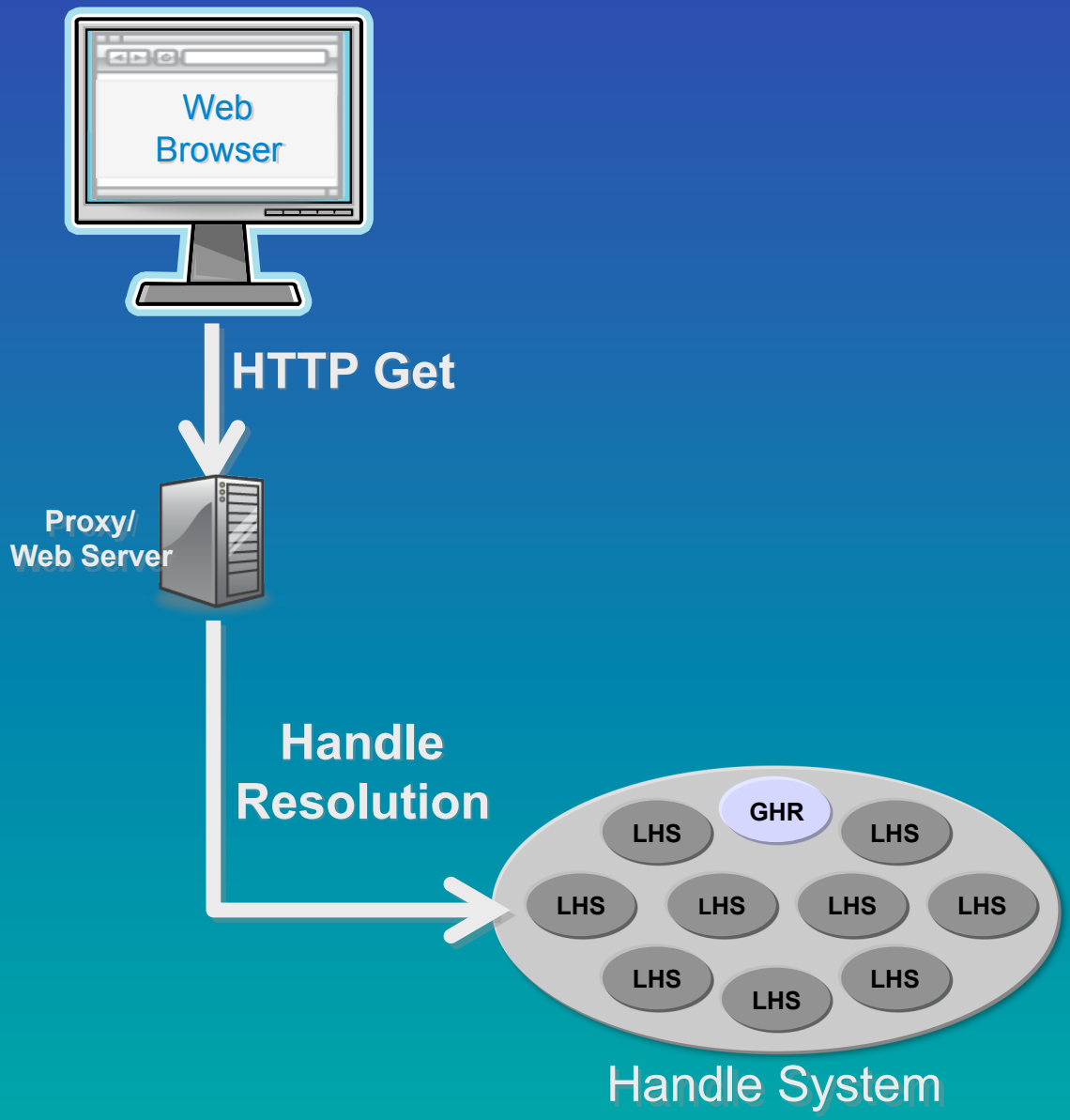
Handle Clients

Request to Client:
Resolve hdl:10.1000/1

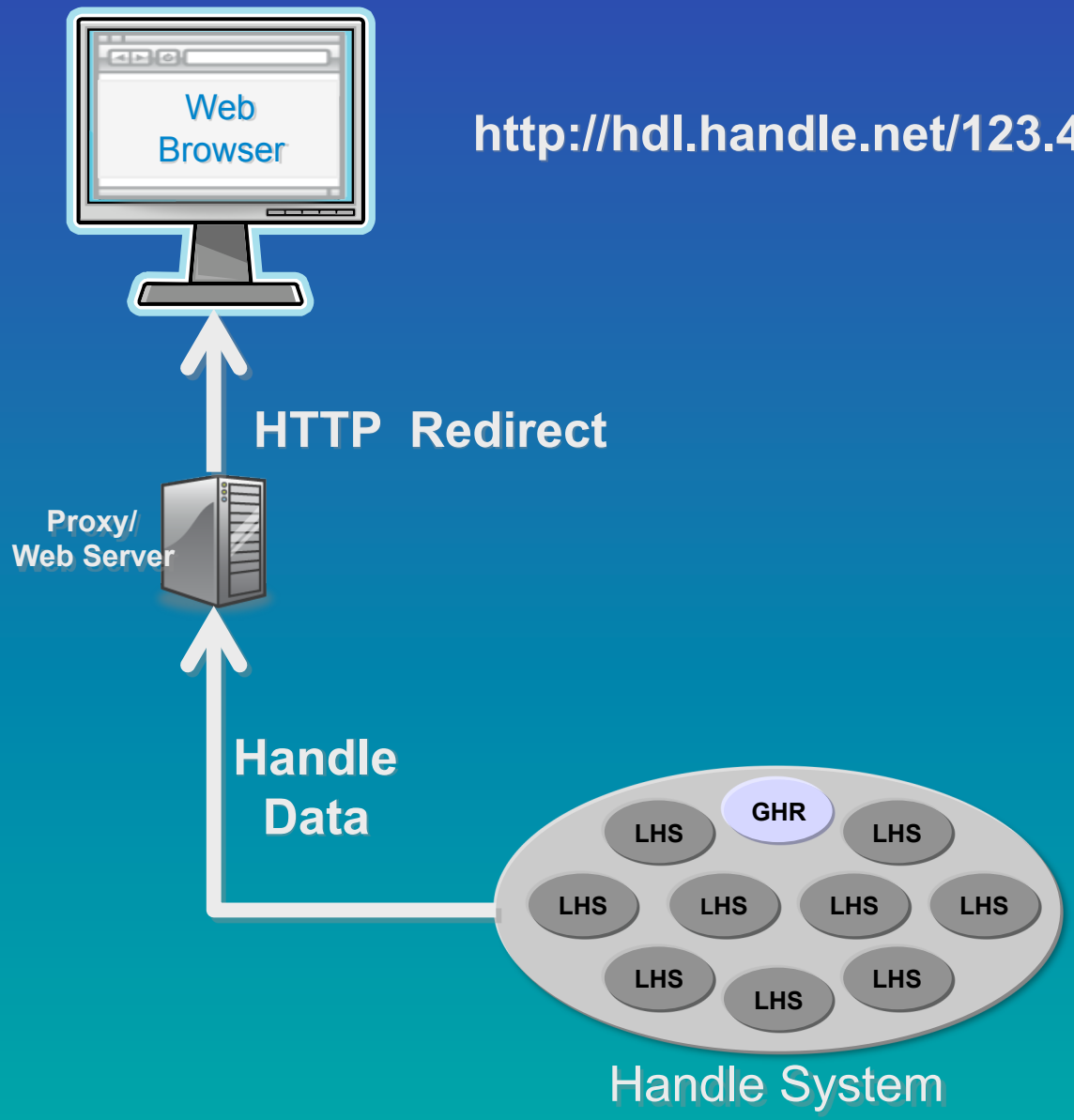
Global Handle Registry

4. Server responds with handle data

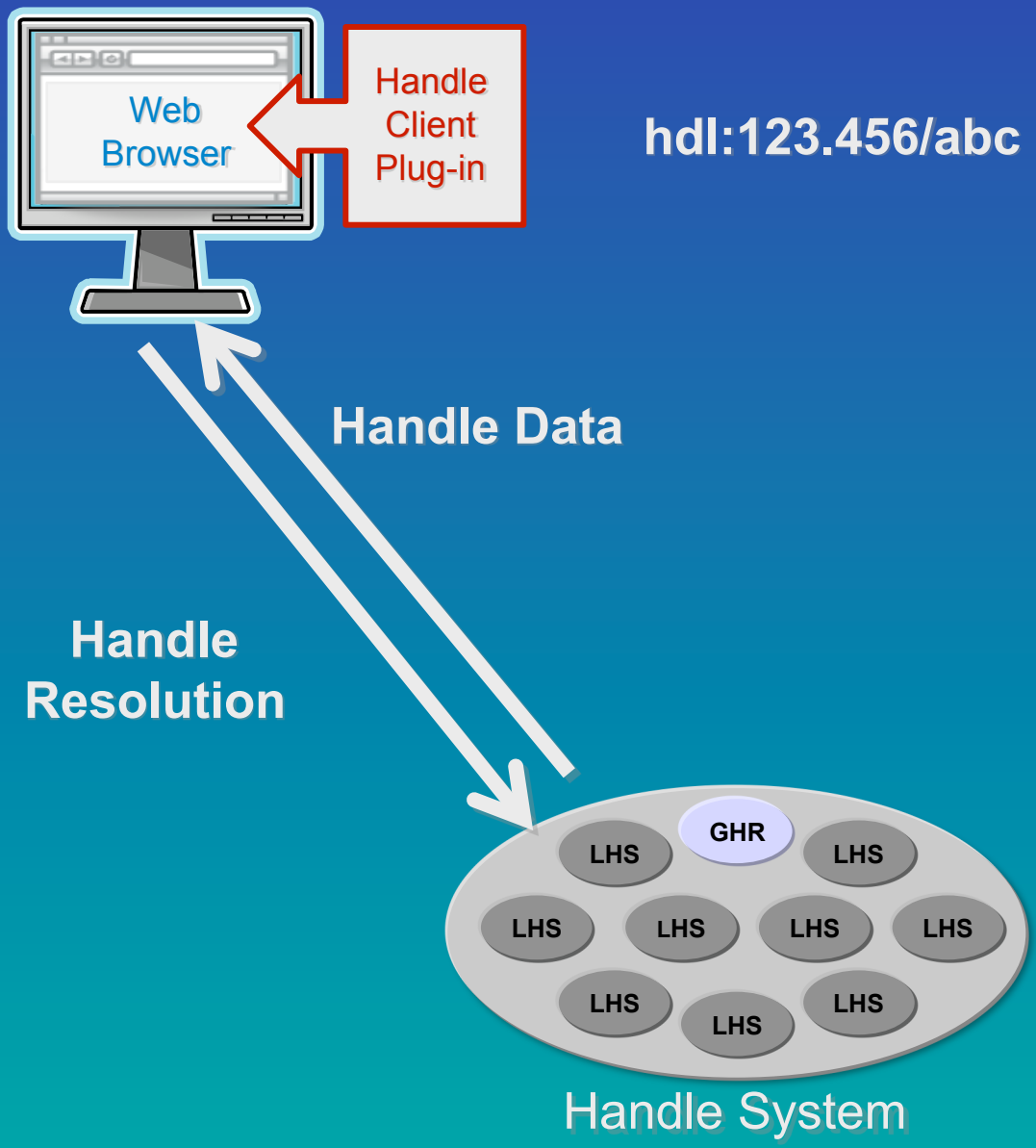
Handle Clients



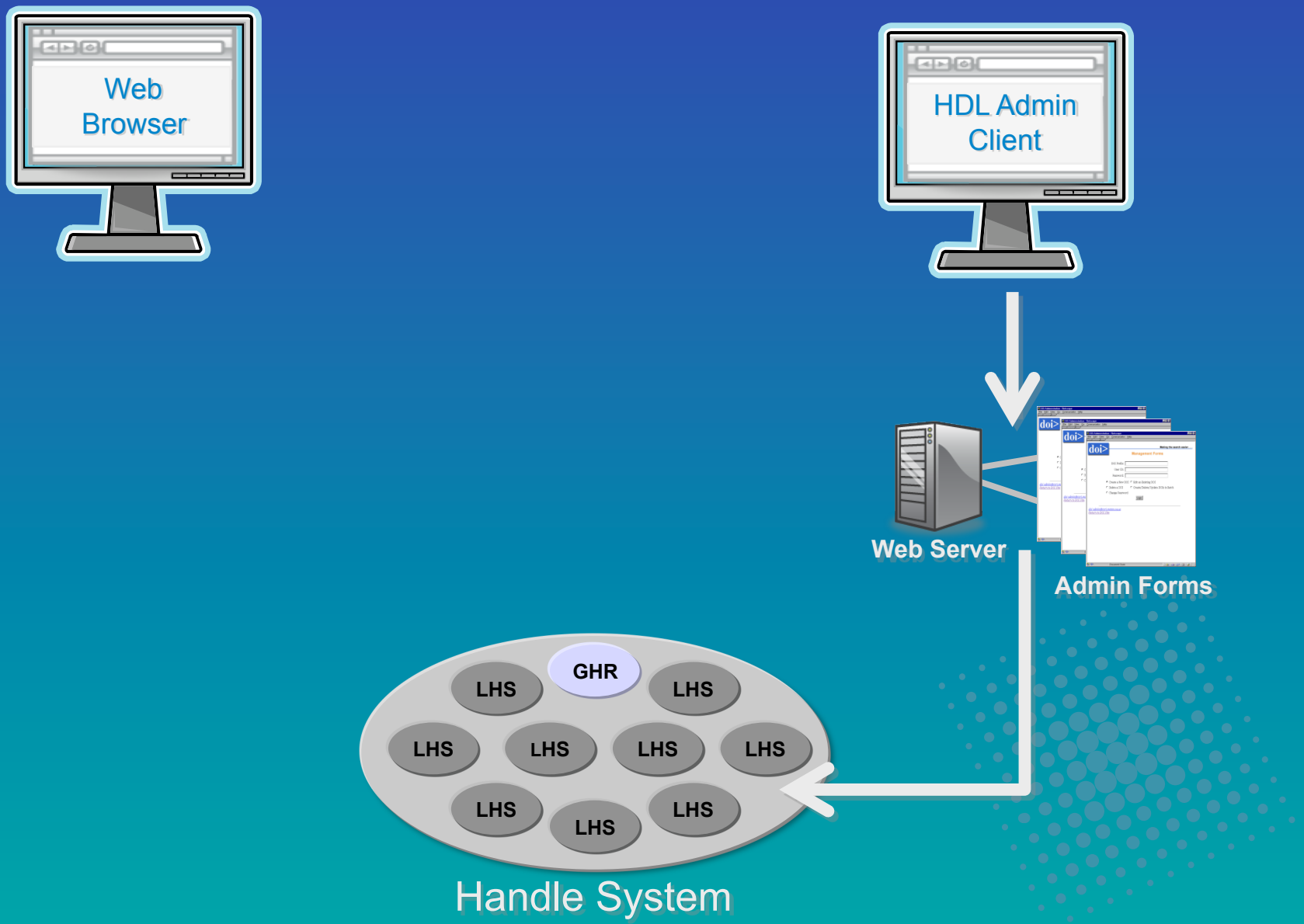
Handle Clients



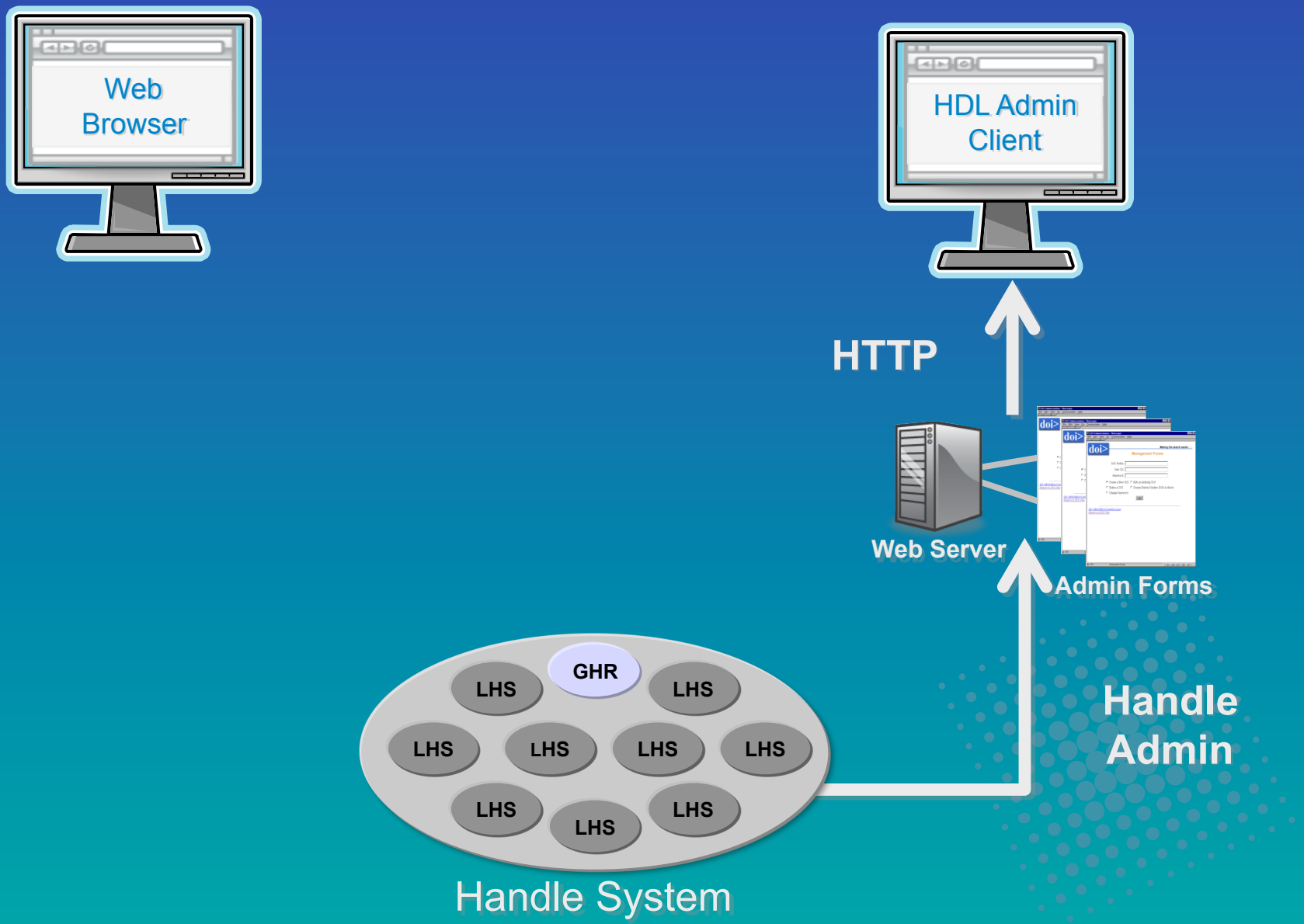
Handle Clients



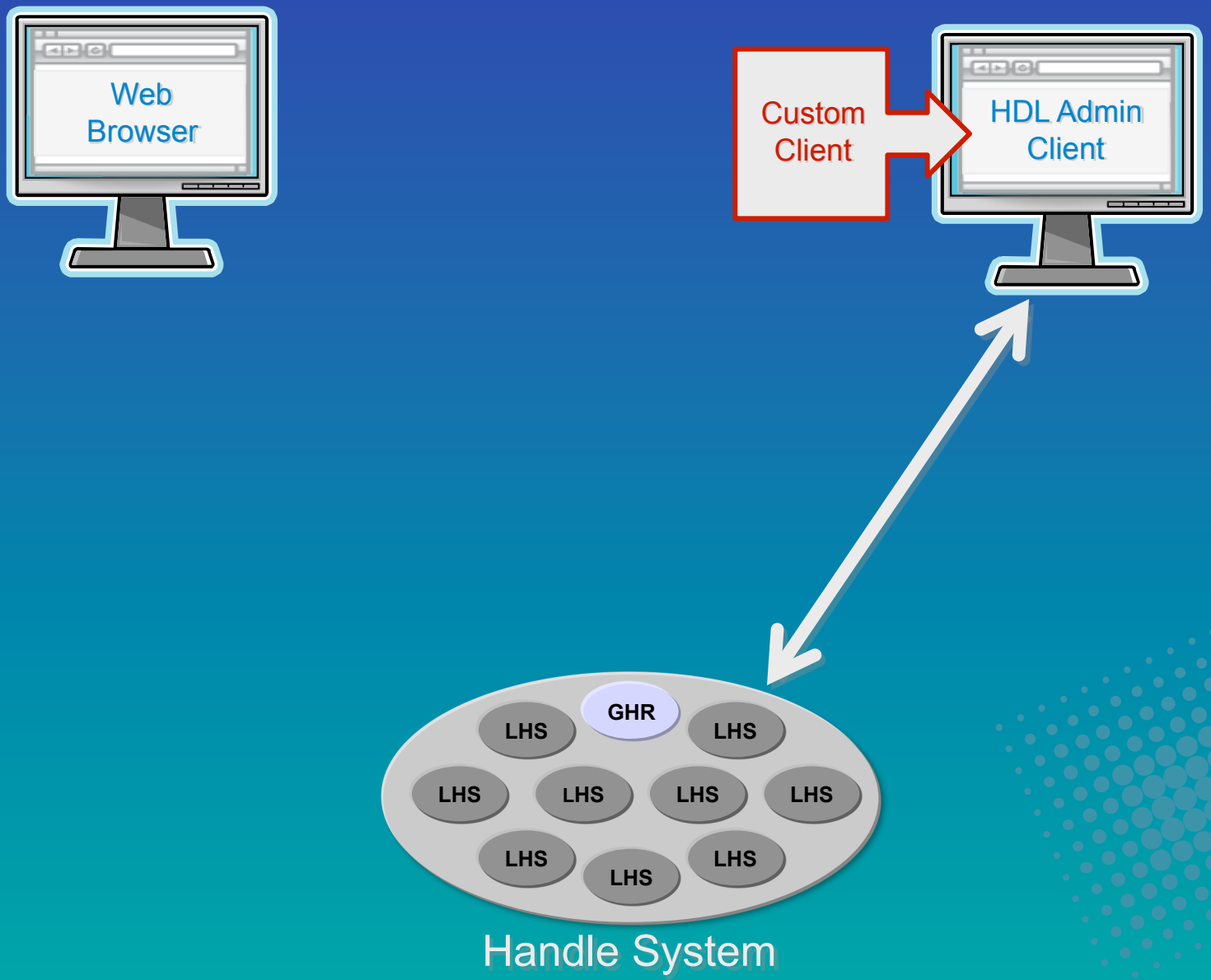
Handle Clients



Handle Clients

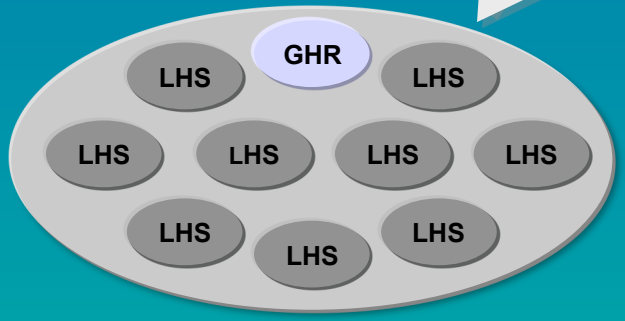


Handle Clients



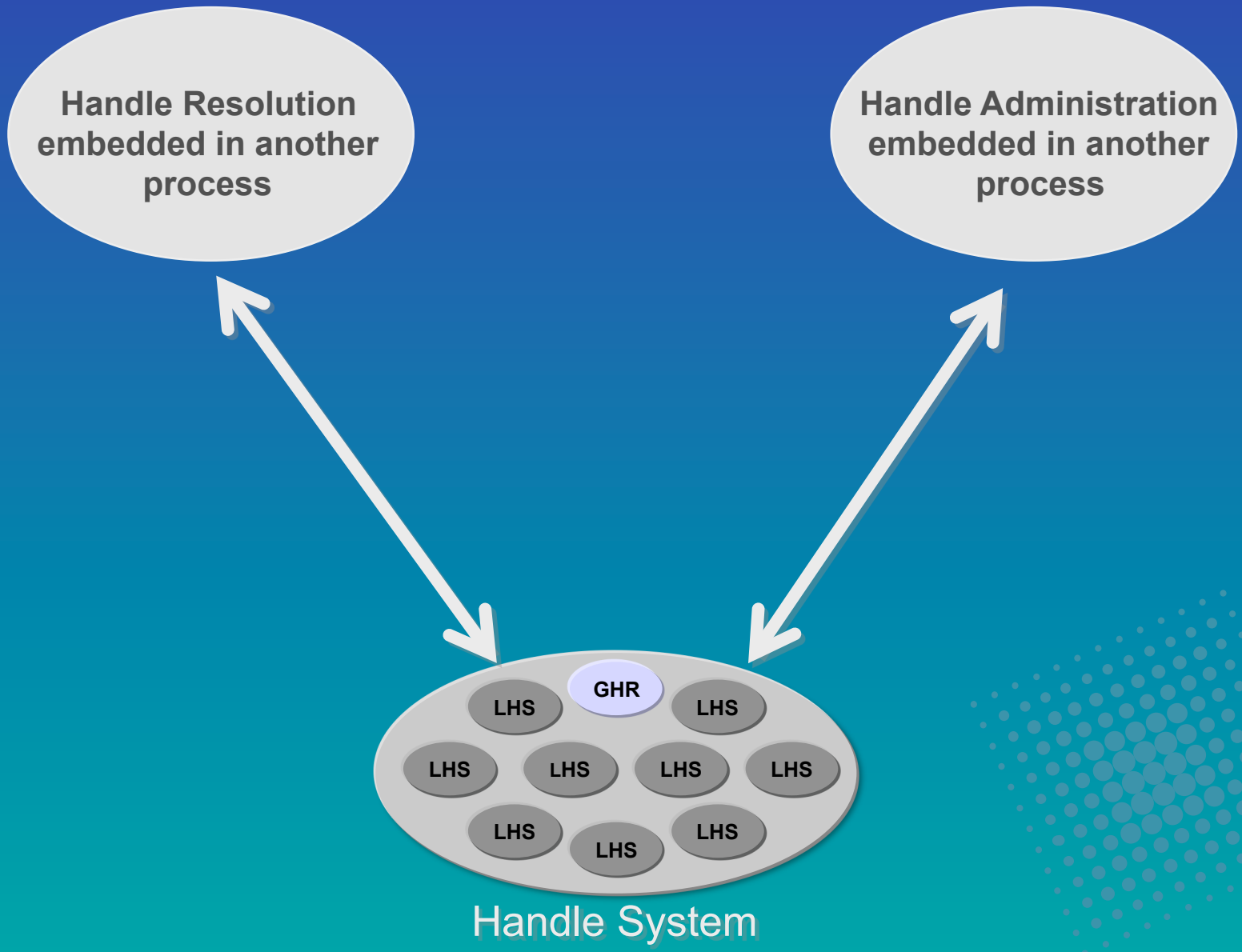
Handle Clients

Handle Administration embedded in another process



Handle System

Handle Clients



Handle System Usage

- Library of Congress
- DTIC (Defense Technical Information Center)
- IDF (International DOI Foundation)
  - CrossRef (scholarly journal consortium, representing >2K publishers & societies)
  - CAL (Copyright Agency Ltd - Australia)
  - MEDRA (Multilingual European DOI Registration Agency)
  - Nielsen BookData (bibliographic data - ISBN)
  - R.R. Bowker (bibliographic data - ISBN)
  - Office of Publications of the European Community (OPOCE)
  - German National Library of Science and Technology (TIB)
  - Wanfang Data
- OECD
- National Agricultural Library/USDA
- DSpace (MIT + HP)
- ADL (DoD Advanced Distributed Learning initiative)
- Los Alamos National Laboratory Research Library
- Australian Dept. of Ed., Sci, and Training (DEST) - PILIN project
- Clarin (Common Language Resources and Technology Infrastructure)
- GENI (Global Environment for Network Innovations)

Handle System Usage
April 09

- Assigned Prefixes
  - DOI – 210,281
  - Other – 1,266
- Handles
  - DOI - 37 M
  - Other - Additional millions (total per prefix known only to prefix manager; LANL adding 600M but privately)
- Handle Services
  - Global
    - Core: four service sites (three CNRI, one Crossref, others coming)
  - Locals
    - >1000 registered LHS's
- Traffic
  - Global: tens of millions per month
  - CNRI-run proxy servers: tens of millions per month

Handle System Management and Standards

- Specification
  - RFC 3650: Overview
  - RFC 3651: Namespace and Service Definition
  - RFC 3652: Protocol
- DoDI 1322.26
- ISO standards track for DOI
- HSAC - Handle System Advisory Committee
  - Approx 15 members representing big users
  - Goal: evolve to oversee the system

Handle System Public Licensing

- License
  - HS Version 6.2 released June 2006 under public license
  - Commercial use welcomed
    - no longer restricted to research and/or education
  - No licensing fees for software or underlying technology
- Service Agreement
  - Service Agreement is required if you use the software/underlying technology to resolve identifiers
  - One time $50 registration fee per prefix
  - Annual $50 maintenance fee per prefix
  - Fees needed to help support global root and administration

Using a Resolution System with Existing Identifiers

- No lack of identifiers
  - ISO TC 46
    - ISAN, ISBN, ISSN, ISRC, ISMN, ISIL, ISNI, ISCI
- 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