



# Integrating PIDs in a Distributed Archive Environment

Daan Broeder

MPI for Psycholinguistics

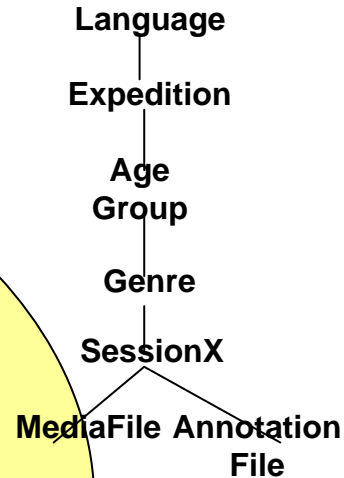
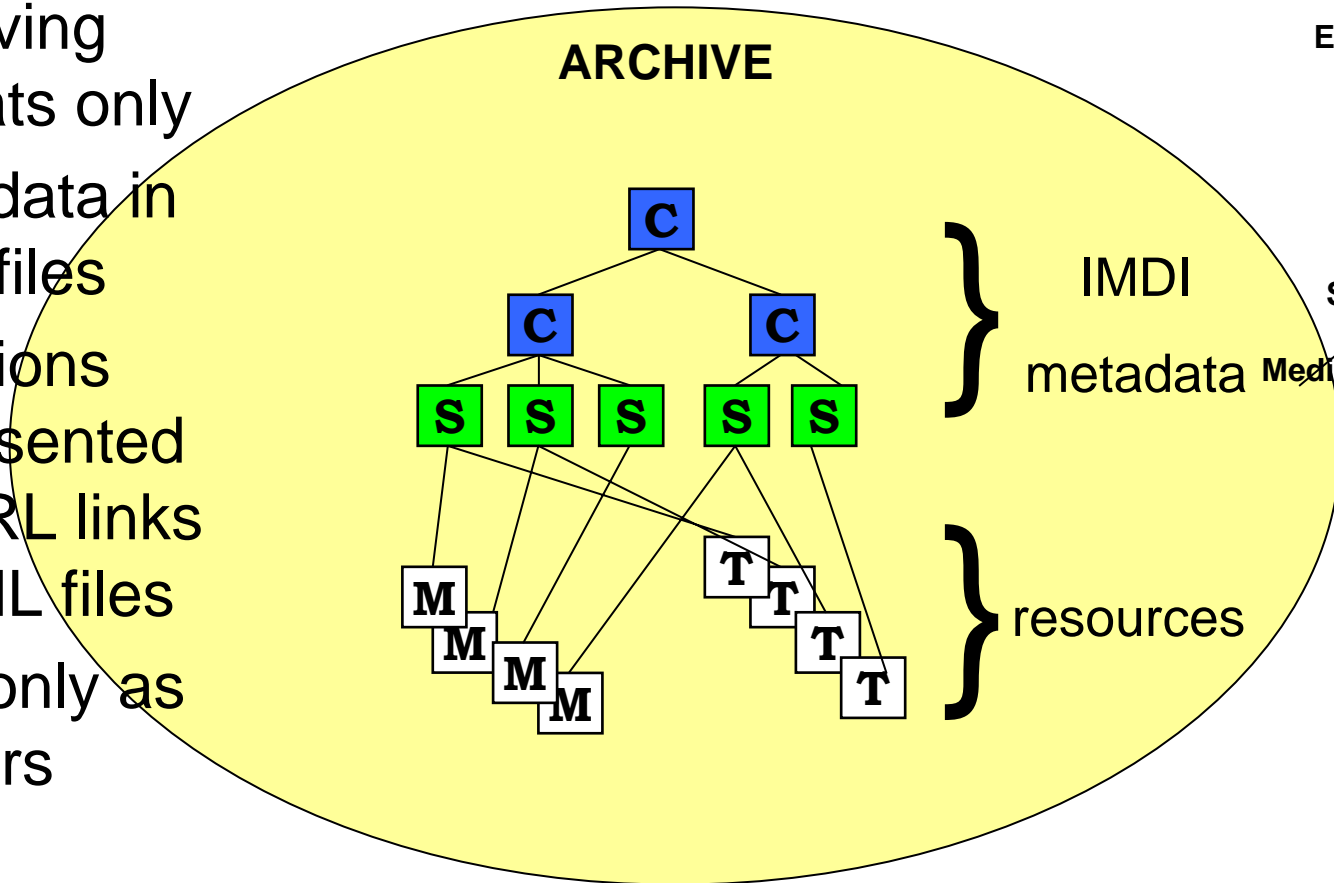


# Introduction

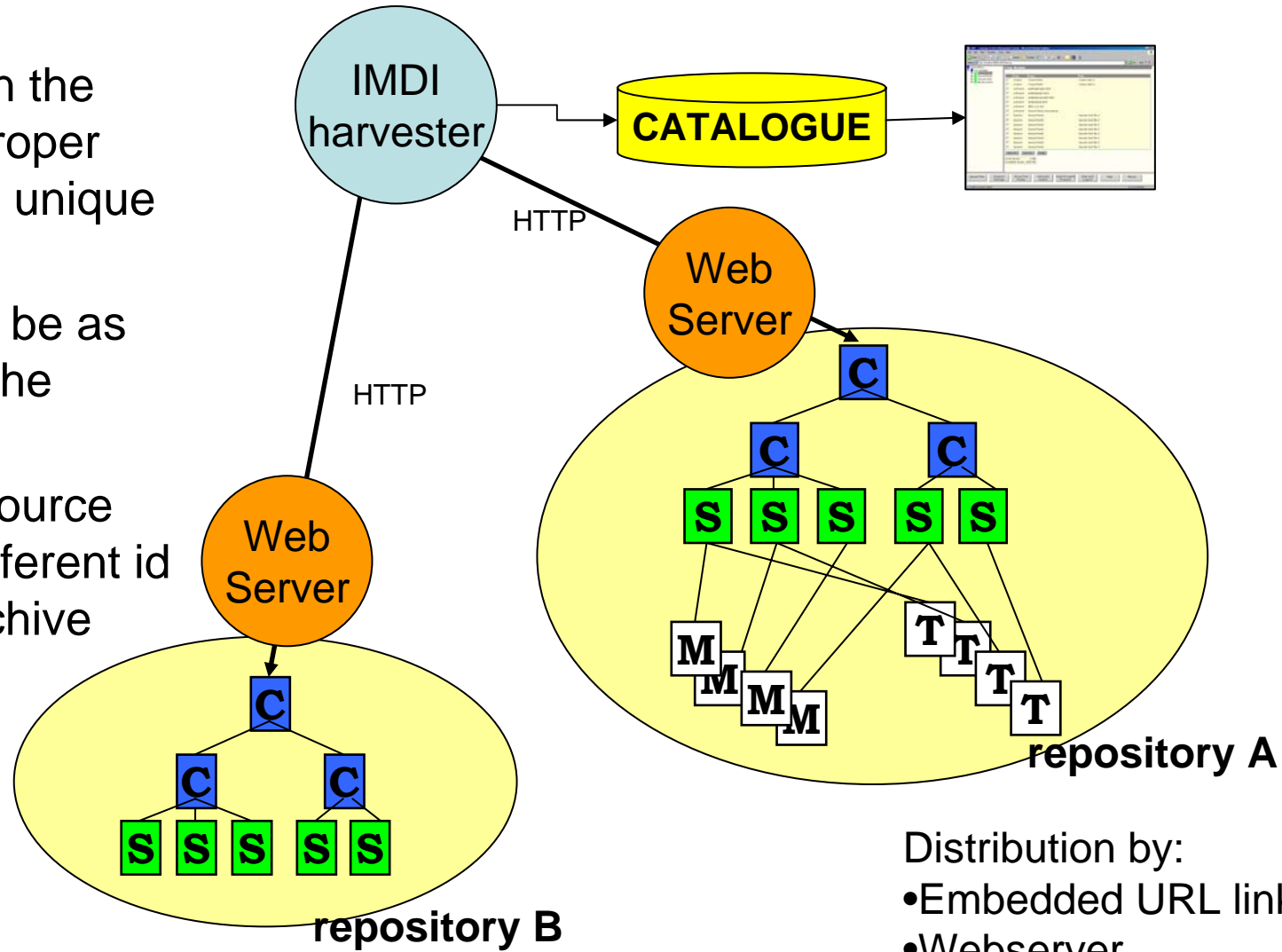


- The MPI for Psyl. Archive Environment
- Introducing PIDs in the DAM-LR project
- Copying resources between archives
- Referencing resource parts

- Archiving formats only
- Metadata in XML files
- Relations represented by URL links in XML files
- DBs only as helpers



- Every object in the catalogue a proper (archive wide) unique identifier.
- This id should be as persistent as the archive
- The same resource can have a different id in different archive catalogues

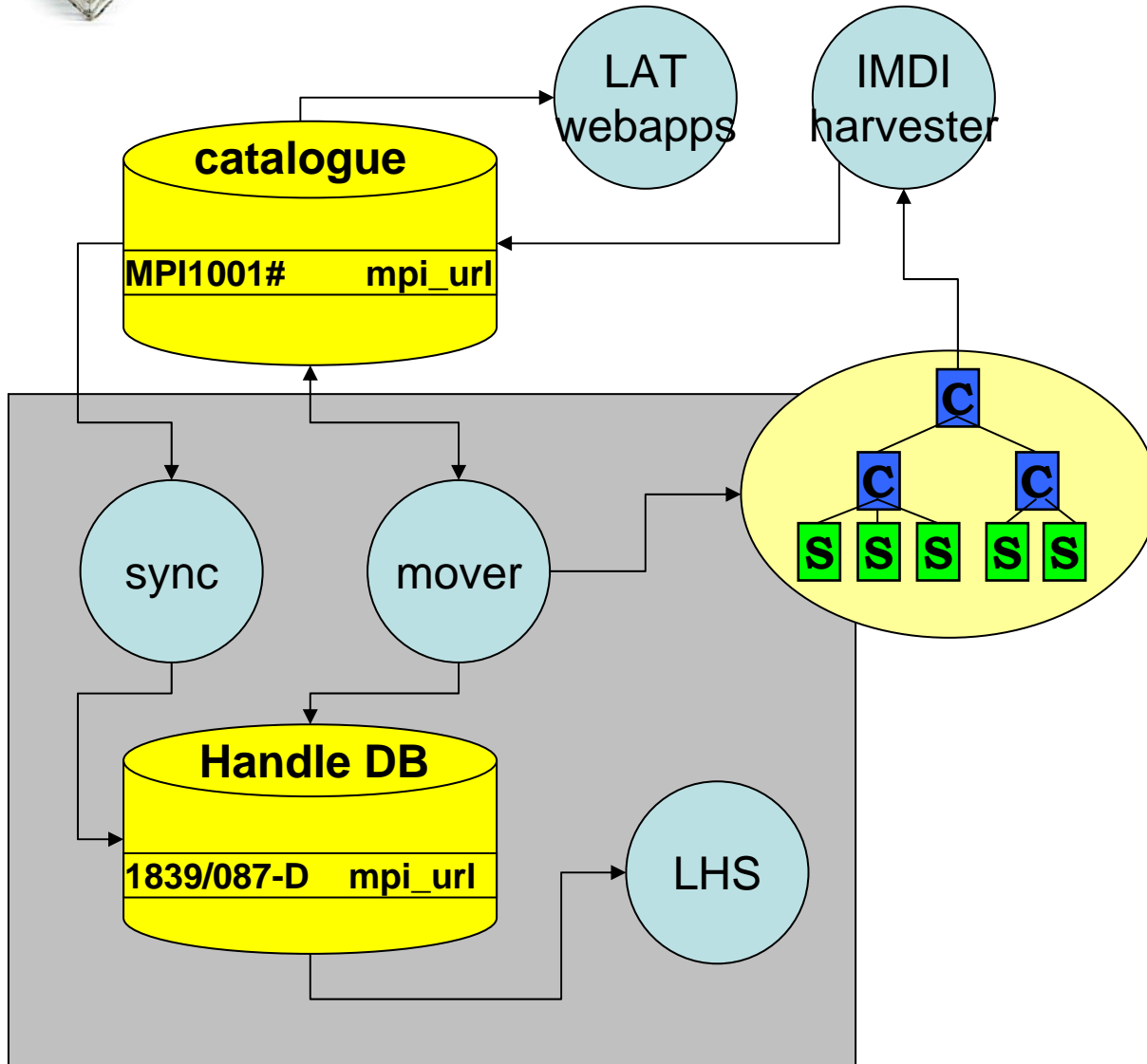


- Distribution by:
- Embedded URL link
  - Webserver
  - Low tech!!!



## Four Language Resource archives creating a federation

- Shared metadata domain
  - Single resource catalogue
- Identity federation using Shibboleth
  - Single user identity in the federation
  - SSO
- Enable the use of PIDs for resource reference and citation. Requirements a.o.
  - Self-sufficiency
  - Robustness/high availability
  - Support multiple resource instances
  - Use for dissemination other info: access rights
  - -> Handle System



## Integration

- it should be an optional extension
- Make sure HS is not SPF
- IMDI/LAT SW functions also without HS

## Issue handles for objects

Only for local resources

## Need special tools

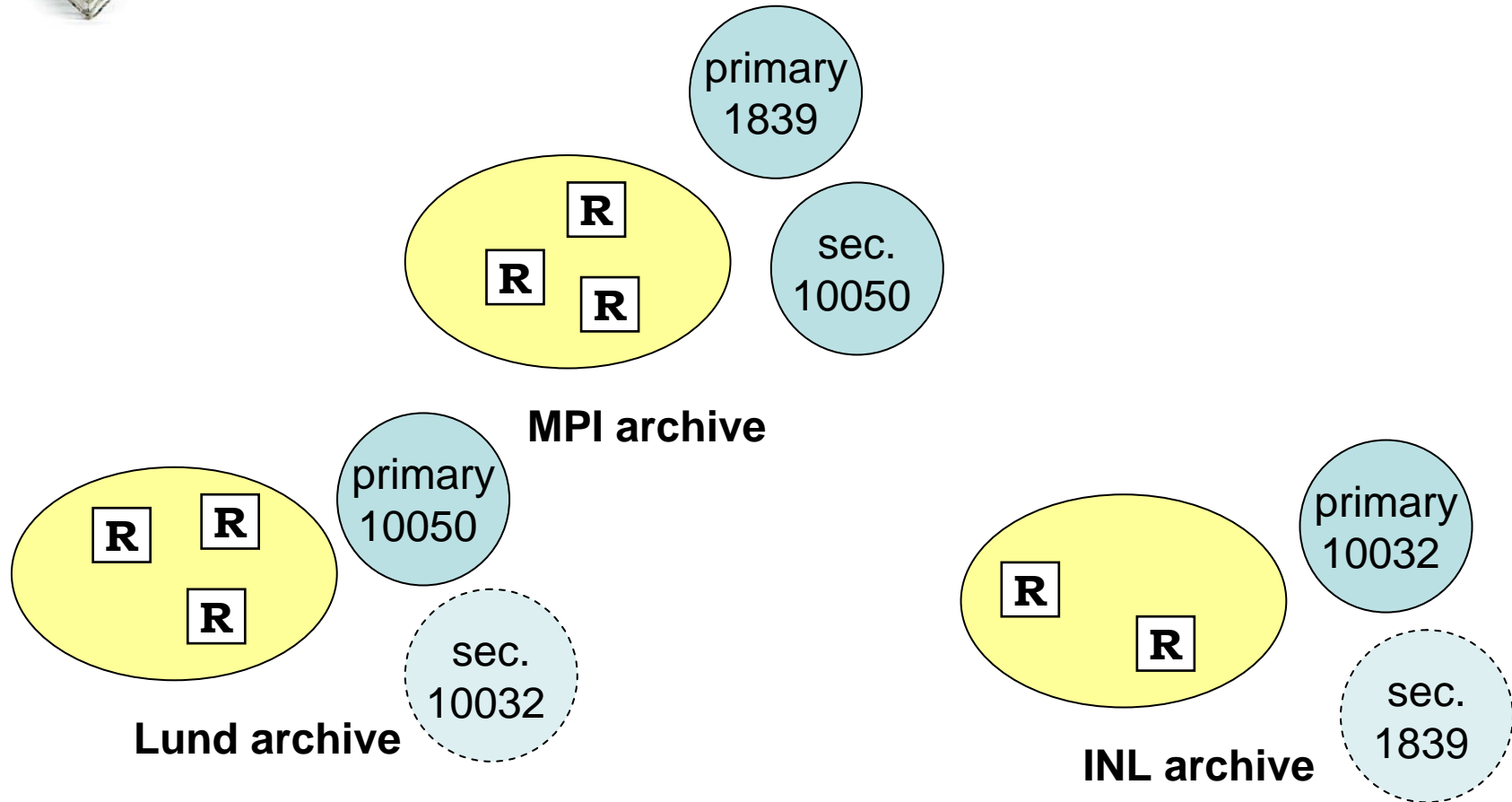
- Mover
  - Updates Handle DB + catalogue
  - Updates IMDI XML files\*
- Restore operations
  - Recreate the Handle DB (and others) from scratch



# handles in IMDI files



```
<?xml version="1.0" encoding="UTF-8"?>
<METATRANSCRIPT ArchiveHandle="hdl:1839/00-0000-0000-0005-82B0-2"
  Date="2006-07-18" FormatId="IMDI 3.0"
  Originator="Editor - Profile:SESSION.Profile.xml" Type="SESSION" Version="1"
  xmlns="http://www.mpi.nl/IMDI/Schema/IMDI"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.mpi.nl/IMDI/Schema/IMDI ./IMDI_3.0.xsd">
<Session>
  <Name>DBD_RIF_14_12_01_064</Name>
  <Title>Dutch Bilingualism Database, Ethnic Dutch, Session 64</Title>
  .....
<MediaFile>
<ResourceLink ArchiveHandle="hdl:1839/00-0000-0000-0004-DC6B-0">
  http://corpus1.mpi.nl/qfs1/media-archive/dbd\_data/boumans/T-Cult/Metadata/./Media/dbd\_rif\_14\_12\_01\_064.wav</ResourceLink>
  .....
```



IMDI-Browser about user: anonymous login logout

- IMDI-corpora
  - WelcomeToIMDIDomain.html
  - index.html
  - AILLA
  - ANDES
  - Bavarian Archive for Speech Signals (BAS)
  - CGN en IFA Corpora
  - CORP-ORAL
  - Coralrom
  - Corpora Soas
  - DBD
    - Conditions\_for\_use\_DBD.pdf
    - Conditions\_for\_use\_DBD.txt
    - dbd
    - Berber
      - E-rramdani
      - El Aissati
        - tcult (2000)
          - Morocco
          - The Netherlands
            - Utrecht (Lombok)
              - DBD\_RIF\_14\_06\_07
                - nordin\_B.cha**
                - DBD\_RIF\_14\_09\_07
                - DBD\_RIF\_14\_09\_07
                - DBD\_RIF\_14\_09\_07
                - DBD\_RIF\_14\_09\_07
                - DBD\_RIF\_14\_09\_07
                - DBD\_RIF\_14\_09\_07
                - DBD\_RIF\_14\_12\_07
                - DBD\_RIF\_14\_12\_07
              - DBD\_RIF\_14\_12\_01\_064
- Dutch Ethnolects
- Moluccan Malay
- Moroccan Arabic
- Papiamentu
- Russian
- Sranan:Sarnami
- Test corpus
- Turkish
- DoBeS archive
- ECHO
- ESF corpus
- Endangered Languages
- GTRP Corpus

## Bookmark an archive node or resource

You selected a node from the archive:  
name: nordin\_B.cha  
title: Resource "nordin\_B.cha" from "nordin"  
internal nodeid: MPI521258#  
handle: 1839/00-0000-0000-0007-F42A-6  
url: http://corpus1.mpi.nl/qfs1/media-archive/dbd\_data/Aissati/tcult\_2000\_/The\_Netherlands/Utrecht\_Lombok\_/Annotations/nordin\_B.cha

To save a reference to the IMDI Browser with the tree opened to this node, bookmark the following link:  
[nordin\\_B.cha](#)

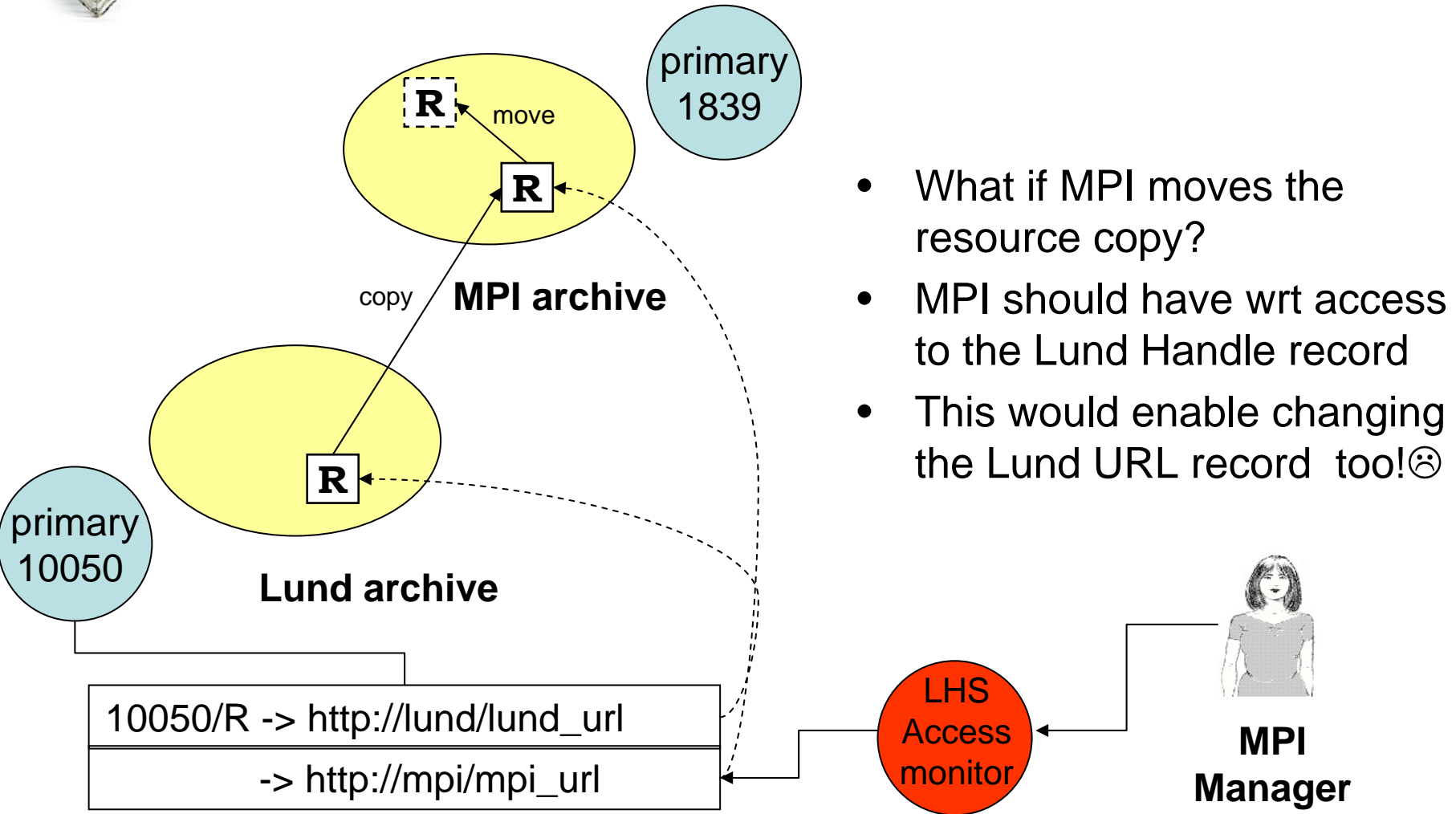
Some browsers support direct use of "handle URIs", if that is the case you can bookmark the following link:  
Handle link to resource: [nordin\\_B.cha](#)

To save a link to the resource using the central handle resolver, bookmark the following link:  
Handle link via central resolver to resource: [nordin\\_B.cha](#)

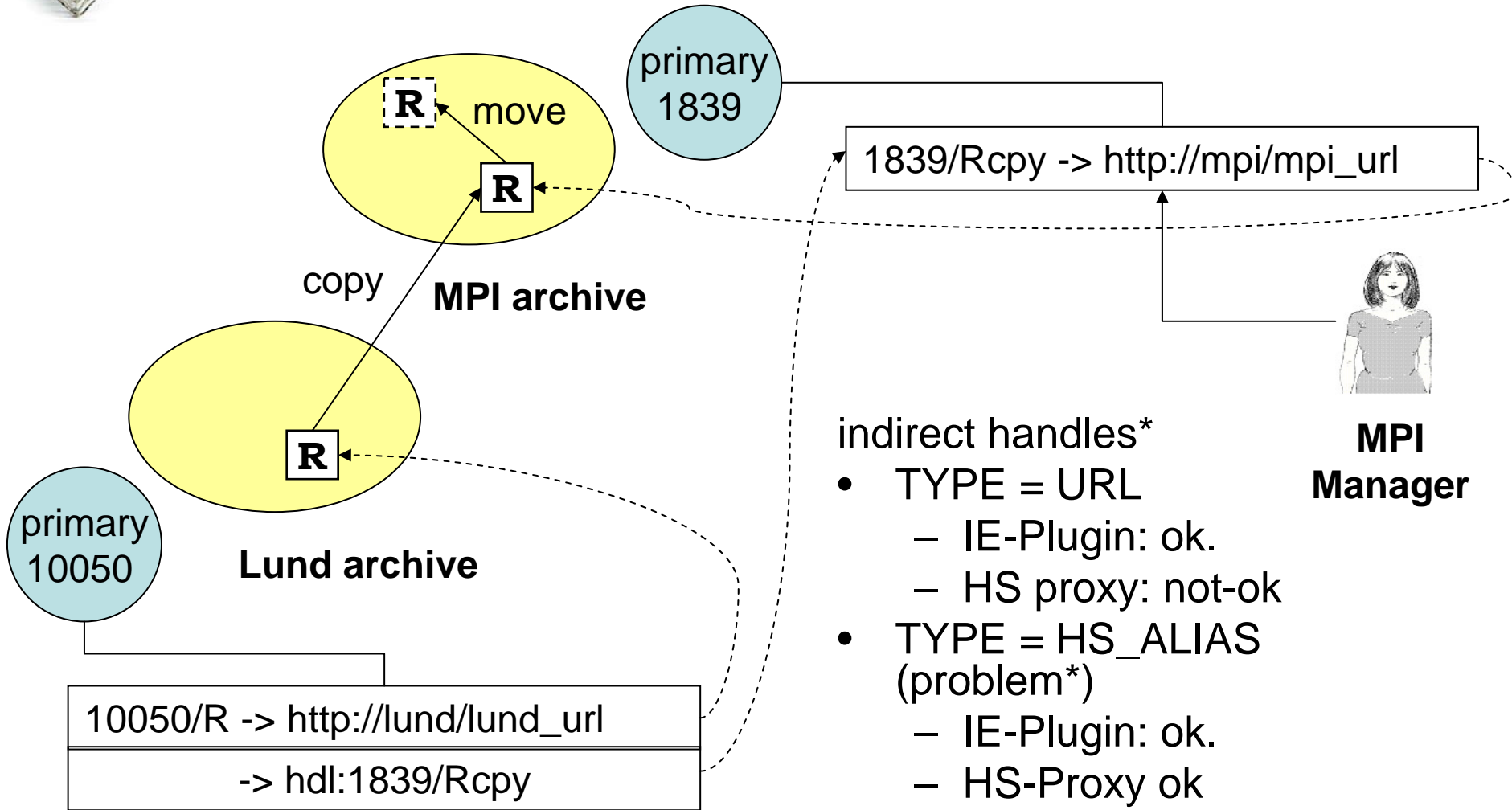
To save the URL for this resource which is NOT guaranteed to be persistent, bookmark the following link:  
URL link to resource: [nordin\\_B.cha](#)

## Cite an archive node or resource

You can create citations for this resource by embedding the following text in your document:  
Title: Resource "nordin\_B.cha" from "nordin"  
Author(s): ,  
Archive: unknown  
Resource format: text/x-chat  
Published: Tue Apr 17 14:25:05 CEST 2007  
Reference: 1839/00-0000-0000-0007-F42A-6  
Link: http://hdl.handle.net/1839/00-0000-0000-0007-F42A-6

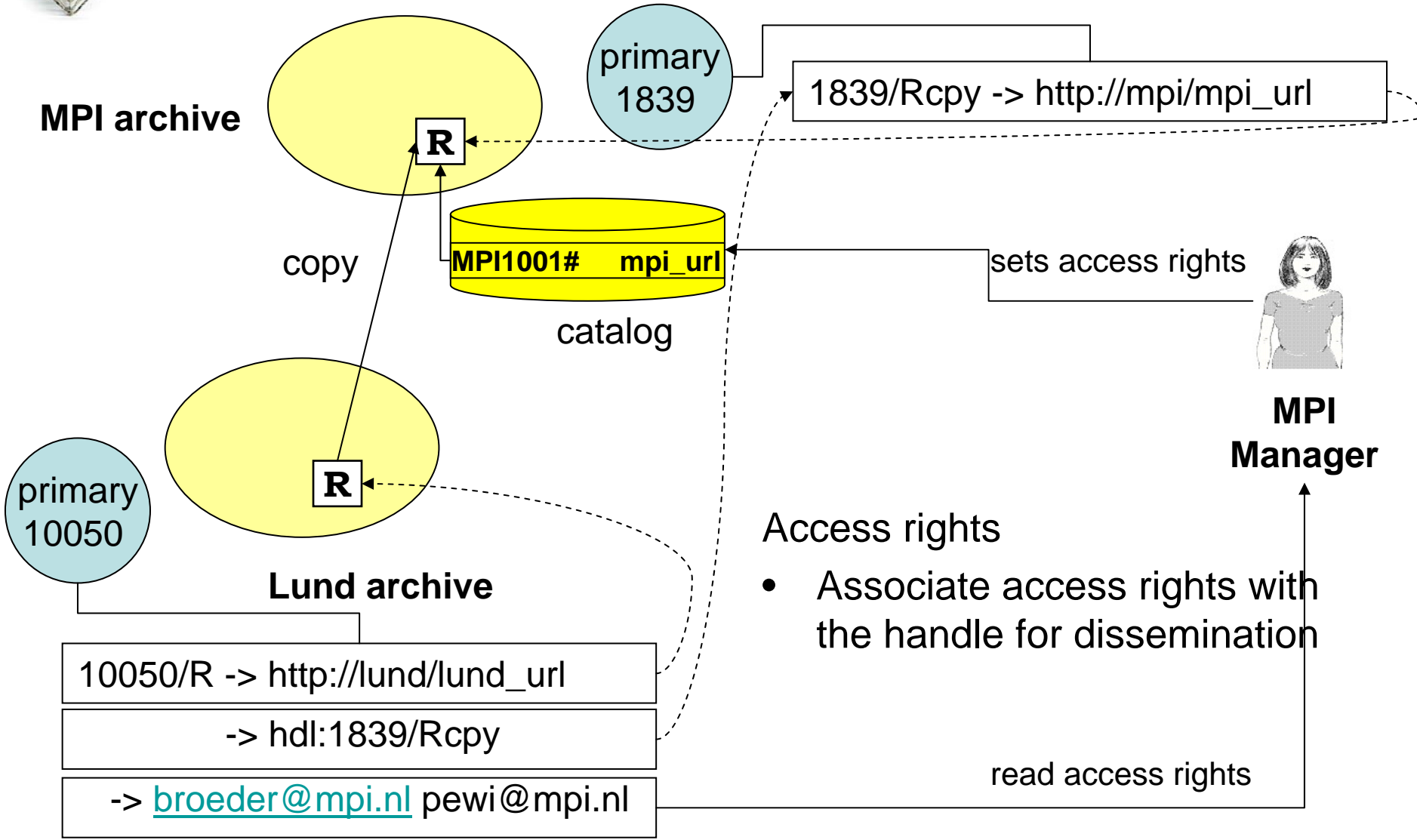


- What if MPI moves the resource copy?
- MPI should have wrt access to the Lund Handle record
- This would enable changing the Lund URL record too! 😞



## indirect handles\*

- TYPE = URL
  - IE-Plugin: ok.
  - HS proxy: not-ok
- TYPE = HS\_ALIAS (problem\*)
  - IE-Plugin: ok.
  - HS-Proxy ok
- Status of 1839/Rcpy handle?
  - Use in documents?





# What's in our handles



## Handle System®

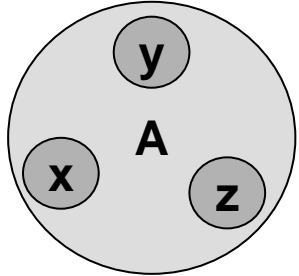
Handle Values for: 1839/00-0000-0000-0001-5206-0

Index	Type	Timestamp	Data
100	HS_ADMIN	NA	handle=0.NA/1839; index=200; [add val,read val]
1	URL	NA	<a href="http://corpus1.mpi.nl/qfs1/media-archive/dbd_data/E-rAMDani/E-rAMDani_2000/Grade_1_children_Morocco/01_Zakaria/Media/">http://corpus1.mpi.nl/qfs1/media-archive/dbd_data/E-rAMDani/E-rAMDani_2000/Grade_1_children_Morocco/01_Zakaria/Media/</a>
2	CRAWLTIME	NA	2006-04-01 17:01:35.659
3	FILESIZE	NA	33132040
4	ONSITE	NA	true
6	FILETIME	NA	2005-07-18 15:52:02.0
12	READRIGHTS	NA	hasloe@mpi.nl aliesm@mpi.nl mategg@mpi.nl broeder@mpi.nl micre@mpi.nl linmee@mpi.nl rskiba@mpi.nl backus@mpi.nl woo
13	WRITERIGHTS	NA	hasloe@mpi.nl mategg@mpi.nl broeder@mpi.nl rskiba@mpi.nl woonic@mpi.nl yahya@mpi.nl jacrin@mpi.nl albert@mpi.nl pberc

[Handle System Web Site](#)

Please contact [hdladmin@cnri.reston.va.us](mailto:hdladmin@cnri.reston.va.us) for your handle questions and comments.

1839/A  
1839/x  
1839/y  
1839/z



- Complex object A has parts x, y and z.
- Wasteful to issue a pid for each part (think of 100k entries in a lexicon)
- Need support for part identifier

1839/A: + 1839/A#x, 1839/A#y, 1839/A#z

