- Architecture
- State Distributed Solution
- Definitions

DAM-LR
A Grid
Project

DAM-LR Objectives

DAM-LR Archive integration aims at direct user benefits:

- Access mechanism: single user identity, single sign-on.

- Resource discovery: sharing a single metadata set for searching & browsing

- Referencing "archived resources" using a single system of identifiers.
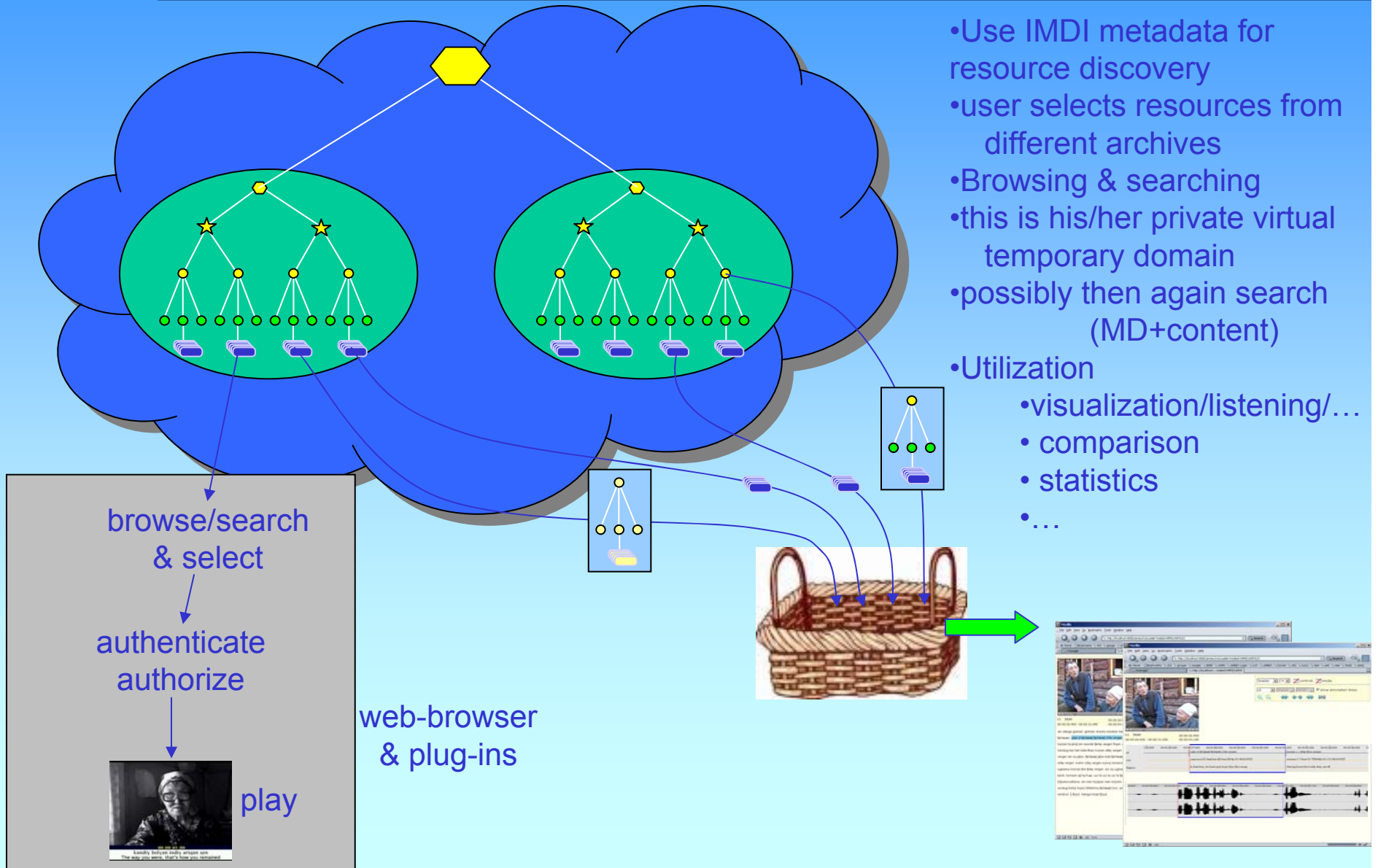
DAM-LR Architecture

- Secure communication and identification using PKI & certificates. All partners become RA recognized by a CA on TERENA/TACAR list of EUGridPMA.
- Integrated Metadata domain using IMDI
- Unique persistent resource identifiers management using the HS
- Federation wide single user identity using Shibboleth
- [Federation wide authorization system using the HS]

Guideline for overall architecture:

- if possible the use or implementation of a component or service should not be obligatory
- Some archives may not be able to support all components
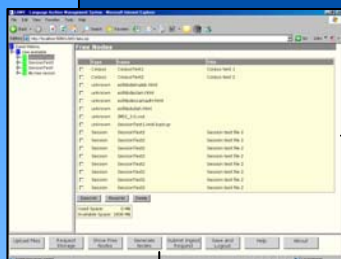
Shared Metadata Domain

- Use IMDI metadata for resource discovery
- user selects resources from different archives
- Browsing & searching
- this is his/her private virtual temporary domain
- possibly then again search (MD+content)
- Utilization
  - visualization/listening/…
  - comparison
  - statistics
  - …

browse/search & select

authenticate authorize

play

web-browser & plug-ins

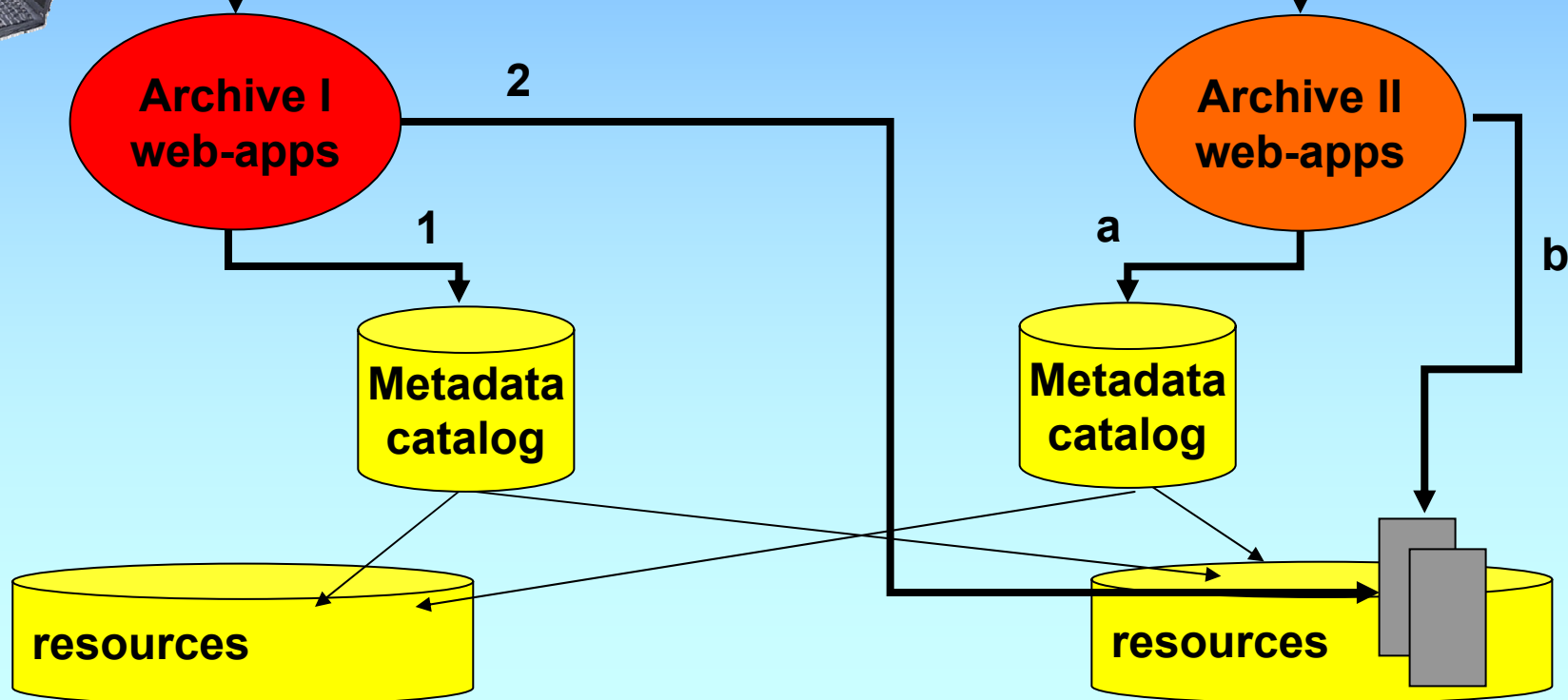**n1**   this is the scenario that show the use of the shared metdata domain
          nn, 21-05-2006

Archive Interoperability: Resource utilization

Scenario requires:

- IMDI metadata – correct type/format specs

- Shibboleth *

Archive I web-apps

2

1

Metadata catalog

resources

Archive II web-apps

a

b
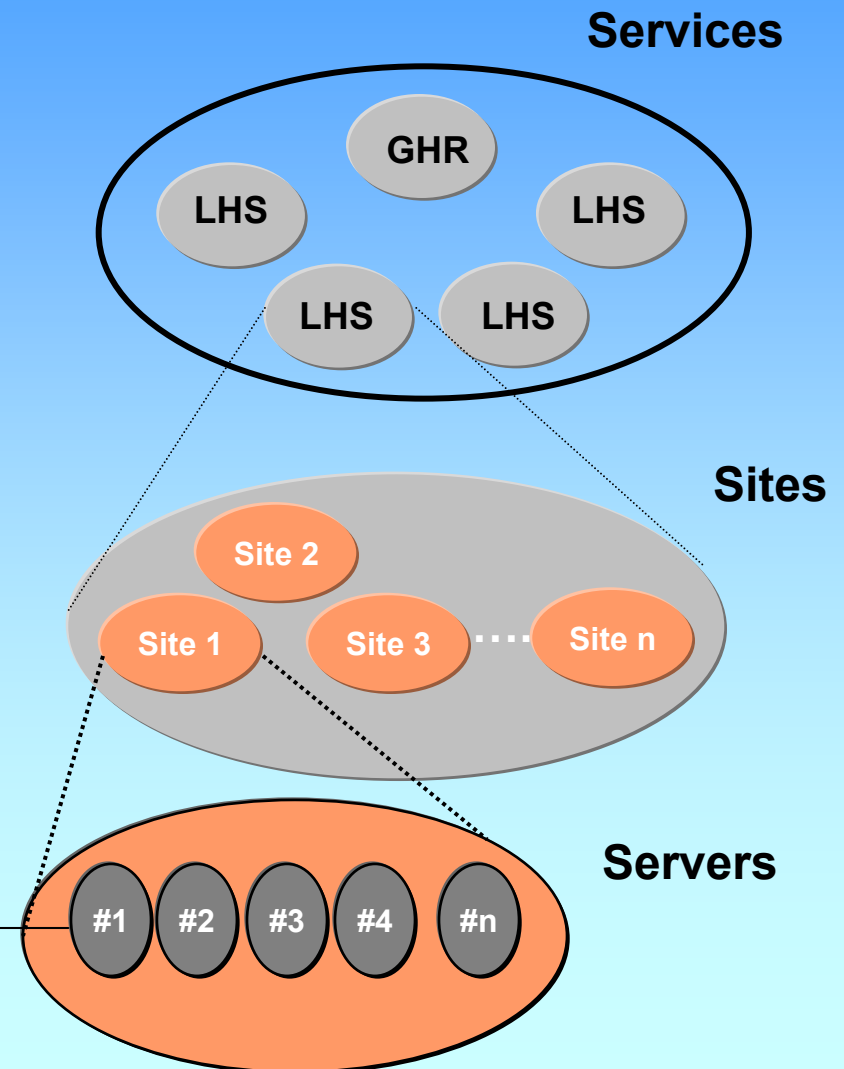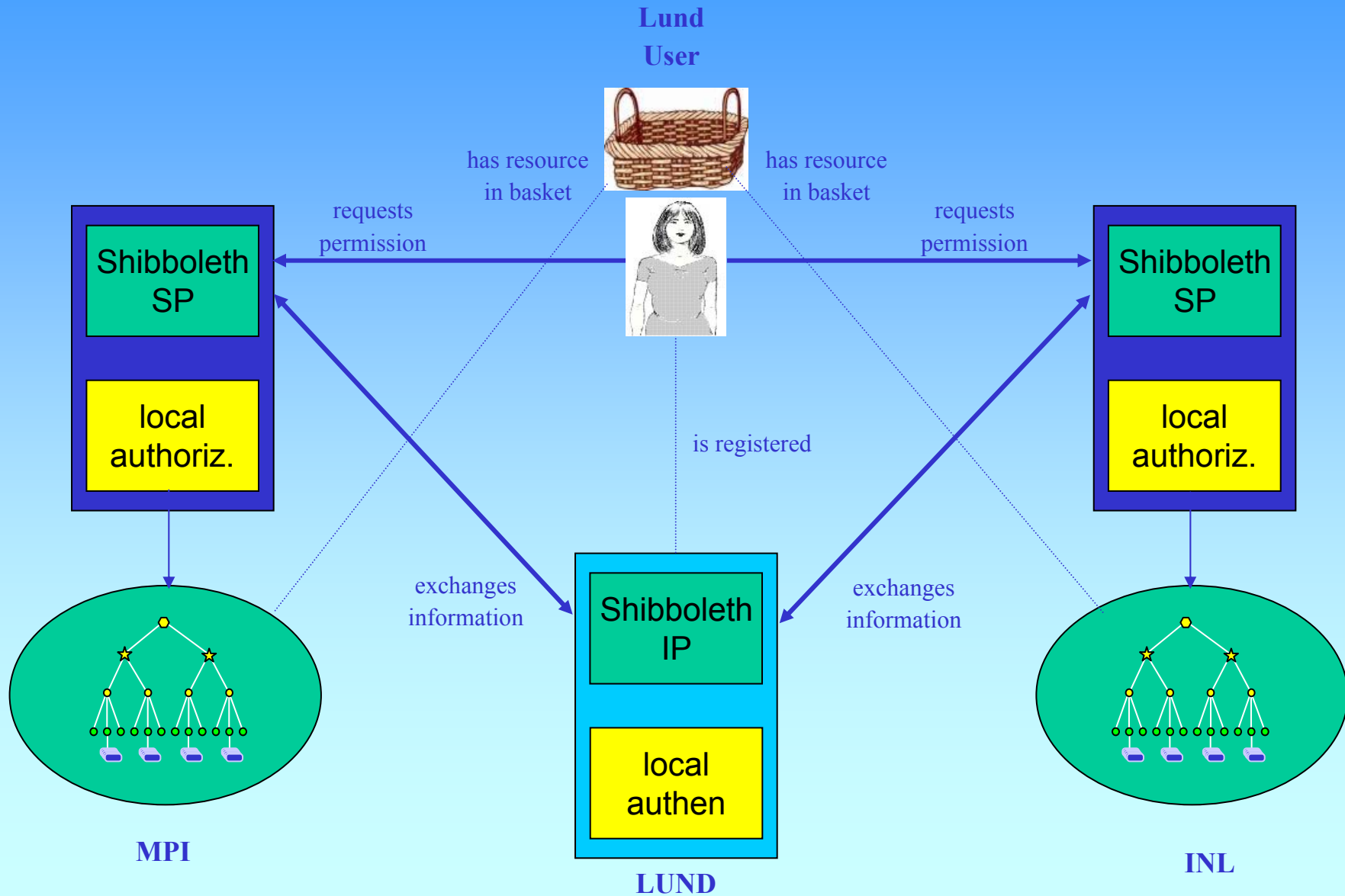
Metadata catalog

resources

Handle System

- HS system manages URIDs: solution for avoiding dead links
- Every object is issued a URID and URID/URL combinations are stored in handle DB.
- to obtain a valid URL a resolving step is needed.
- URID (Handle) is combination of pre- end postfix: **1839/00-0000-0000-0000-0000-4**
- maximal independence and flexibility: every partner manages its own prefix and handle server
- DAM-LR partners will replicate each other LHS to assure redundancy

Services

GHR

LHS   LHS

LHS   LHS

Sites

Site 2

Site 1   Site 3   · · · ·   Site n

Servers

#1   #2   #3   #4   #n

18390/000-4   URL   4   http://corpus1.mpi.nl/IMDI/top

                  URL   8   http://dam-lr.sol.lu.se/IMDI/top

DAM-LR Scenario with Shibboleth
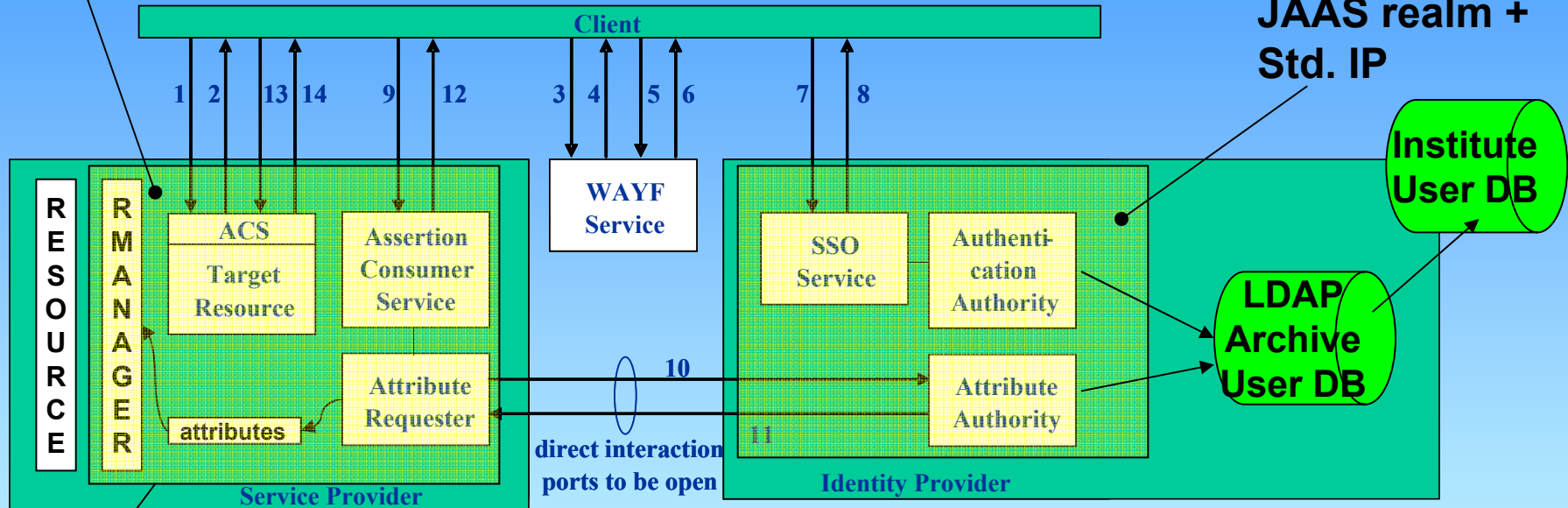
Lund
User

has resource in basket

has resource in basket

requests permission

requests permission

is registered

Shibboleth SP

local authoriz.

Shibboleth SP

local authoriz.

exchanges information

exchanges information

Shibboleth IP

local authen

MPI

LUND

INL

Shibboleth Scenario

Apache WS + mod_shib

TOMCAT + JAAS realm + Std. IP

Client

1 2 13 14 9 12 3 4 5 6 7 8

RESOURCE

R MANAGER

ACS
Target Resource

Assertion Consumer Service

attributes

Attribute Requester

WAYF Service

10

direct interaction ports to be open

Service Provider

SSO Service

Authenti-cation Authority

Attribute Authority

11

Identity Provider

Institute User DB

LDAP Archive User DB

htaccess authorization

1 Get Resource
2 Redirect (302)
3 Get Form
4 Send Form (200)
5 Submit Form
6 Send Cookie and redirect (302)
7 Request Authentication

8   Authentication Response
9   Send an Assertion Profile
10 Request Attributes
11 Send Attributes
12 Redirect with attributes
13 Send attributes for check
14 Provide Resource

State of Distributed Solution

Feb 2007	INL	Lund	MPI	SOAS
IMDI metadata	CGN and IFA corpus ready. Maybe more resources to come?	ready	ready	IMDI provider Yes, partially IMDI mapping others?
Portal	ready	ready.	ready	IMDI portal in place
Certificate	ready	?	ready	yes
HS	Yes (run proxy too)	Need update ref. impl.	Own HS: ready Backup resolvers:?	yes
Shibboleth	IP yes, SP ?	IP yes, SP yes, limited attributes	IP yes, SP yes (test server), limited attributes	?

Yet to do for DAM-LR

- Improve stabilize local archiving systems
- Add further resources & metadata
- Finalize the concatenated attribute setup:
  - Allow use of federation wide unique uids in the authorization records in htaccess file.
- Complete and share the shibboleth's metadata.xml file, identifying all the DAM-LR partners with their certificates.
- Setup multiple WAYFs
- Adapt AMS to use federation-wide unique uids *
- Develop RRS to identify/authenticate external user requests for access.*

* MDI context

Yet to do for DAM-LR

- Extensive test integrated systems (WP11)
- Improve metadata catalogs. Better structuring. DAM-LR branded portal + showcase
- Estimate cost for integration single archive
- Create/update installation guides
- Final version definition report containing only final agreements
- Minimal requirements for participation in DAM-LR federation

* MPI prototype

Definitions I

- IMDI metadata (version 3.0) for
  - Exchange format
  - Harvesting via top node URL
  - Specify type/format of resources *
- Handle System
  - One prefix per partner archive
  - Handle postfix format is free
- PKI Certificates
  - the partners in DAM-LR will become at least RA and will get their certificates from the corresponding national root authorities that are supported by TERENA TACAR

- ## User Authentication
  - – Method is left open. Prototype uses Tomcat CMS

- ## User Identification
  - – Shibboleth provides the following user attributes based on eduPerson set:
    - first name, last name
    - uid -> eduPersonPrincipalName
    - affiliation -> eduPersonAffiliation
    - email -> email
    - *** {hostingInstitute}_{uid} -> PrincipalName
    - hostingInstitute
    - *class+
    - *status -> status

* extra wrt. eduPerson
*** new wrt. def. report

- Authorization
  - Dissemination of auth info via HS Format: *"userID1 userID2 … userIDN"*
  - Use federation wide unique identifiers:
    - e.g. "INL_LA:MyUID"
- hostingInstitute codes:
  - INL -> INL_LA
  - Lund-> SOL-LU
  - MPI -> MPINLA
  - SOAS -> ?

Maybe to do for DAM-LR

- Investigate and create an efficient way for copying resources or synchronizing whole corpora between archives.
- Make the shared metadata domain more efficient by making the different metadata catalogs housed by partners that copy the MPI reference model "complementary" rather than "duplicated".
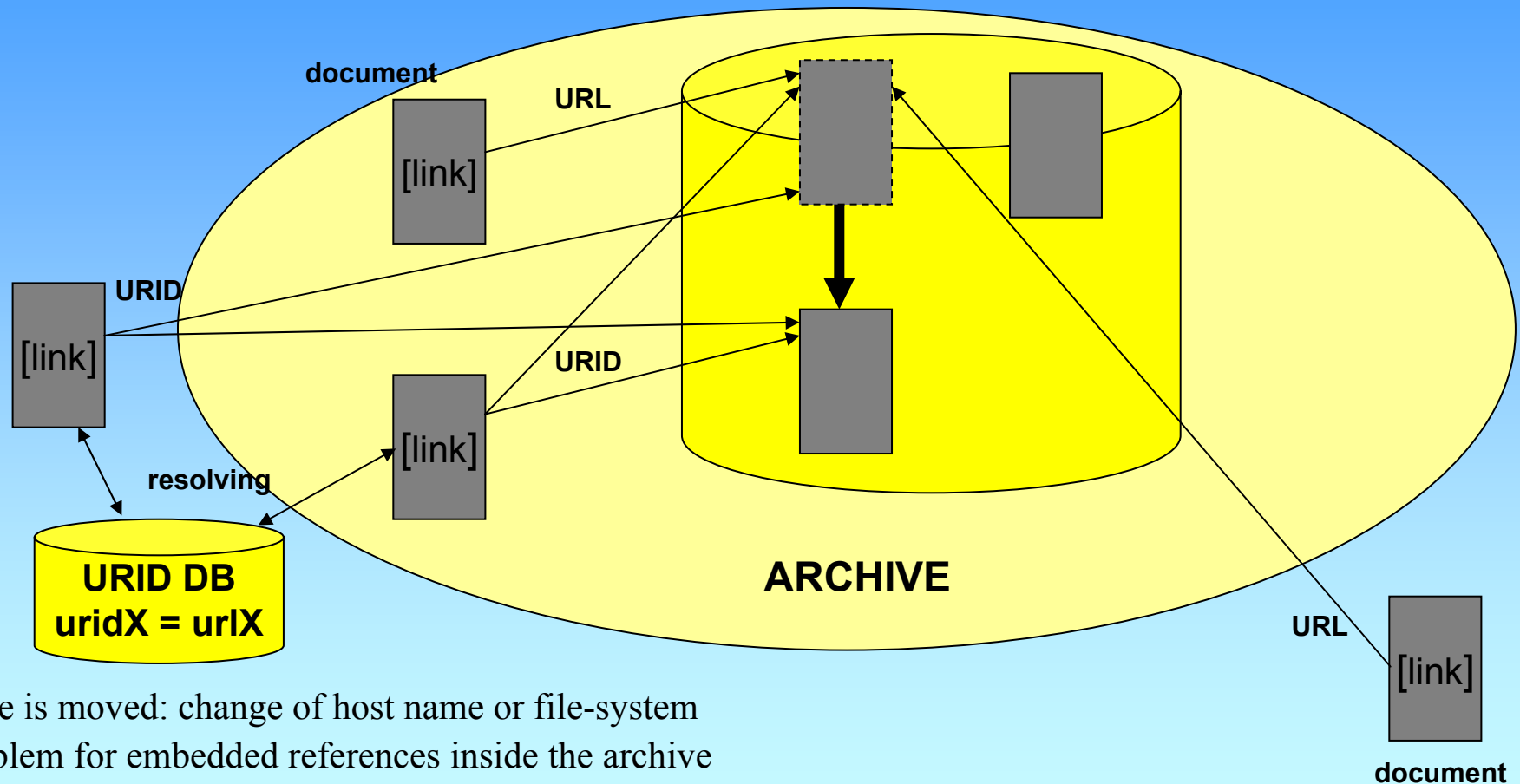- Adapt selected tools for Shibboleth scenario: ANNEX,. ….

The End

Unique Resource Identifiers (URIDs)
(or how to avoid linkrot)

Resource is moved: change of host name or file-system

- Problem for embedded references inside the archive
- But especially outside the archive

  Solution: separate object identity and object location, store mapping in separate system

- Only change resource location using special mover/copier tool

| 18390/000-4 | URL | 4 | http://corpus1.mpi.nl/IMDI/top |
| | URL | 8 | http://dam-lr.sol.lu.se/IMDI/top |