DAM-LR
Distributed Solution

Working on a Federated Archive

- ideas -

Daan, Freddy, Peter

Federation Goal in DAM-LR

- single sign-on
- integrated metadata layer
- one basket idea
- federated authorization

User

integrated shell

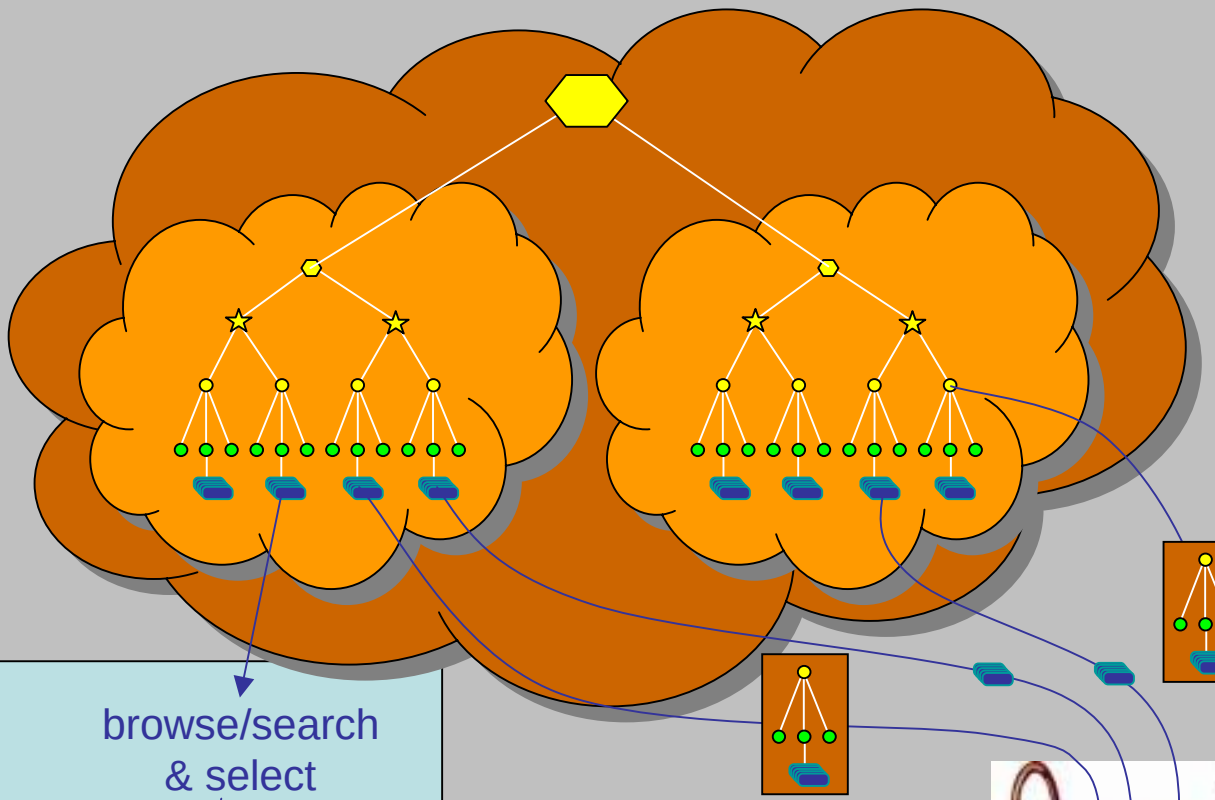
management and access shell

Archive X

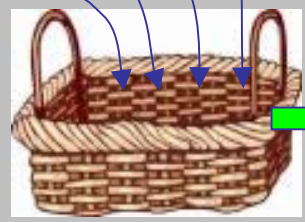
management and access shell

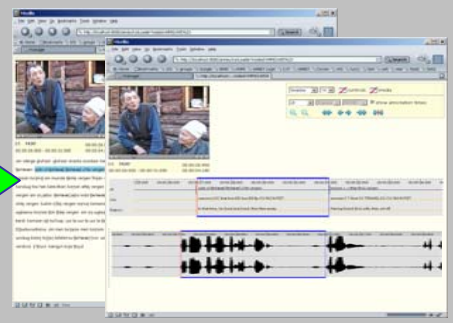
Archive Y

Private Domain - DELAMAN Vision



- user selects resources from everywhere
- this is his/her private virtual temporary domain
- can store it as profile
- invoked by session
  - one authentication step
  - one authorization step
- then search (MD + content)
- visualization/listening/…
- comparison
- …





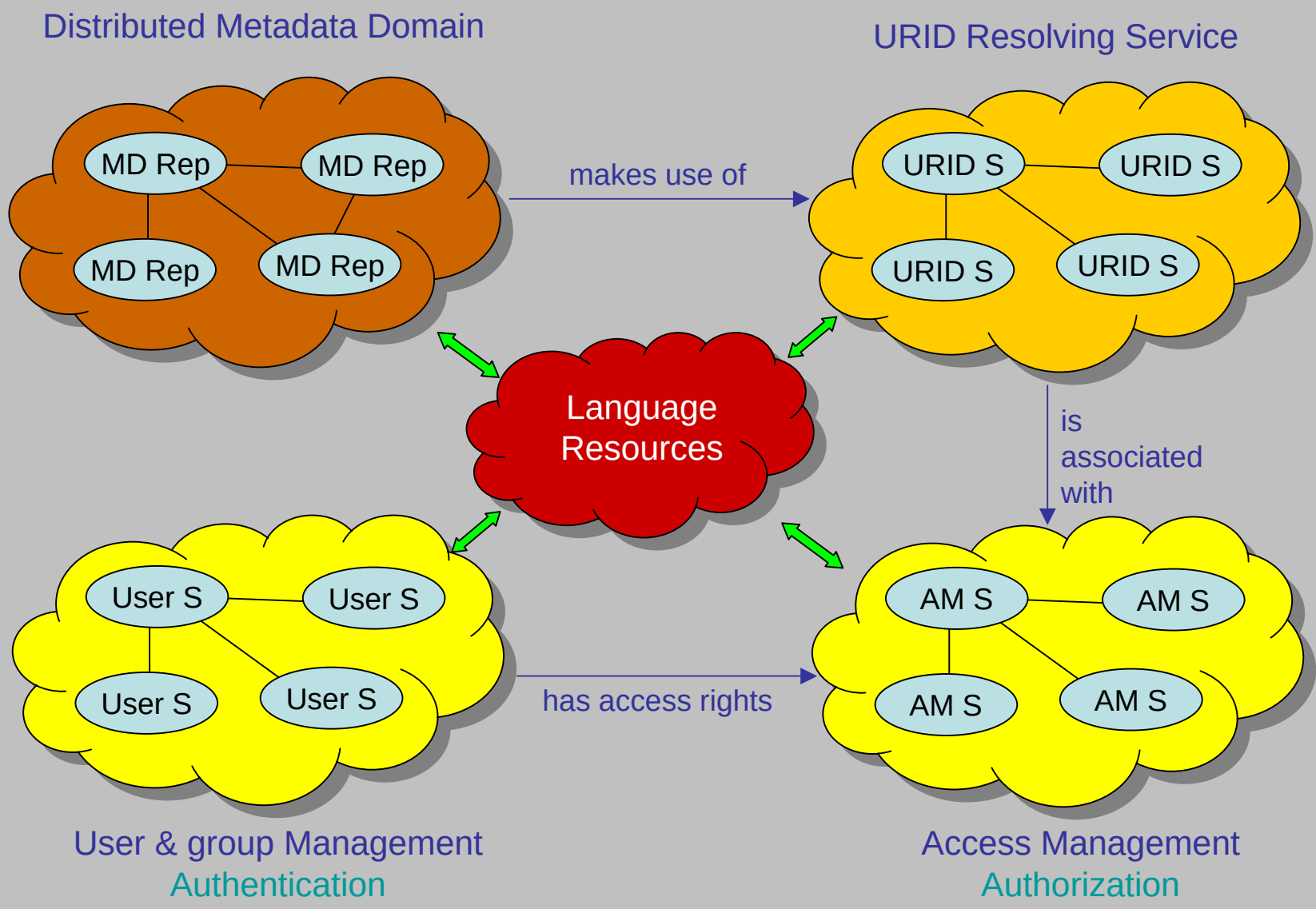
browse/search & select

authenticate authorize

play

web-browser & plug-ins

Pillars of the Solution



DAM-LR Project Goals

- joining the collections (or part of the collections) of
  - MPI,
  - Lund,
  - SOAS and
  - INL
- at the end: demonstrate a functioning Federation
- but: everyone has to be able to work stand-alone!

URID Pillar

Unique Resource Identifiers
URIDs

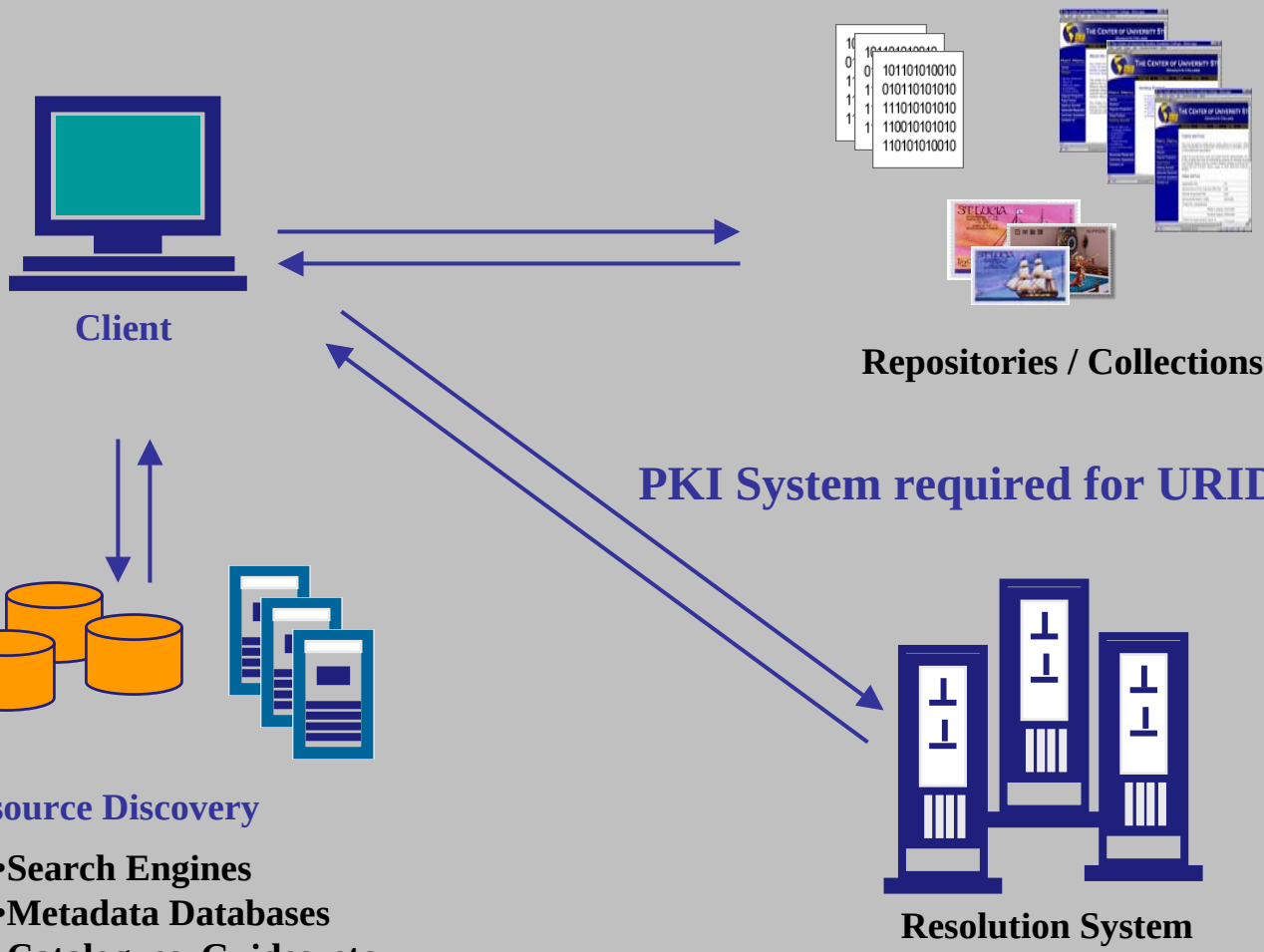
Slides partly taken from Larry Lannom

URID: Reasons

- Managing Digital Objects is a huge task
- Give resources a name and talk to it
  - don’t worry about where it is
  - don’t worry about what it’s made of
  - similar to ISBN

- Rise above details of application versions and content formats
- Make archive ready for all sorts of references
  that have to be stable independent of physical changes
- Just one place to carry out changes – tractability issue

URID: Information Flow



URID: Nothing for free

- Fundamental indirection system for Digital Archive management on the net
- No free lunch ☹
  - Added layer of infrastructure
  - Must be managed
  - Must run with high availability
  - Must not prevent users from accessing resources directly
  - has to function reliably
- The Handle System is a candidate for URID resolving
  - Installed and tested it already at MPI

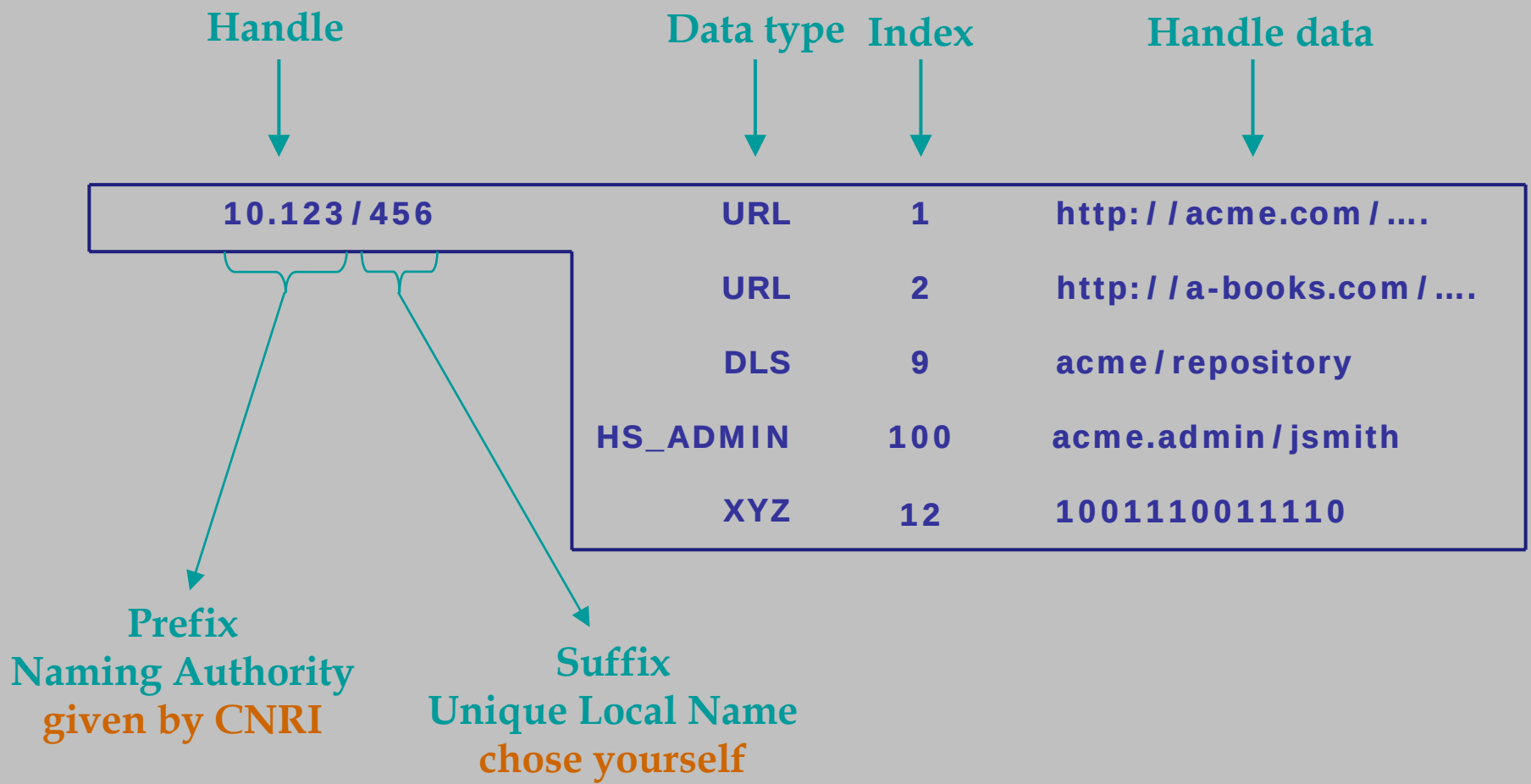
CNRI Handle System

- Distributed, scalable, secure
- Enforces unique names
- Enables association of one or more typed values, e.g., URL, with each name
- Optimized for speed and reliability
- Open, well-defined protocol and data model
- Provides infrastructure for application domains, e.g., digital libraries, electronic publishing ...

Handle System Usage

- Library of Congress
- DTIC (Defense Technical Information Center)
- IDF (International DOI Foundation)
  - CrossRef (scholarly journal consortium)
  - Enpia (Korean content management technology firm)
  - CDI (U.S. content management technology firm)
  - LON (U.S. learning object technology firm)
  - CAL (Copyright Agency Ltd - Australia)
  - TSO (U.K. publisher & info mgmt service provider)
  - MEDRA (Multilingual European DOI Registration Agency)
  - Nielsen BookData (bibliographic data - ISBN)
  - R.R. Bowker (bibliographic data - ISBN)
  - Office of Publications of the European Community
- NTIS (National Technical Information Service)
- DSpace (MIT + HP)
- CORDRA (ADL's Federated Content Repository Model)
- Globus Toolkit (in development)

Handles Resolve to Typed Data

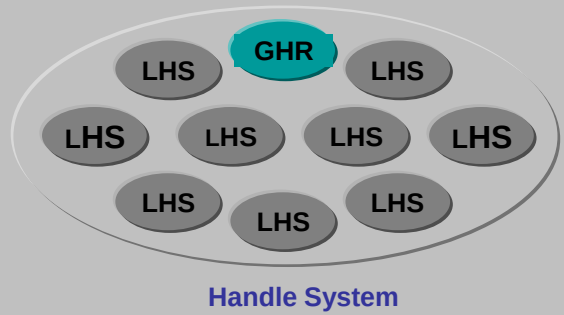


The Two Types of Handle Query

1. Request all data

Give me all data associated with handle 10.1000/123.

Handle	Index	Type	Data
10.1000/123	3	URL	URL1(Server in US)
	2	URL	URL2 (Server in Asia)
	5	URL	URL3 (Server in Europe)
	10	PK	public key
	9	EM	email address
	4	IP	rights data

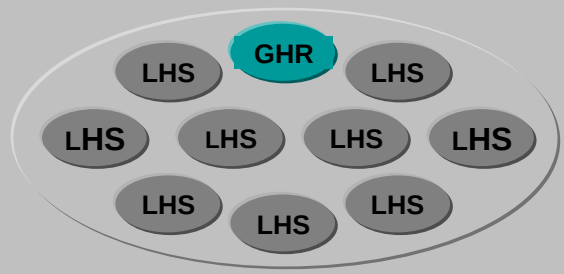


Handle System

2. Request all data of a given type

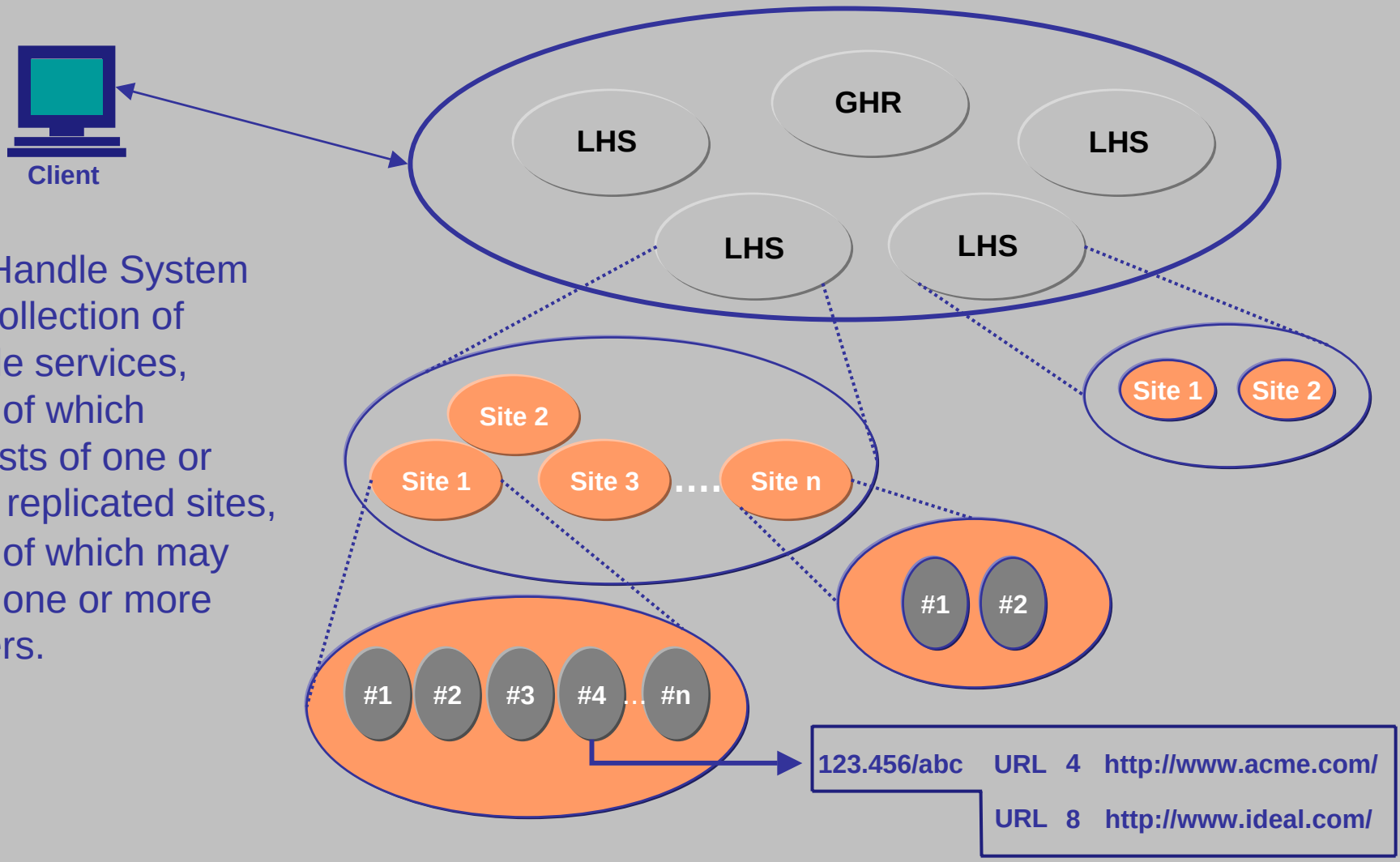
Give me all data of type URL associated with handle 10.1000/123.

Handle	Index	Type	Data
10.1000/123	3	URL	URL1(Server in US)
	2	URL	URL2 (Server in Asia)
	5	URL	URL3 (Server in Europe)



Handle System

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

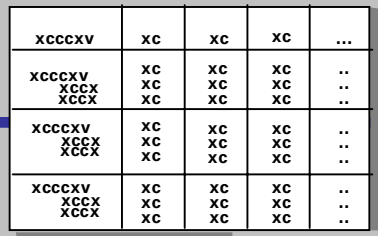
Client

2. Global Responds with
Service Information for 10.1000

Global Handle Registry

xccccxv	xc	xc	xc	...
xccccxv xccx xccx	xc xc xc	xc xc xc	xc xc xc	.. .. ..
xccccxv xccx xccx	xc xc xc	xc xc xc	xc xc xc	.. .. ..
xccccxv xccx xccx	xc xc xc	xc xc xc	xc xc xc	.. .. ..

Service Information
Acme Local Handle Service



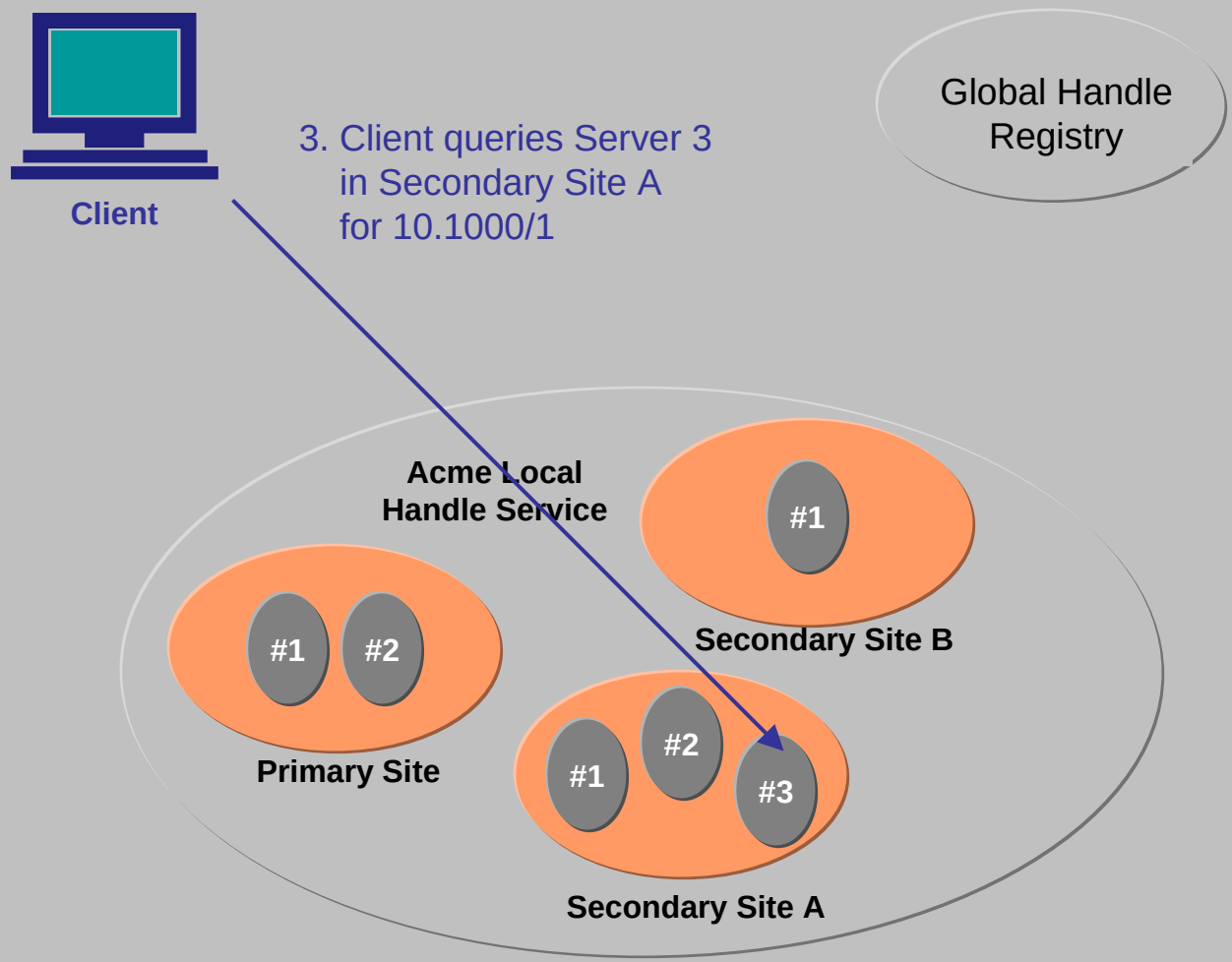
Handle Clients

	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

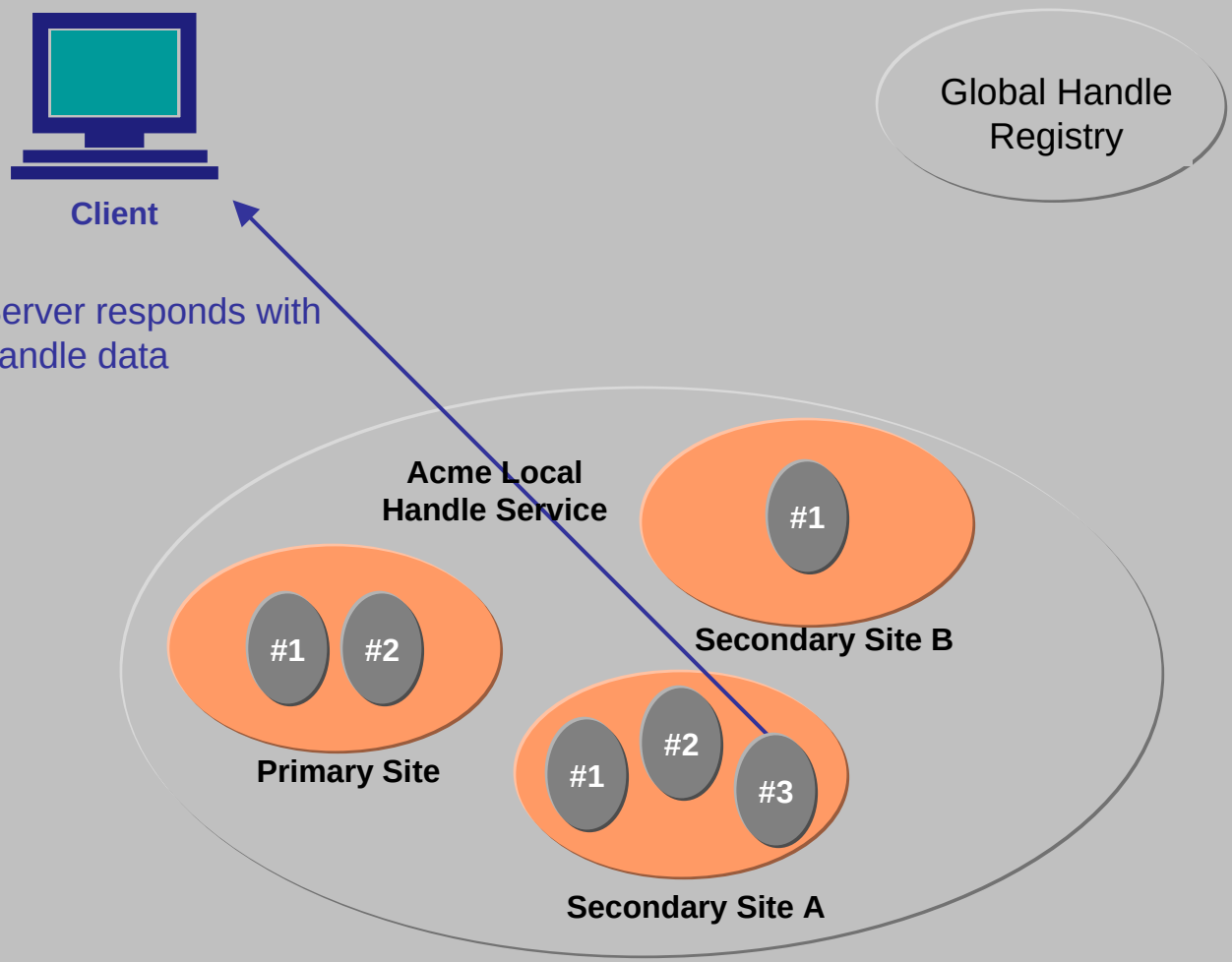
Handle Clients

Request to Client:
Resolve hdl:10.1000/1

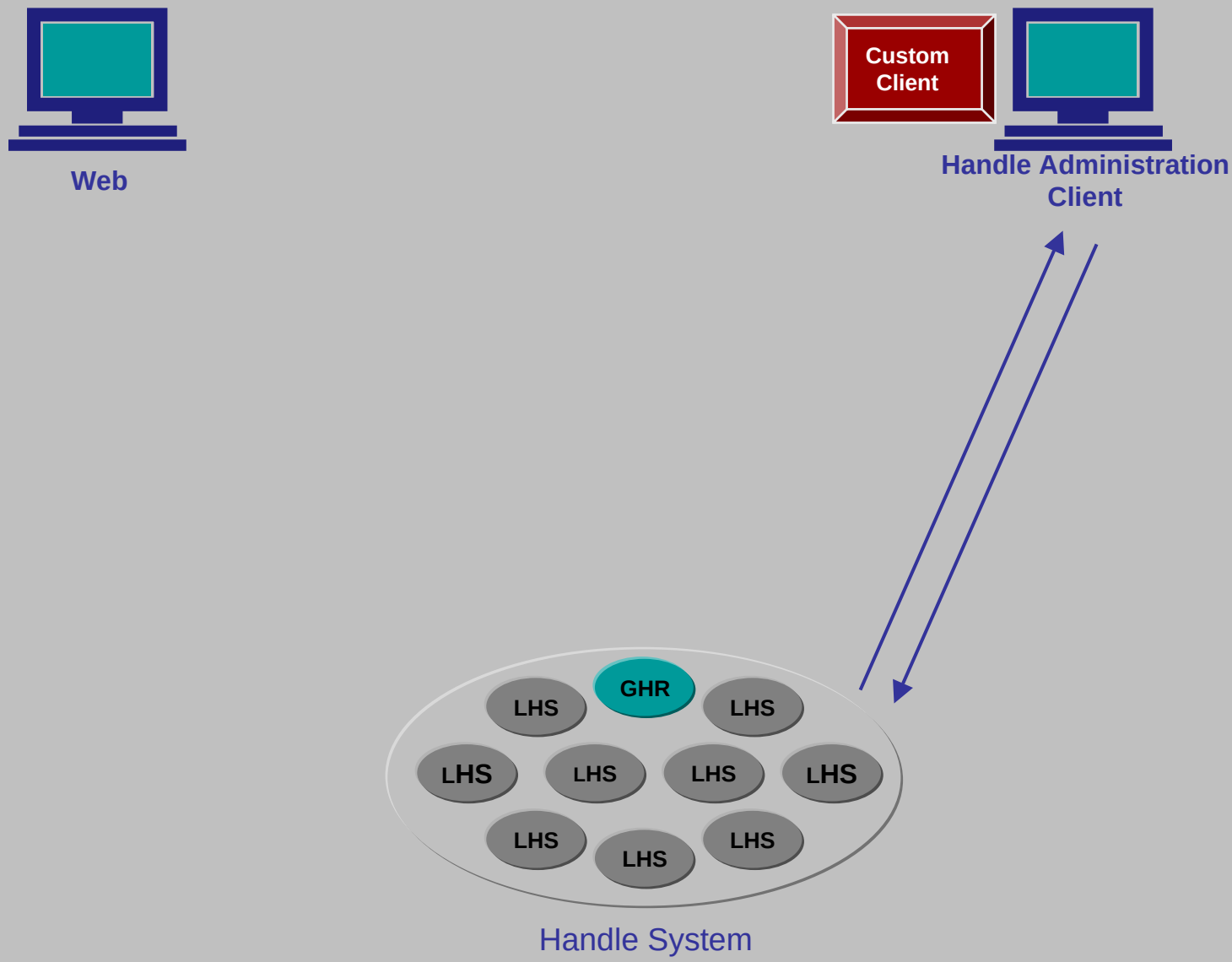


Handle Clients

Request to Client:
Resolve hdl:10.1000/1



Handle Clients



HS Administration

- Ownership is at the handle level
- Administrators defined by handles
- Administrator handles contain keys
- All admin transactions validated via challenge/response from server to client
- Allows distributed administration

URIDs: HS Questions I

- which is the appropriate handle domain?
  - one DELAMAN/DAM-LR domain
    - one resolution domain, several replication servers, traffic local
    - then structure in the suffix

      (<DELAMAN prefix>/<institute><unique number>)
  - one domain per archive root
    - several resolution domains, proliferation of servers, traffic via GHS
    - then simple structure (<institute prefix>/<unique number>)
  - how to choose the random number
    - better no structure dependent on formats, origin etc
- URIDs point to different URLs in case of copies

URIDs: HS Questions II

- should we indicate versions via the Handle System
  - only the latest version has a URID
  - but how to point to older versions?
- should all resources have a URID as a principle
  - or are there dependent resources (pictures in HTML file, etc)
- should we add a link to the Authentication & Authorization Infrastructure?
  - quick solution for finding out whether person is allowed to access
- obvious that server have to identify them trustfully
  - Public Key Infrastructure is needed

AAI Pillar

Authentication and Authorization Infrastructure

AAI

DAM-LR Federation

- single sign-on
- integrated metadata layer
- one basket idea
- federated authorization

User

integrated shell

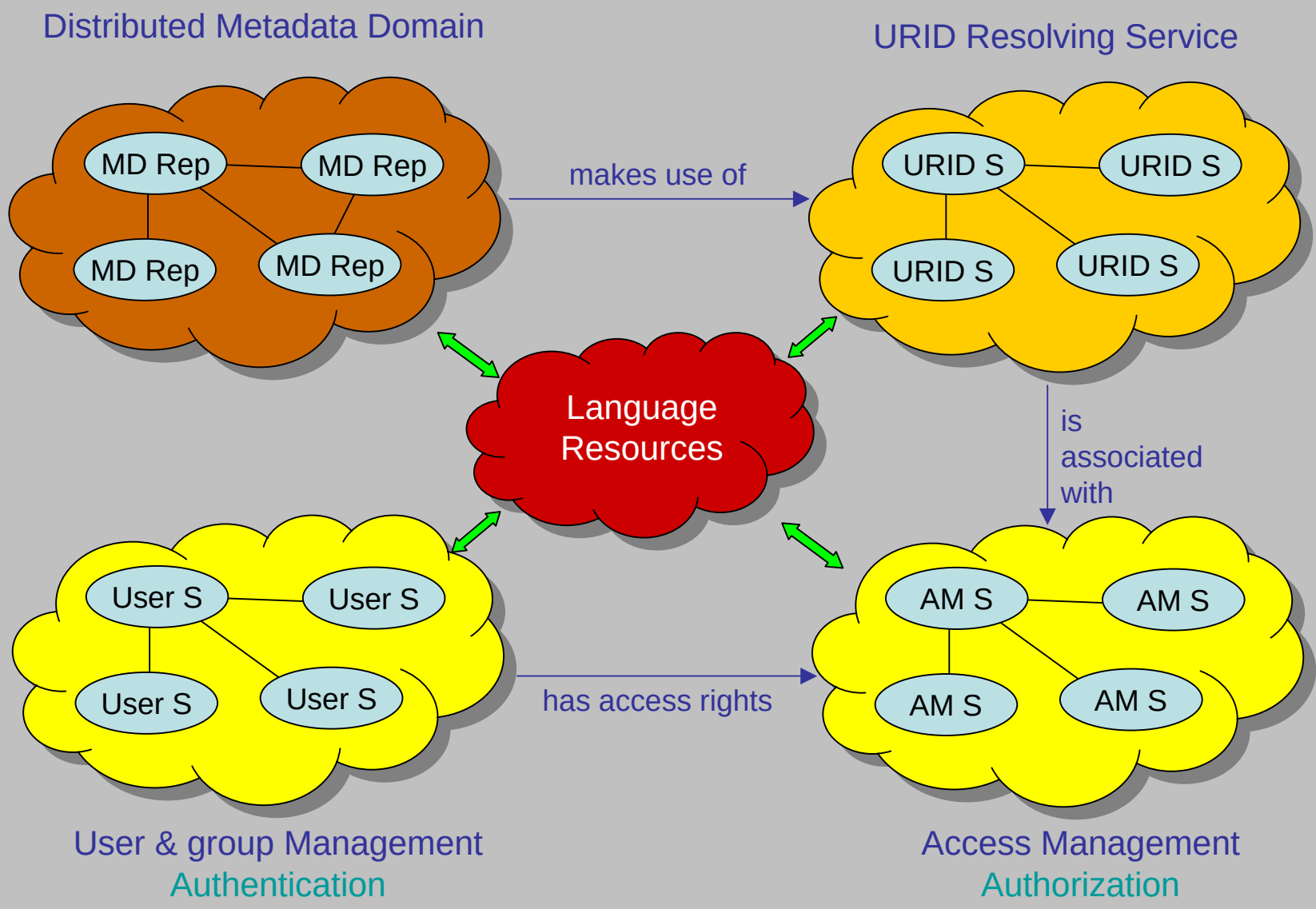
management and access shell

Archive X

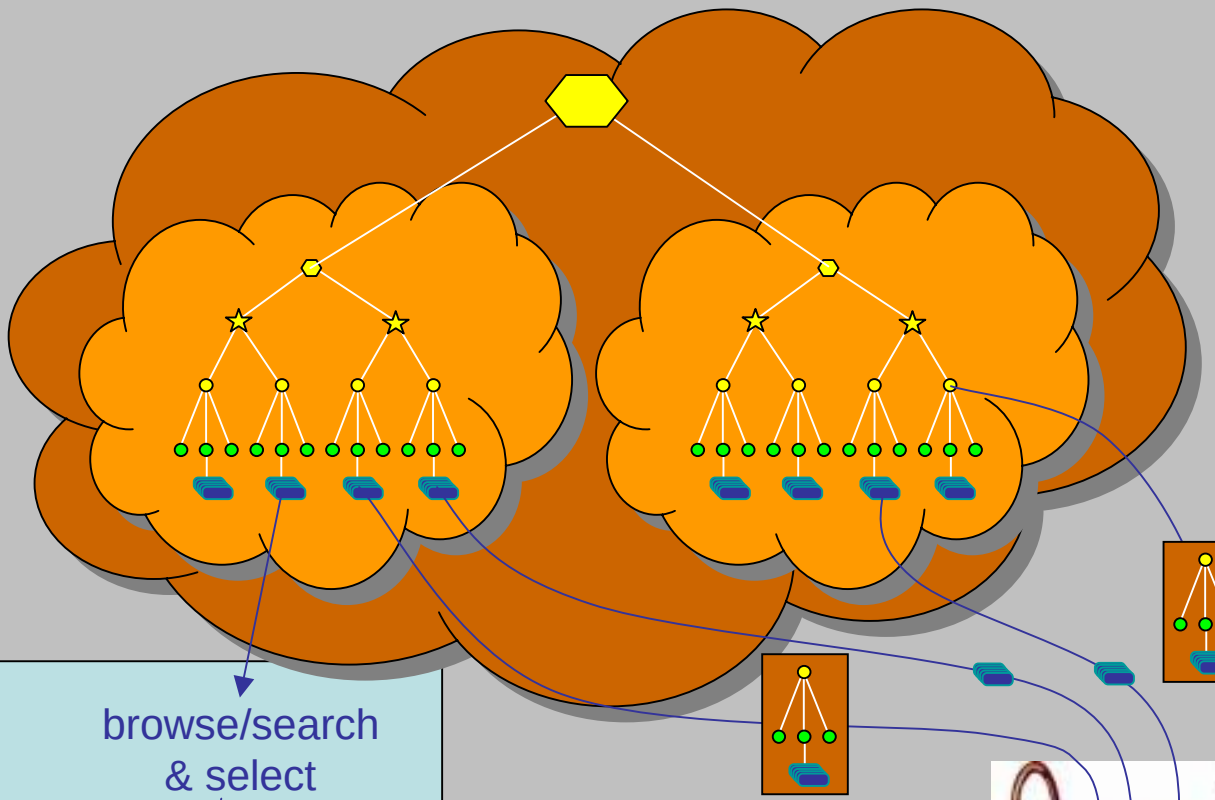
management and access shell

Archive Y

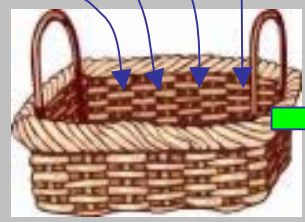
Pillars of the Solution

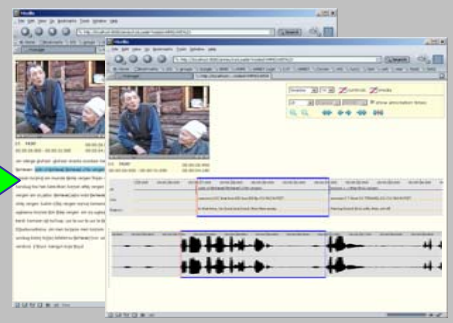


Private Domain - DELAMAN Vision



- user selects resources from everywhere
- this is his/her private virtual temporary domain
- can store it as profile
- invoked by session
  - one authentication step
  - one authorization step
- then search (MD + content)
- visualization/listening/…
- comparison
- …





browse/search & select

authenticate authorize

play

web-browser & plug-ins

AAI: Requirements I

- AA domain must be clear
  - authority remains with the origin when copying
  - AM includes policies (usages, ethical rules, …)
  - at first instance it's up to the institute how authentication is done
  - definition of policies and rights must be efficient (management task)
- how do we share users?
  - only one entry per user (at home institute or by exchange (unsafe))
  - reduce management task
- user is prompted for identity when accessing a resource
  - no problem for a single resource
  - what when having a basket – a private workspace …
  - want to be sure that access is seamless after session start
    - batch type authorization transparent to user
    - caching of tickets for whole session

AAI: Requirements II

what else is of relevance?

Shibboleth is a candidate – have to look at it and test it if DAM-LR will go for it

different authentication systems around

- MPI has one for the archive (as others)
- A-Select (SURFNet) is a very professional one
- have to look at it in more detail

Current AAI Systems

essentially 3 scenarios:

1. authentication = authorization (simple)
2. identity plus a few attributes (commonly used)
   - at MPI identity + acceptances + group association
   - but authentication means authorization (central solution)
3. privacy-preserving negotiation about attributes to be exchanged (ideal and upcoming)

Birdseye View on Shib

What is Shibboleth?

An Internet2/MACE project than provides a framework and technology for inter institutional authorisation for (web) resources. A major feature is to offer authorisation without compromising the users privacy. All based on Trust relations within a federation;

What does Shibboleth offer?

authorisation, attribute gathering and privacy safe transport of attributes; attribute based authorization; users can control which data is sent where

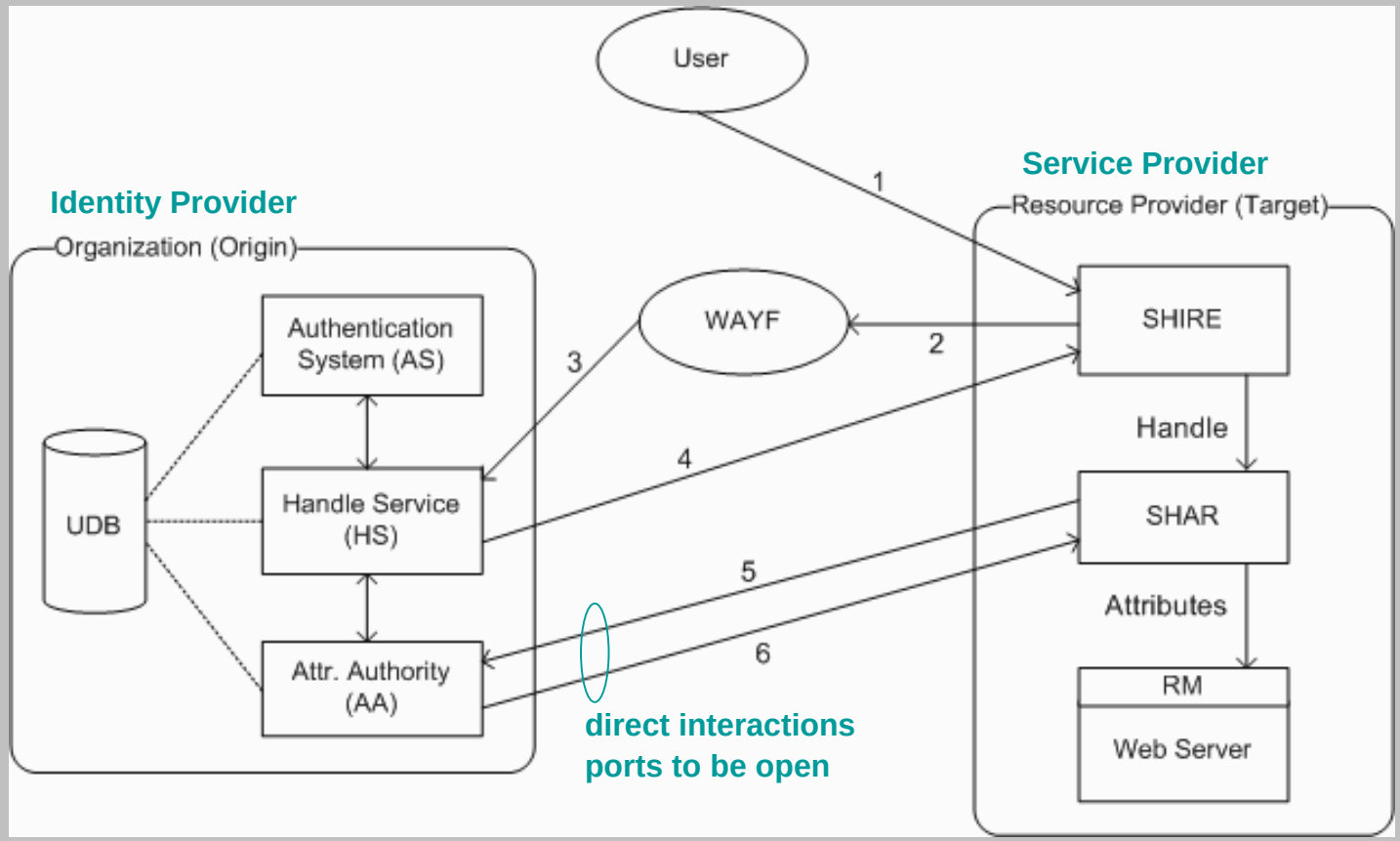
What doesn't Shibboleth do?

Out of the box authentication, choose a Web AMS (f.e. A-Select) can operate with local authentication solutions

Results at a protected resource after Shibboleth process:

user ID-x with the attributes X,Y wants access to resource Z

Shib Configuration + Traffic



1 User finds a resource with MD and selects it
2 Shibs target SW redirects request to WAYF
3 WAYF presents form with all federation members and user selects
4 WAYF passes request to Shibs Origin SW
5 The Handle service connects with the AS system and if all ok provides an “assertion”
6 SHIRE checks correctness and SHAR requests attributes for the handle
7 The local RM checks whether person with that attributes is allowed to access

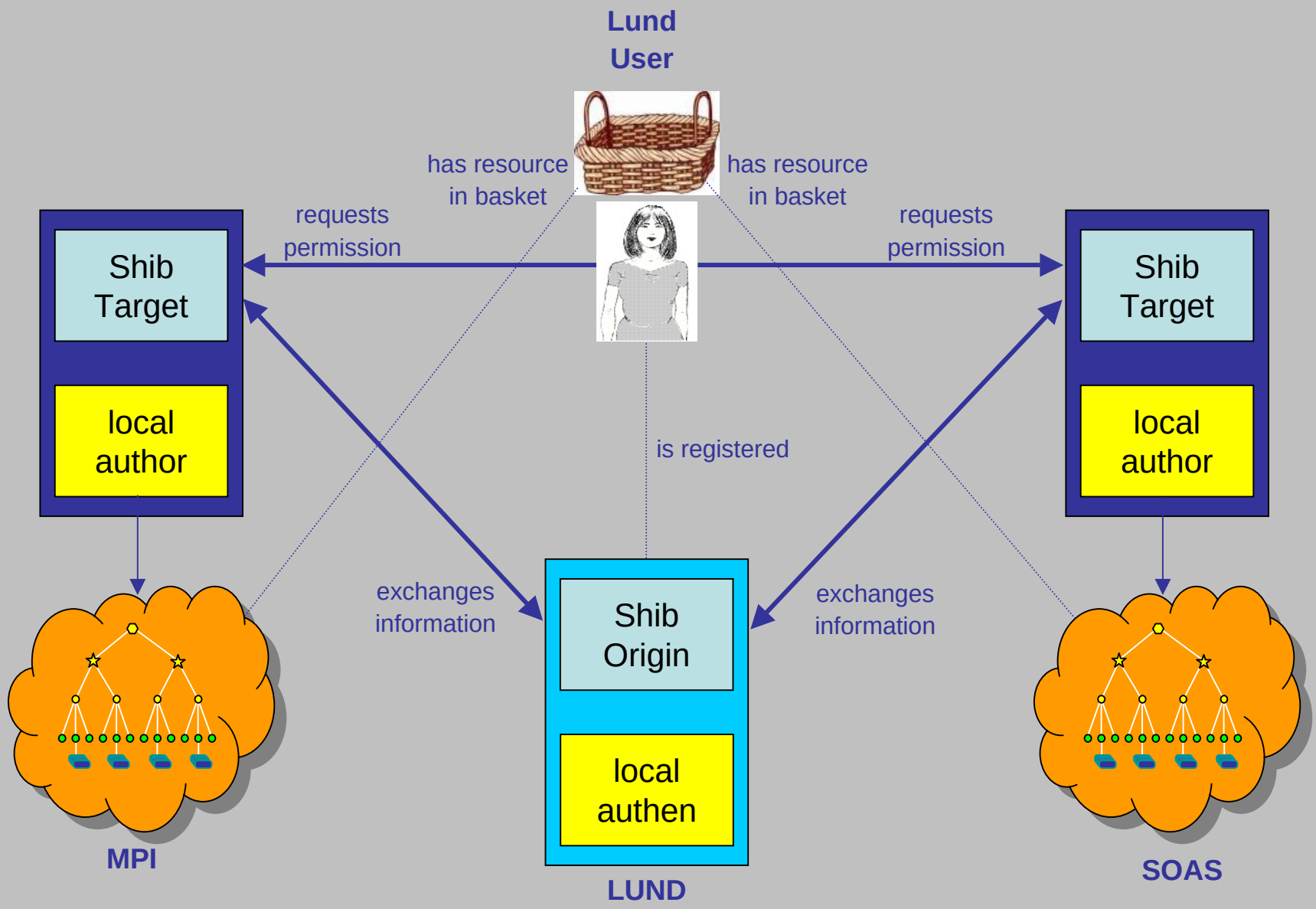
SHIRE **Shib Indexical Reference Establisher**
WAYF **Where are you from**
SHAR **Shib Attribute Requester**
RM **Resource Manager**

Shib is modular (can intervene for our purposes) and Open Source

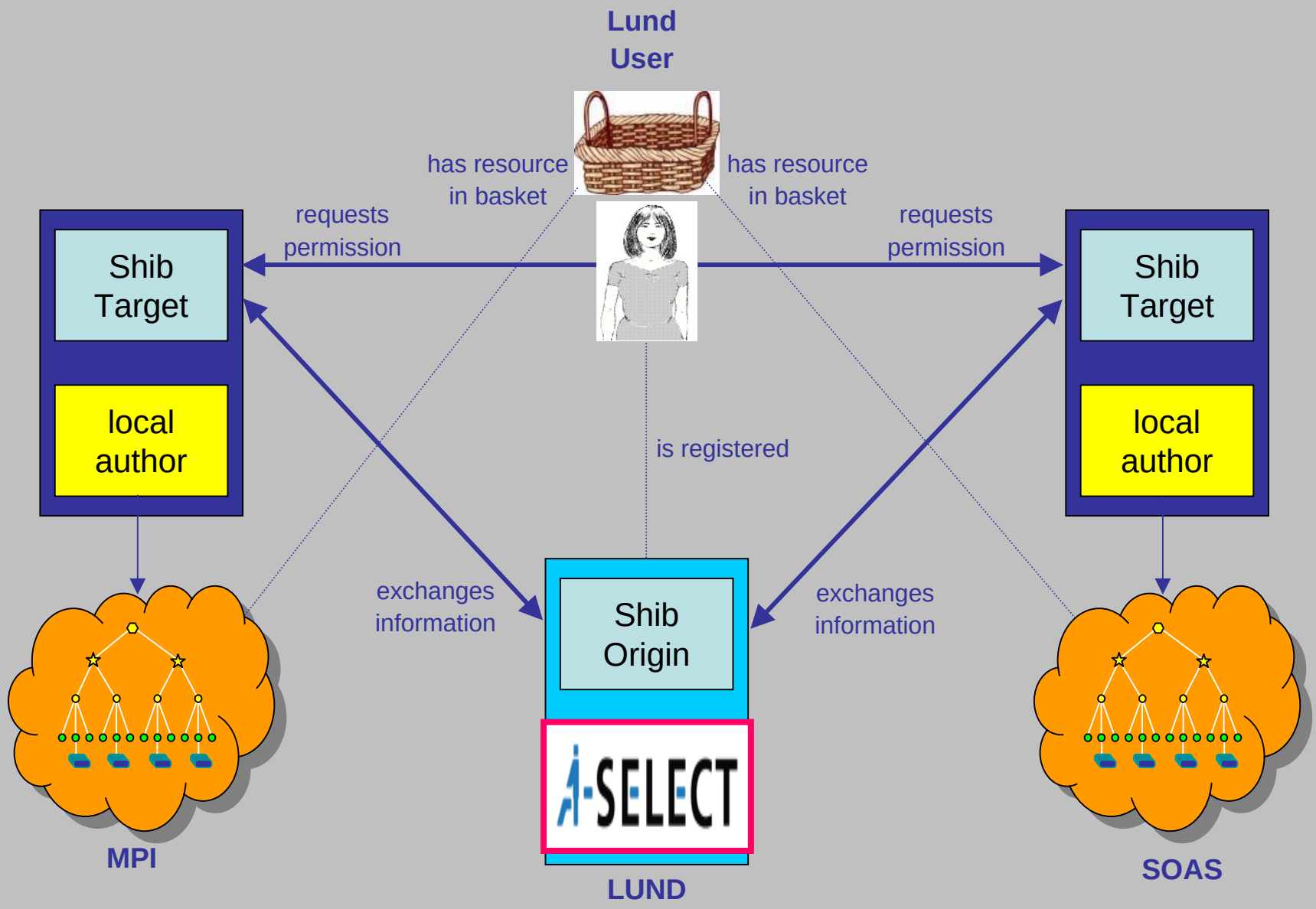
What else about Shib

- WAYF can set a cookie in the user's browser to bypass forms
- if home organization's authentication supports single sign-on and the user has already a session open, no further logins are required
- exchange of info is done via SAML (Security Assertion Markup Language)
- Shib creates interoperability between local AA systems
- suggestions for attributes and roles can be found in EDUCAUSE
- tools for the management of attribute release policies (users have to do more!)
  - ways to set defaults
  - still thinking of optimizations
- Shib allows complex configurations, i.e. a service can be in several federations
- increasing amount of users to be managed – users we don't know
  - can rely on other institutions to manage trustful users
  - for example: indigenous people

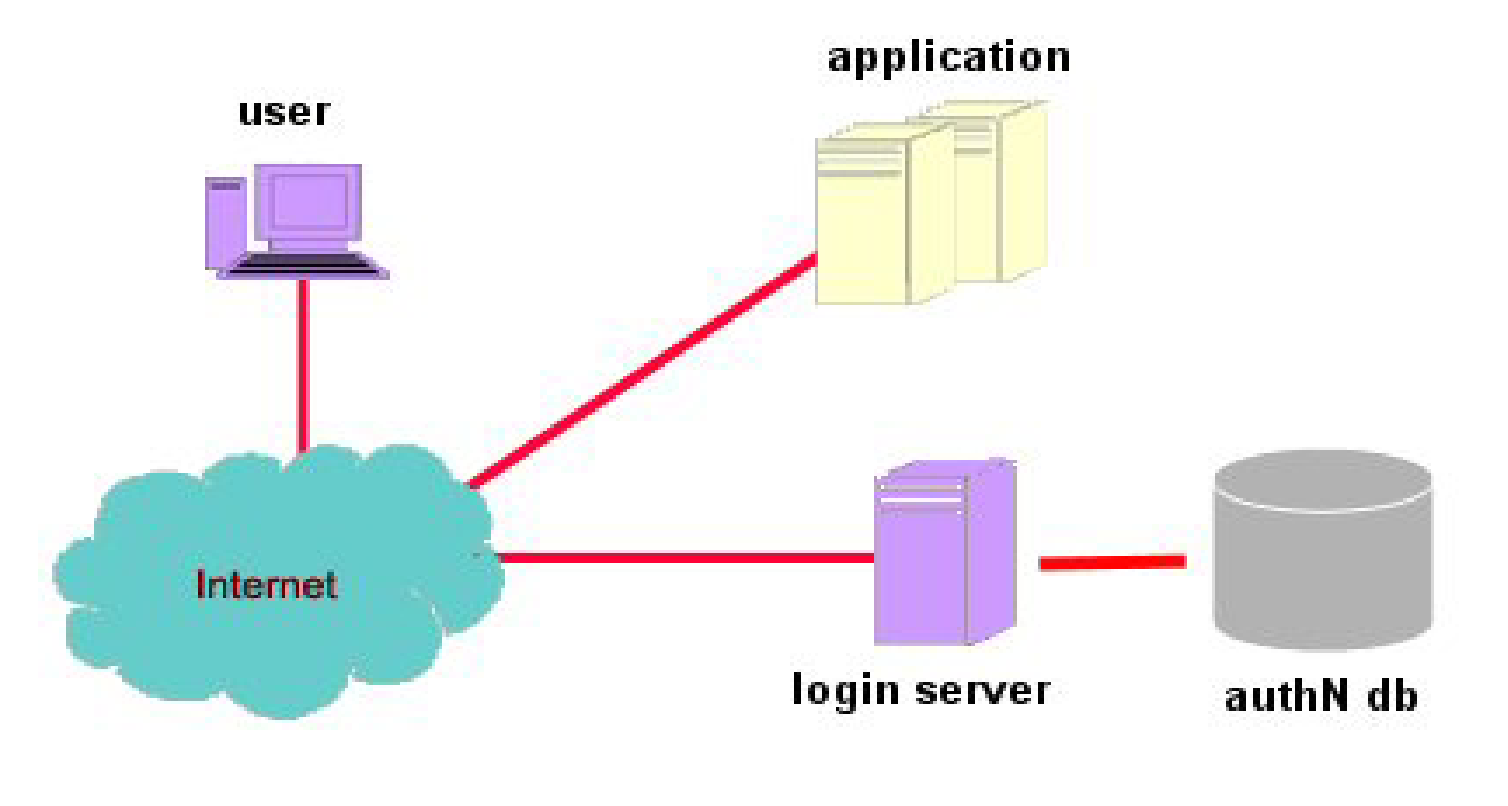
DAM-LR Scenario with Shib



Additional Authentication?



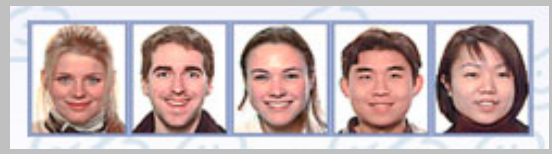
Professional Authentication



authentication by specialized institution

Authentication layers

- IP address
- Username / password
  - LDAP / Active Directory
  - RADIUS
  - SQL
- Passfaces
- PKI certificate
- OTP through SMS
- OTP through internet banking
- Tokens (SecurID, Vasco, …)
- Biometrics
- …

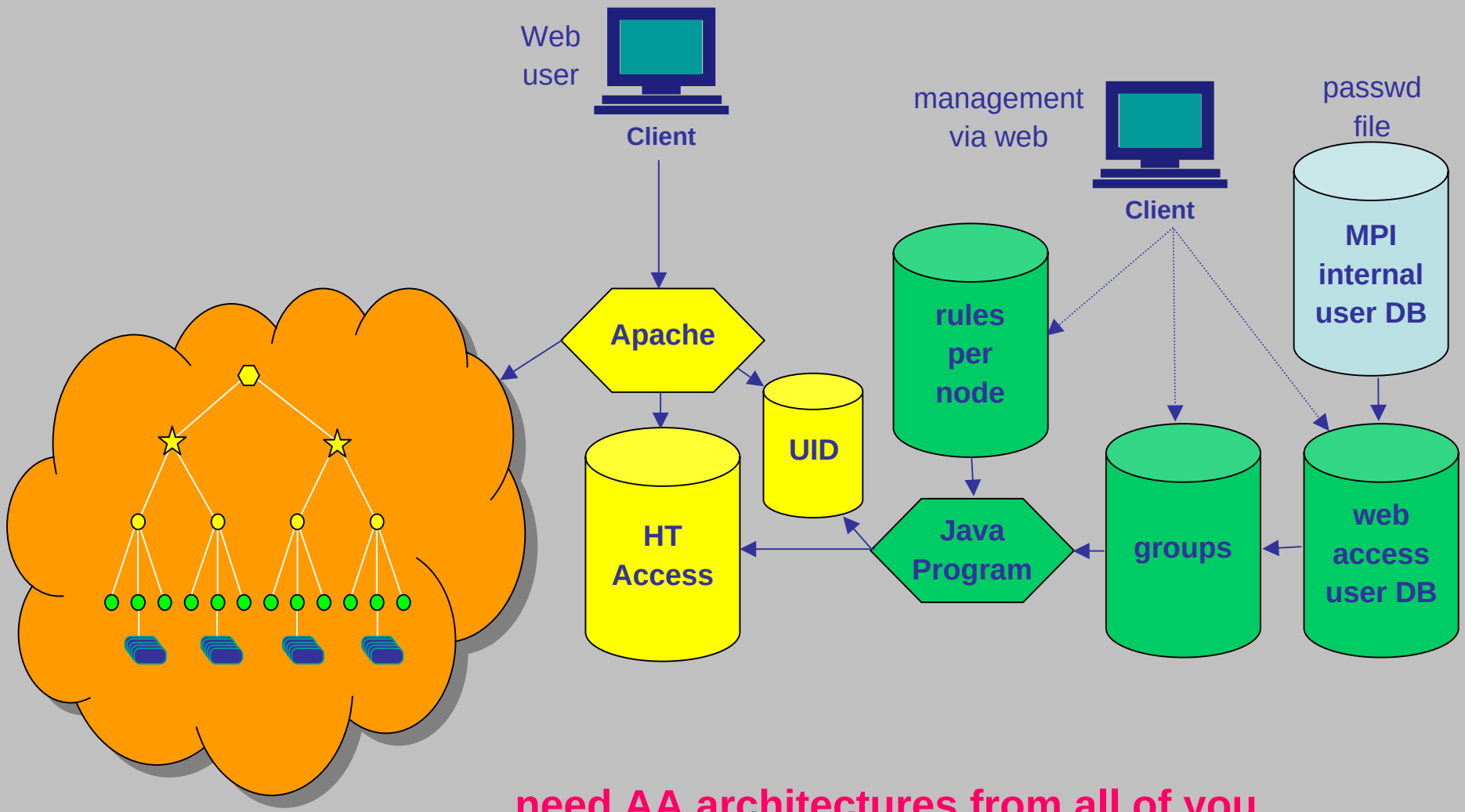


do we need this now? – not per se within DAM-LR

Actions/Questions

- SOAS, LUND, INL and MPI have to form a federation
- have to setup / agree on the following
  - Public Key Infrastructure for all our servers
  - which user attributes do we want to share/use
  - do users want to maintain this
  - does access via attributes work – how granular has it to be (name?)
  - how to exchange user information (how do we know who is registered)
  - how to locate servers (exchange configurations)
  - how to exchange authorization information (when having copies somewhere)
  - how does Shib interact with our local authentication and authorization mechanisms
- have to start looking into these aspects

MPI AA Architecture



need AA architectures from all of you
part of the definition

And now?

Let’s start thinking and analyzing

It’s challenging anyhow
Have to bring linguistics into the Grid game.

will start with
- a Wiki about requirements specs
- a Wiki about Shibboleth

Goal:
- in November ready with evaluation and concept