

# Nature and Relevance of PID Systems

1 October 2009

Larry Lannom

Corporation for National Research Initiatives

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

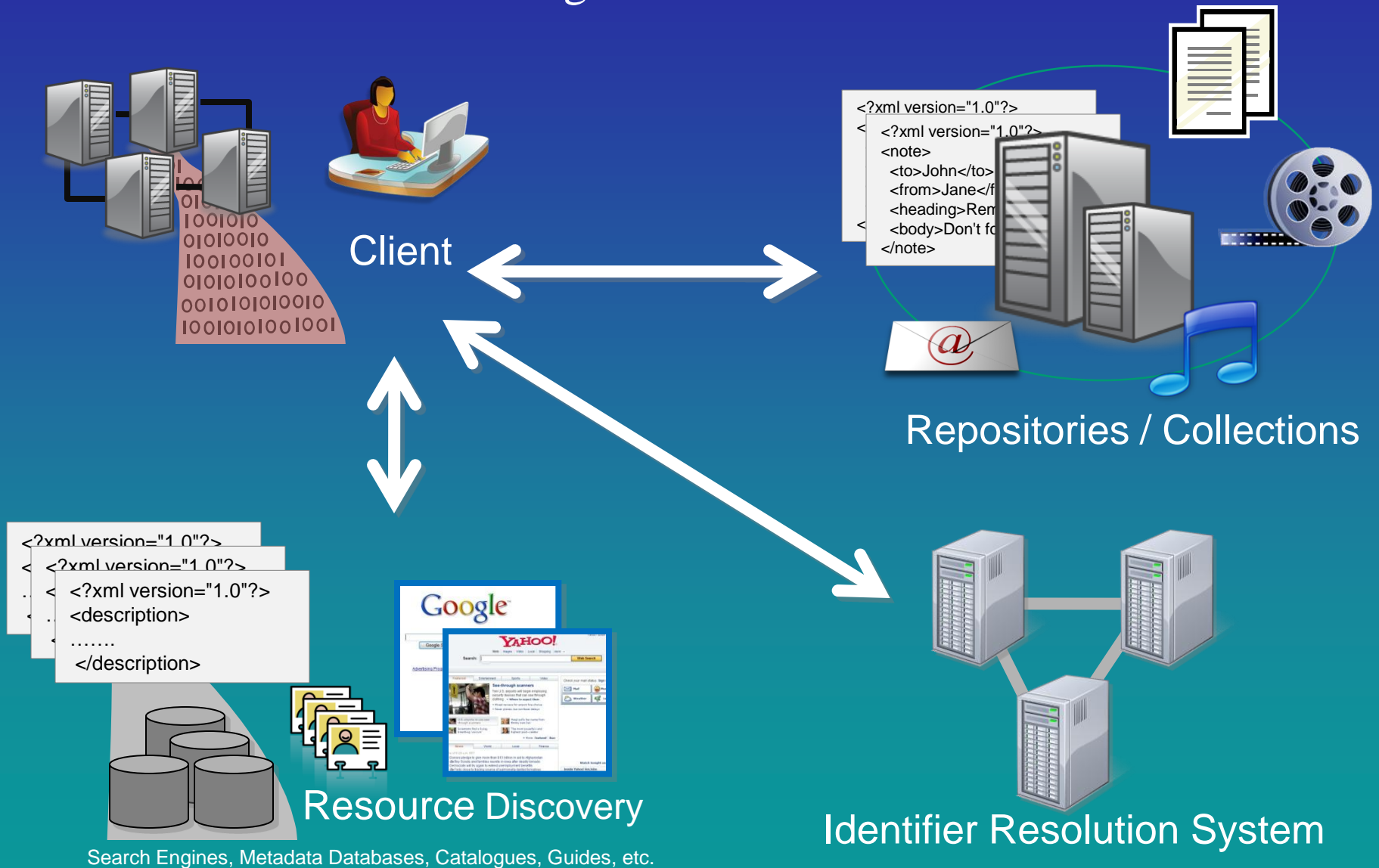
<http://www.handle.net/>

# Background

Larry Lannom: Director of Information Management Technology at Corporation for National Research Initiatives (CNRI).

CNRI: not-for-profit organization formed to undertake research in the public interest. Activities center around long-range infrastructure research in the area of open-architecture networking. Founded and still managed by Robert Kahn, co-inventor of TCP-IP, and director of the research program that resulted in the Internet.

# Role of Identifiers and Identifier Resolution Systems in Information Management on Networks



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

# Why Worry About Identifiers in Research?

- 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

# Requirements: Identifier String

- Not based on any changeable attributes of the entity
  - Location
  - Ownership
  - Any other attribute that may change w/o changing data itself
- 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
- All of the above contribute 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 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 ...

# Handles Resolve to Typed Data

HANDLE



DATA  
TYPE



HANDLE  
DATA



10.123/456

URL

<http://acme.com/...>

URL

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

DLS

acme/repository

HS\_ADMIN

acme.admin/jsmith

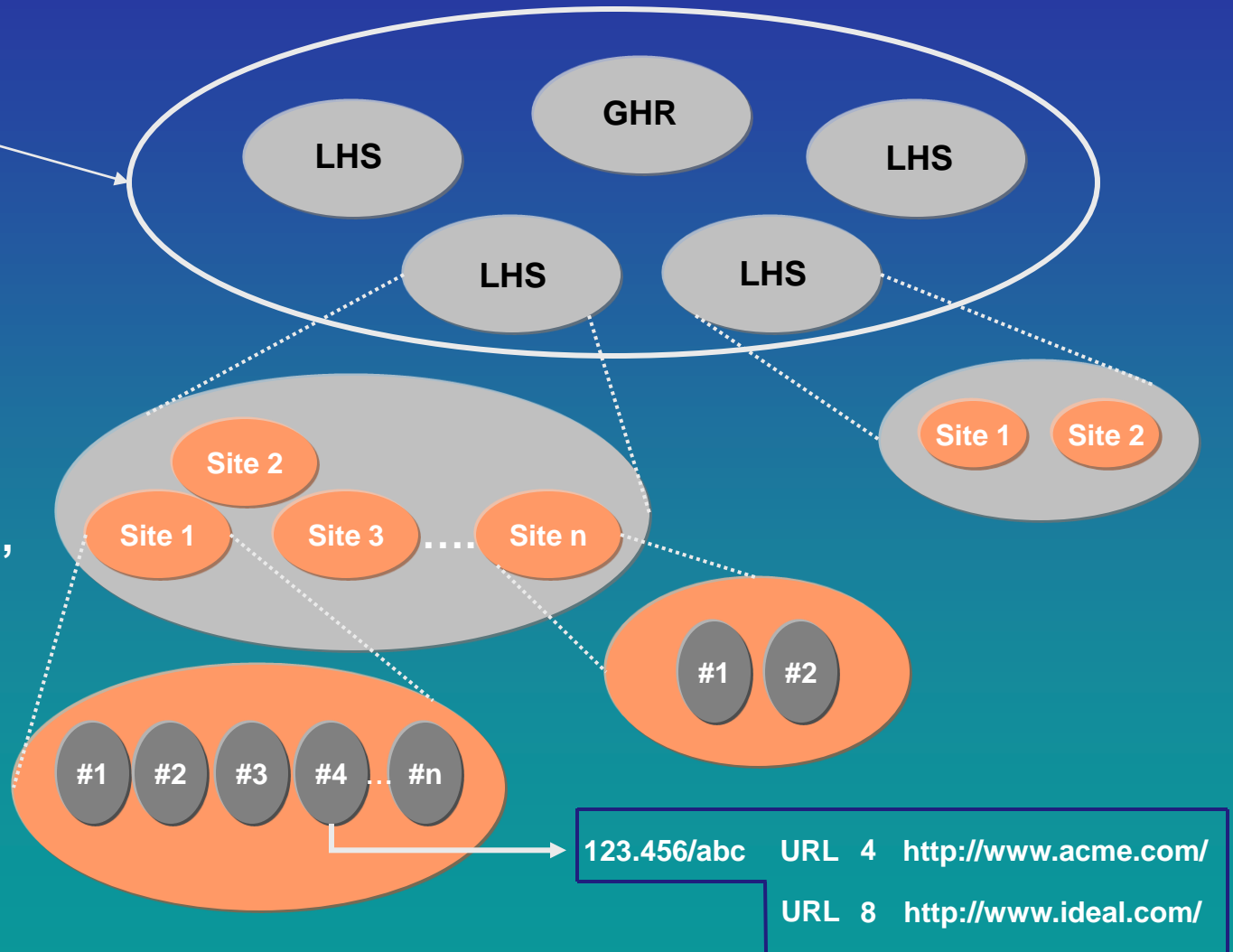
XYZ

1001110011110

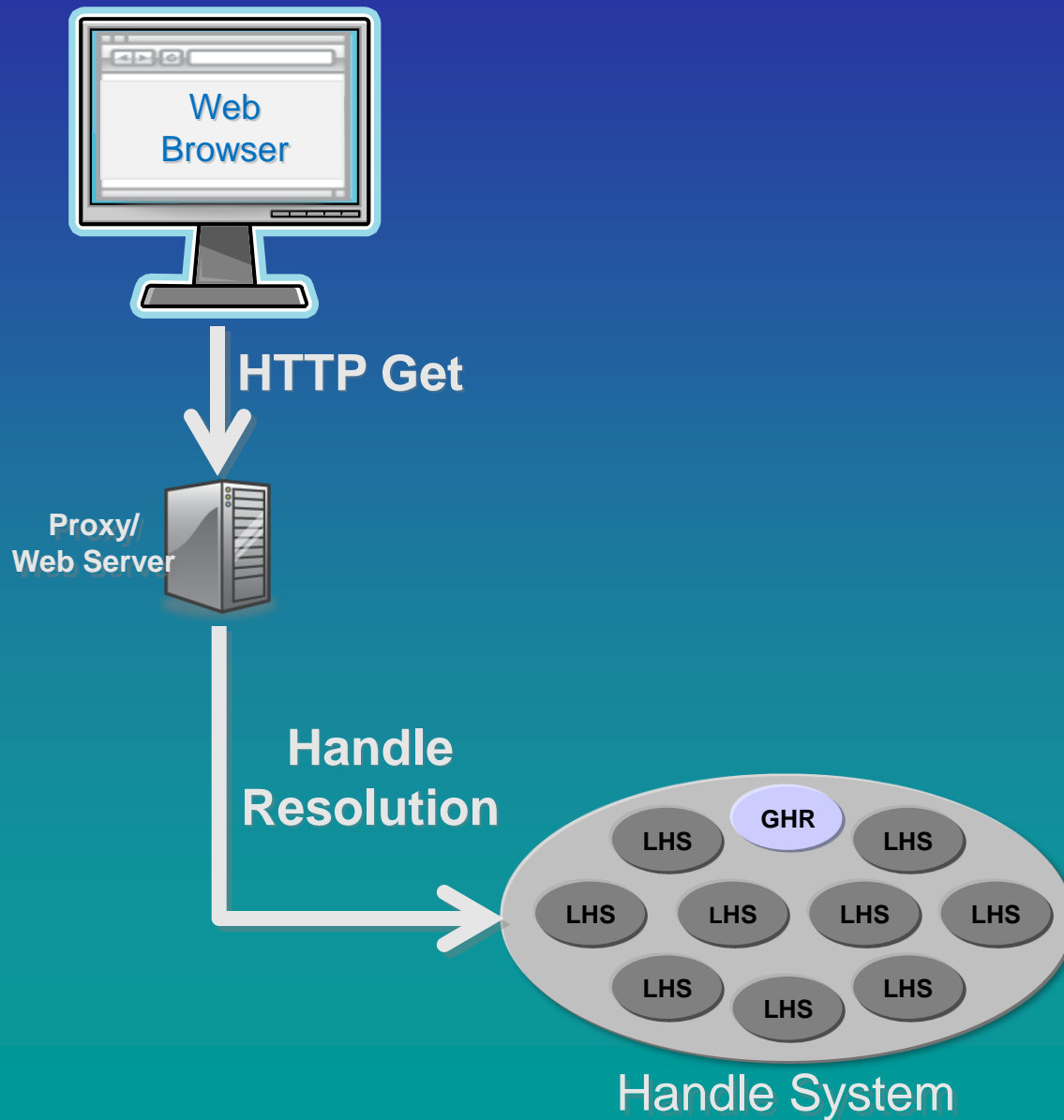
# Handle Resolution



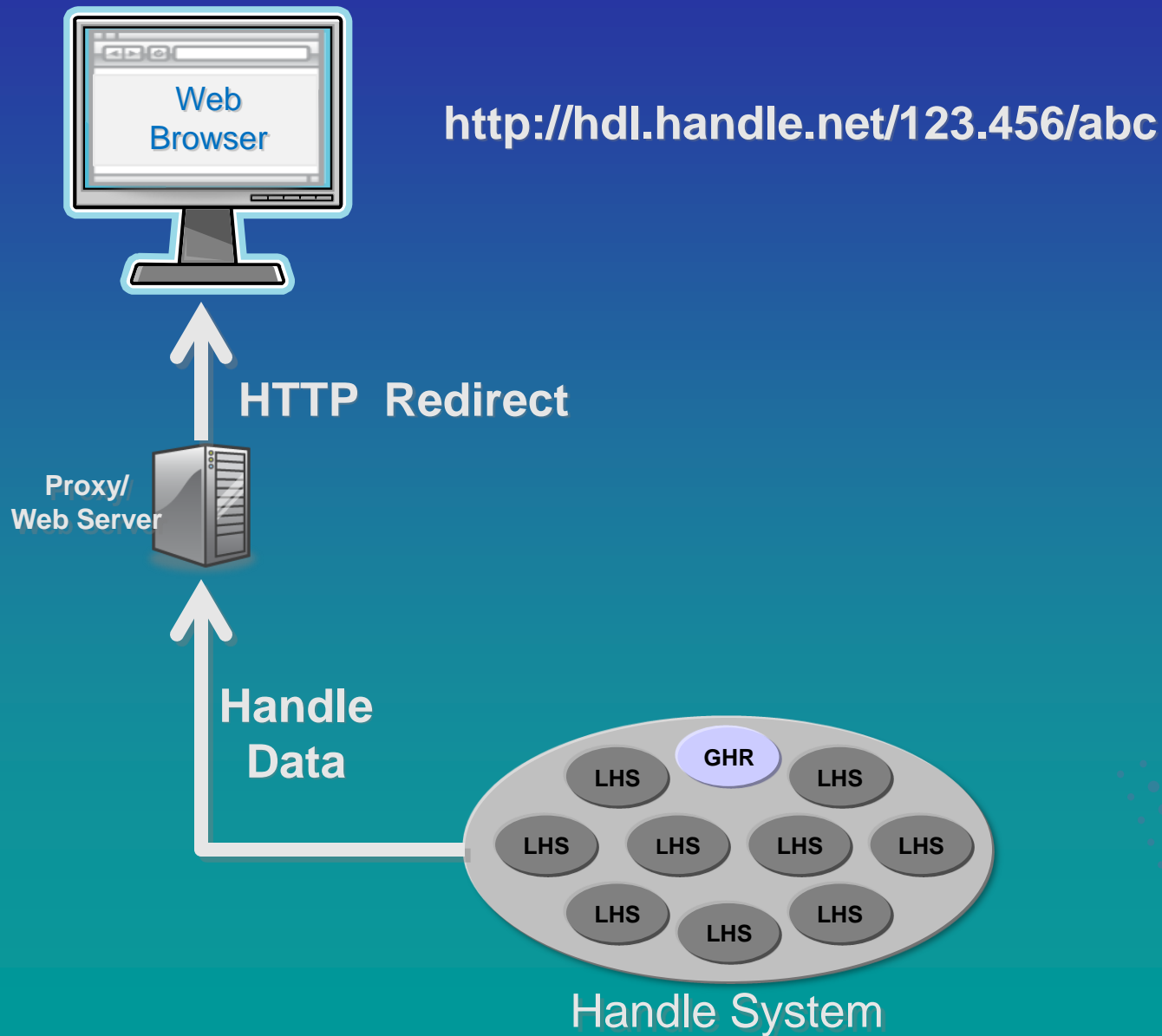
The Handle System is a collection of handle services, each of which consists of one or more replicated sites, each of which may have one or more servers.



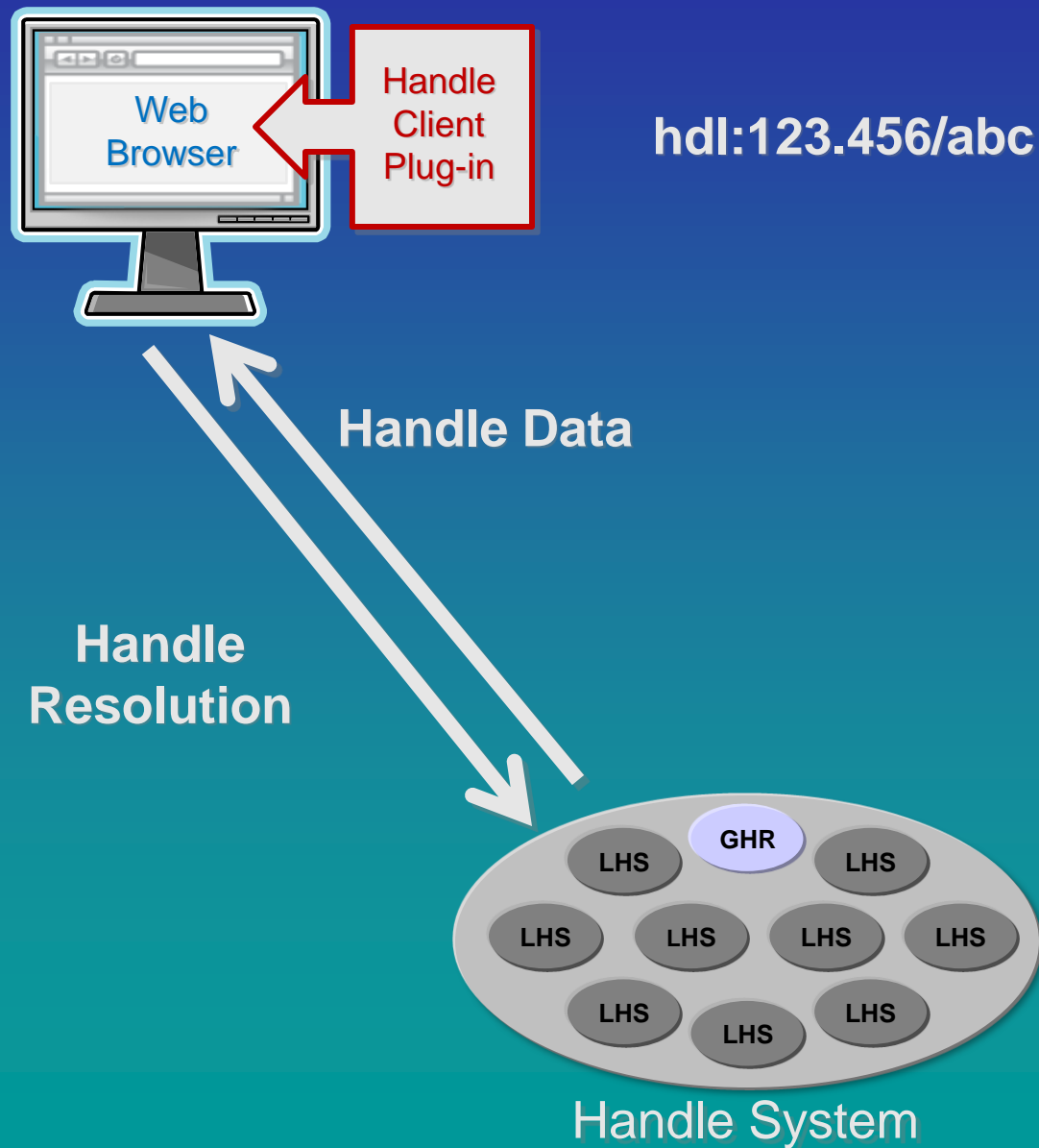
# Handle Clients



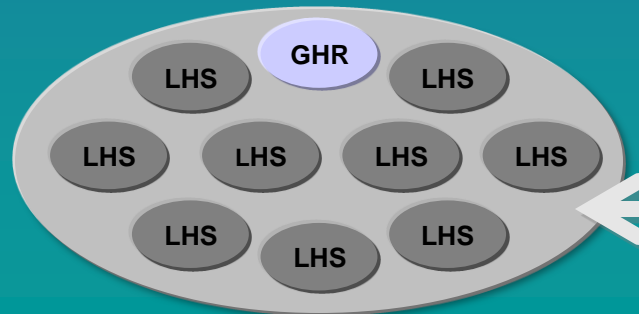
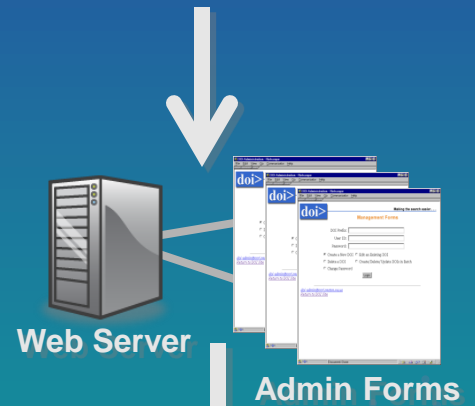
# Handle Clients



# Handle Clients



# Handle Clients

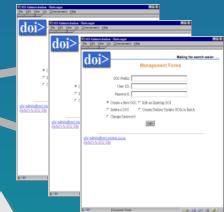


Handle System

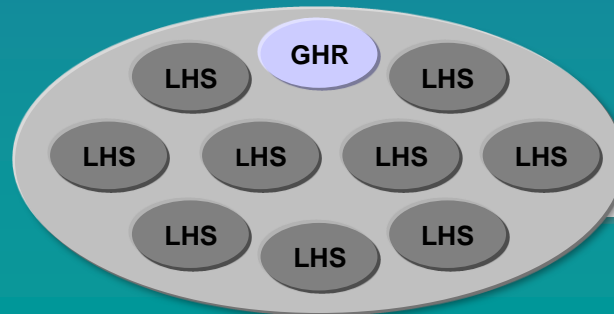
# Handle Clients



HTTP



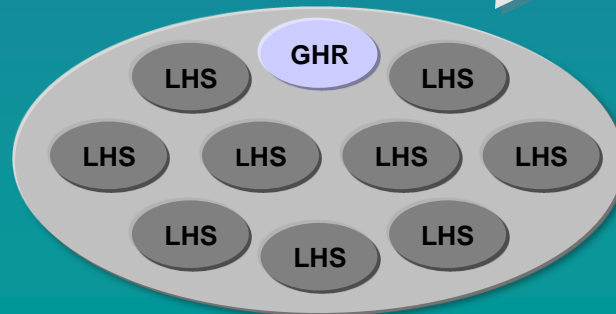
Admin Forms



Handle System

Handle Admin

# Handle Clients



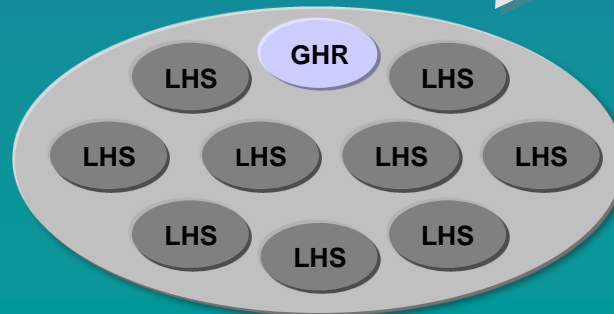
Handle System



# Handle Clients

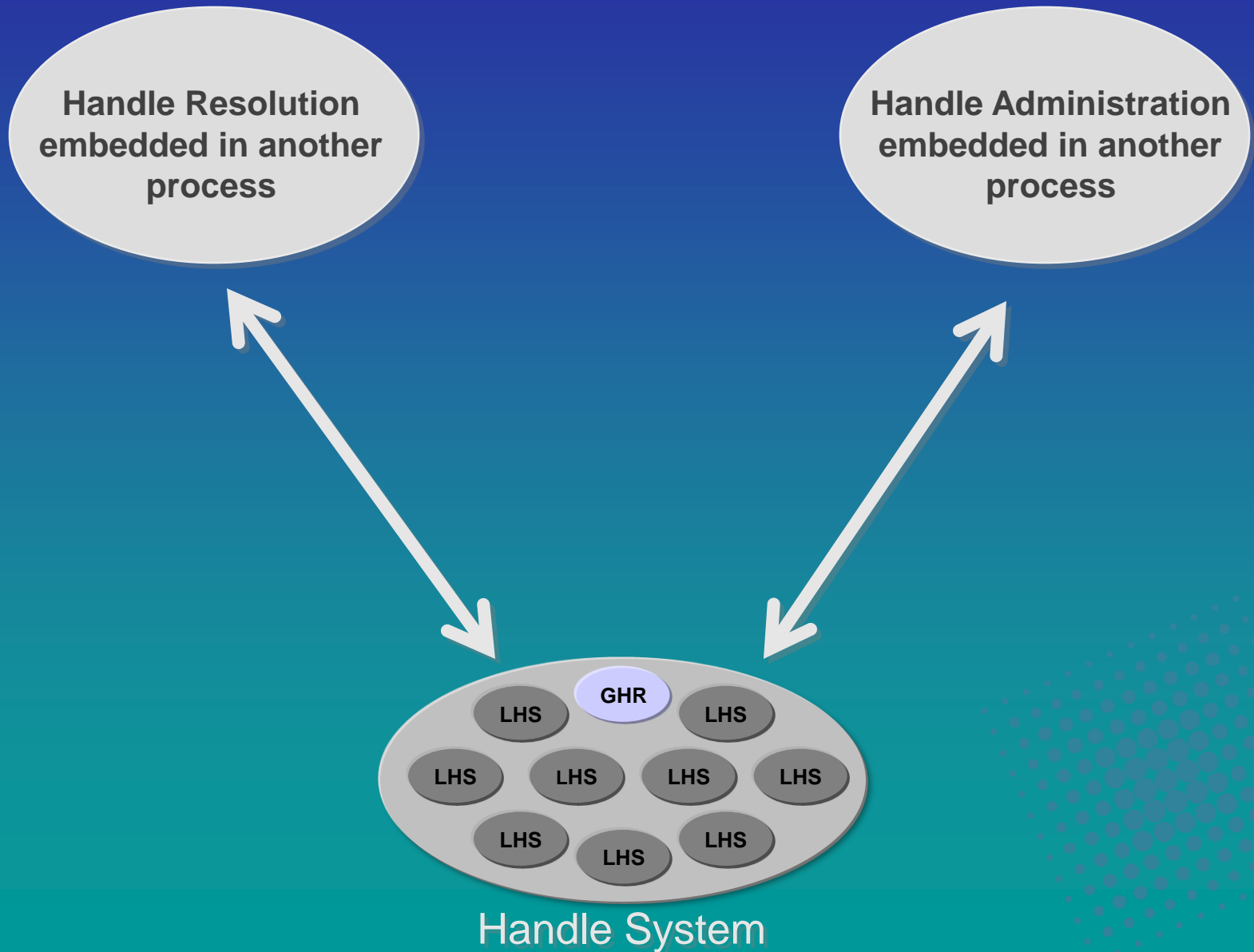


Handle Administration  
embedded in another  
process



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](http://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](http://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

# Handle System Usage

- Library of Congress
- DTIC (Defense Technical Information Center)
- IDF (International DOI Foundation)
  - CrossRef (scholarly journal consortium, representing >2K publishers & societies)
  - mEDRA (Multilingual European DOI Registration Agency)
  - R.R. Bowker (bibliographic data - ISBN)
  - Office of Publications of the European Community (OPOCE)
  - DataCite (8 member consortium started by TIB)
  - Wanfang Data
- OECD
- National Agricultural Library/USDA
- DSpace (MIT + HP)
- ADL (DoD Advanced Distributed Learning initiative)
- Los Alamos National Laboratory Research Library
- Australian National Data Service (ANDS)
- Clarin (Common Language Resources and Technology Infrastructure)
- GENI (Global Environment for Network Innovations)

# Handle System Usage

## August 09

- Assigned Prefixes
  - DOI – 210,390
  - Other – 1,309
- Handles
  - DOI – 42 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
- U.S. Patent 6,135,646
  - Intent was to protect the technology as usage grew
  - Never used by CNRI, but has been referenced by others as prior art
  - It has served its purpose well and it expires in 2013
- HSAC - Handle System Advisory Committee
  - Approx 15 members representing big users
  - Maturation has diminished need for advice
  - Time for the next stage

# Handle System Management: Next Steps

- Global Mirrors
  - Currently 4 sites – 3 run by CNRI, 1 by CrossRef
  - Agreements pending with CNNIC, GWDG
- Default web-to-handle proxy server system (hdl.handle.net)
  - Currently 4, load-balanced and all run by CNRI
  - Open to others running additional instances
  - MPG/GWDG have expressed interest
- ‘Handle System Board’ (for now - final name TBD)
  - Currently in initial planning stages
  - MOU with MPG, with this as its goal, close to signature
  - It would assume ownership of the system, incl. intellectual property
  - It would assume responsibility for the system, incl. financial
  - Invited participation ff. by the hard job of creating the agreement
  - Criteria for invitation – vested interest plus resources

# 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**
  - **Going from \$0 to \$50 did not slow growth**
- **System not self-supporting, but getting closer**
- **The ‘board’ would inherit this mechanism and could make changes**