A DSpace-based Preservation Repository Design

DLF Fall Forum 2006

Joseph G. Pawletko,
Ekaterina Pechekhonova
New York University Libraries

Presentation Overview:

- Motivation
- Architectural Questions
- Our Approach
  - Background: OAIS Reference Model
  - Implementation details
- Current Status and Future Development
- Q & A

Motivation:

- NYU's Digital Library Program is working on several digital preservation grants
  - Hemispheric Institute Digital Video Library (**HIDVL**)
  - Afghanistan Digital Library (**ADL**)
  - NDIIPP: Preserving Digital Public Television (**PTV**)
- Grants have similar objectives
  - preserve and provide access to content

Architectural Questions:

- Build separate, grant-specific applications or a single Preservation Repository (PR) to fulfill grant requirements?
- Build monolithic or distributed applications?

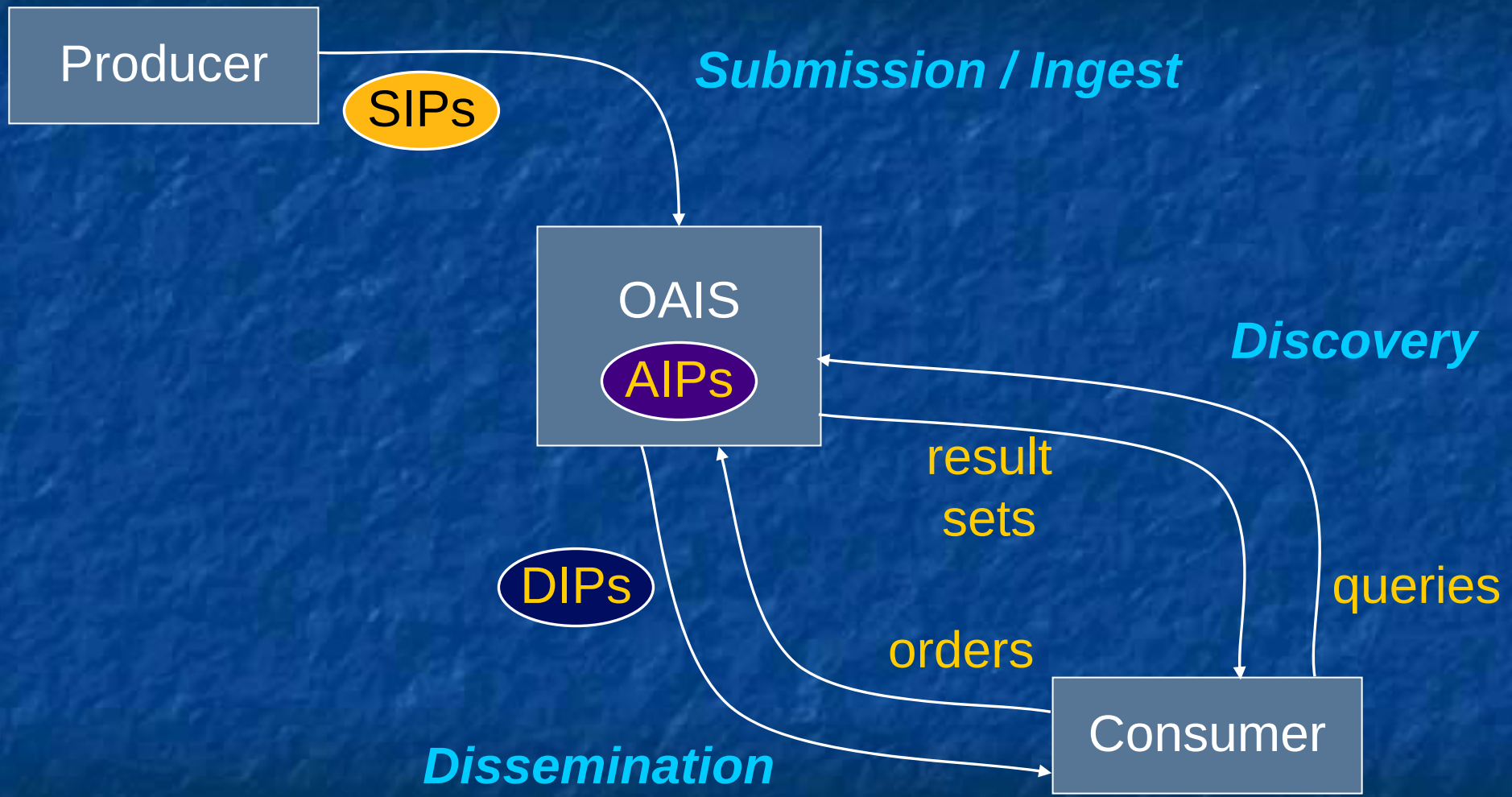
Our Decision:

- Build a single **Preservation Repository**
  - distributed architecture
  - loosely coupled components (stable interfaces)
- Each component provides subset of PR functionality
  - build **project-independent components** if possible
  - build **project-specific components** when required

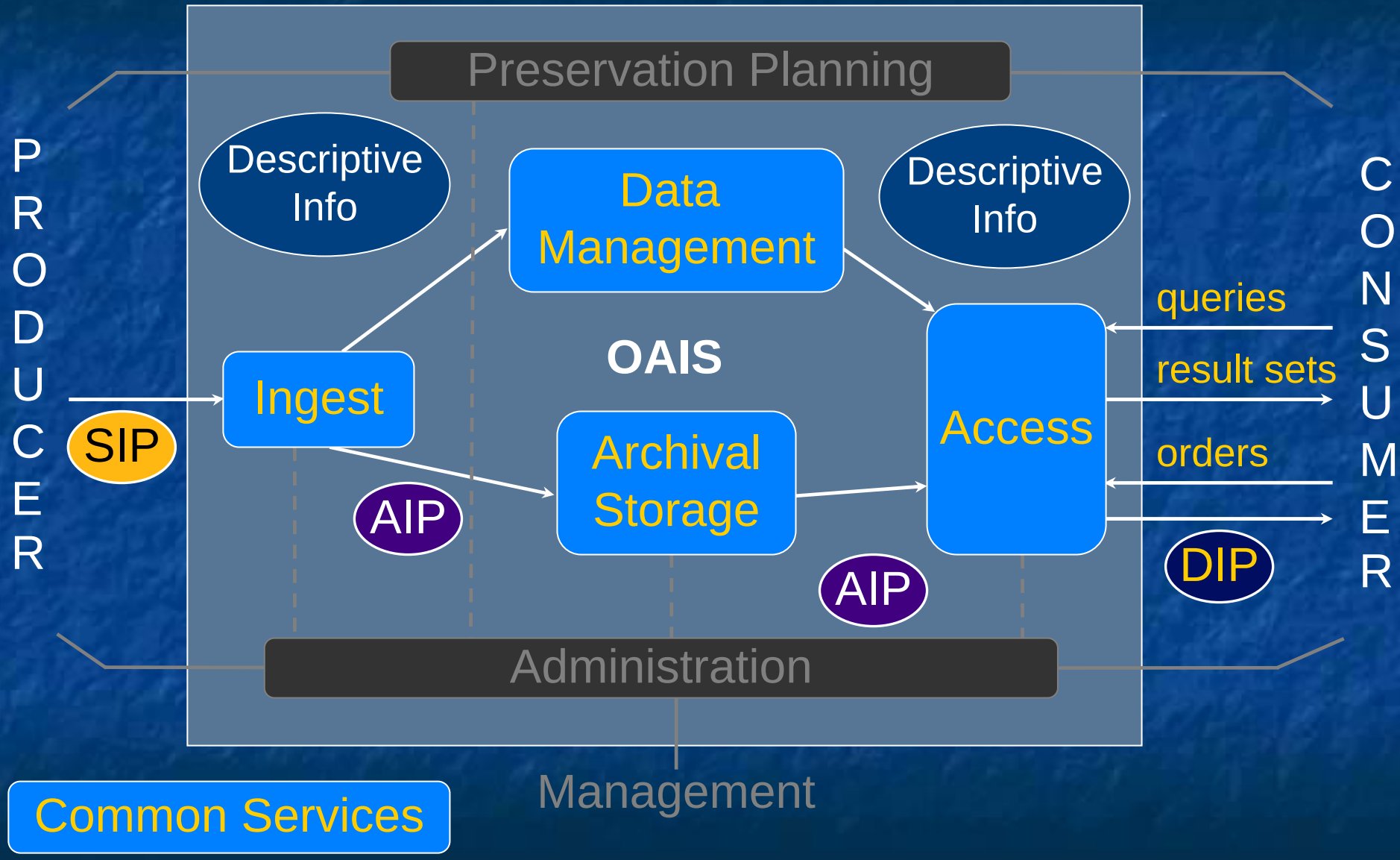
Why Components?

- Technology is changing rapidly…
- Using components…
  - provides upgrade flexibility
  - improves software development return
  - decreases wasted development effort

OAIS Archive External Data [1]



OAIS Functional Entities [2]



NYU-DL PR Implementation

Ingest
Project-Specific

Common Services
Project-Independent

Data Management
DSpace Features

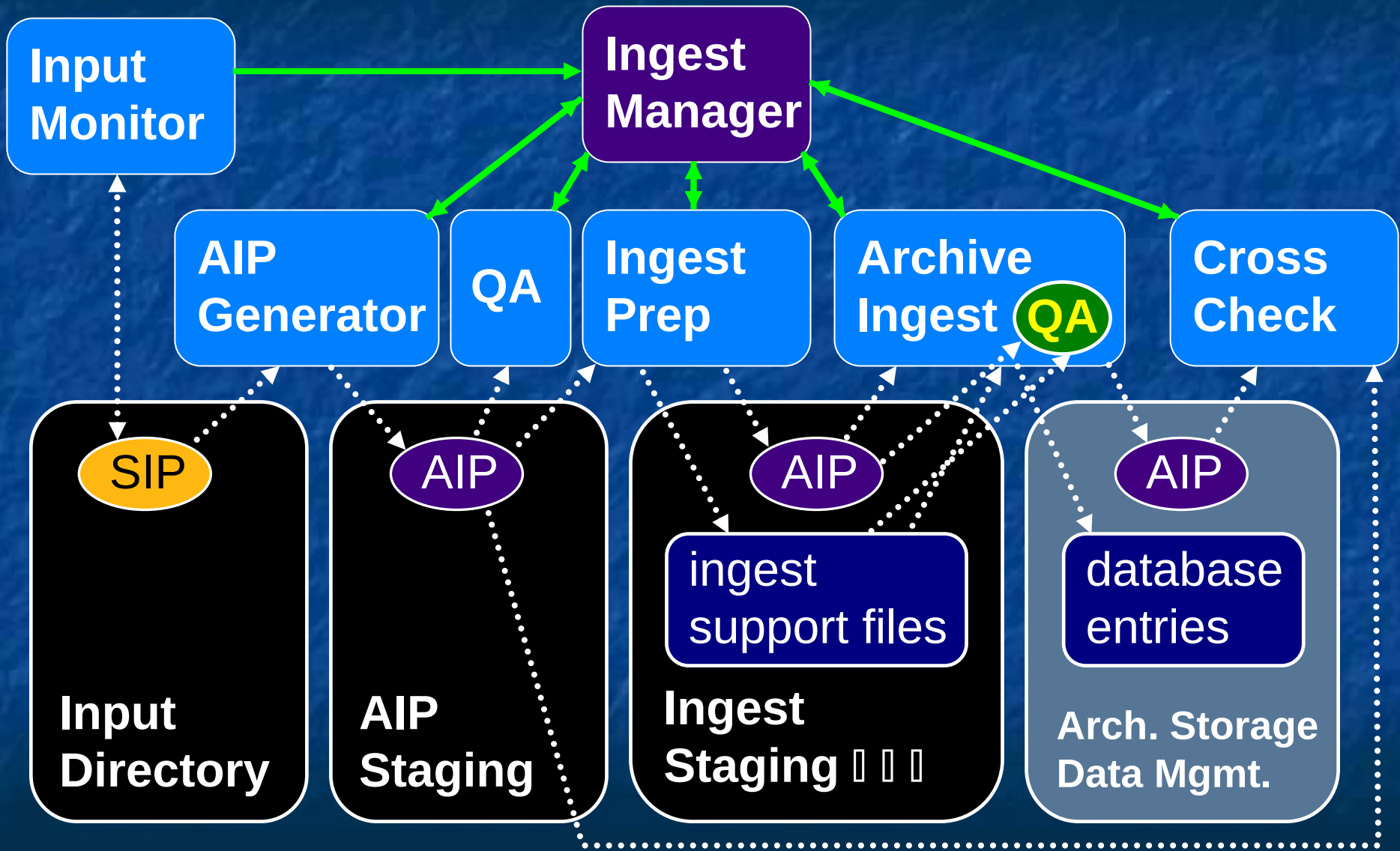
DSpace

Archival Storage
DSpace Item Importer
DSpace API
Maintenance API

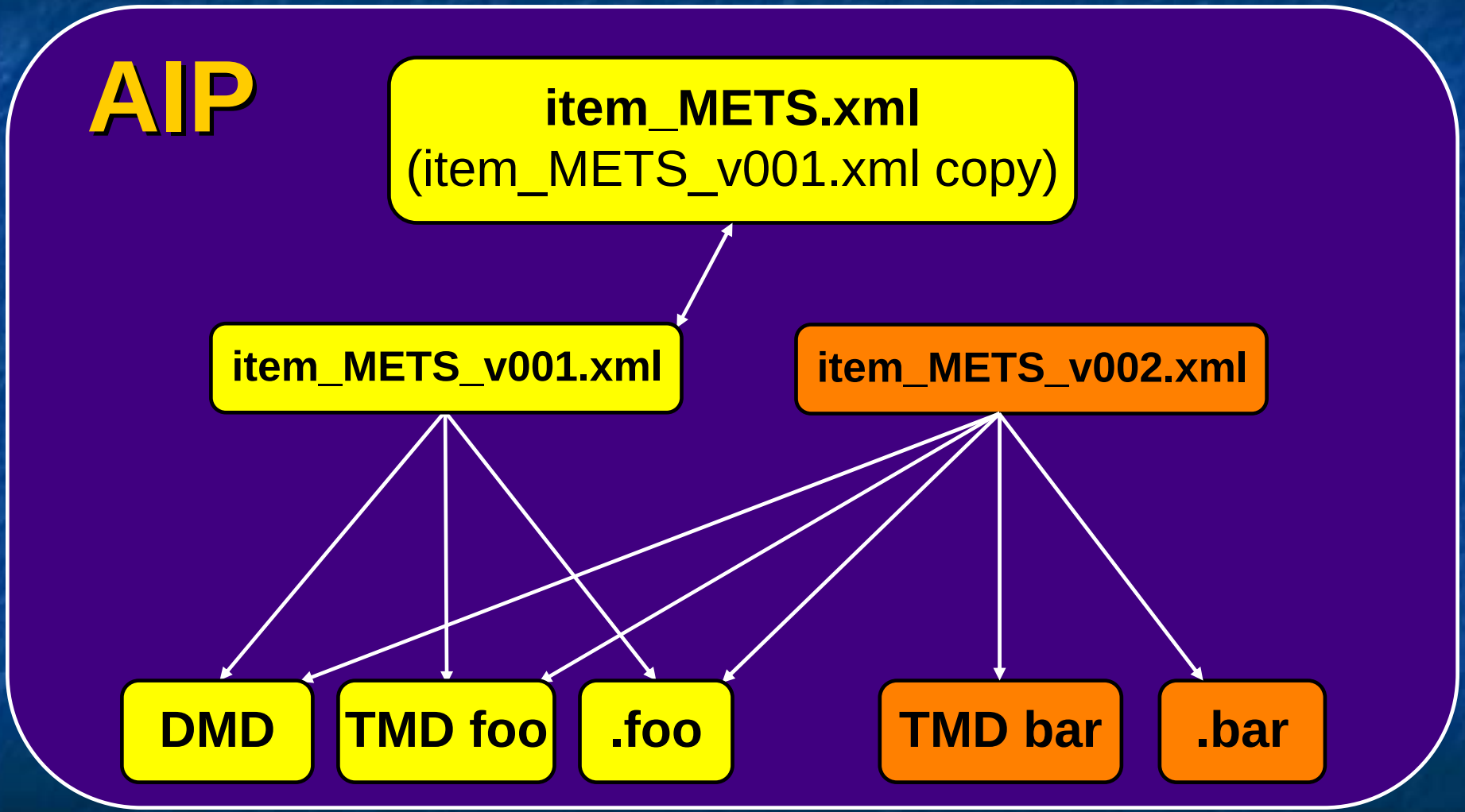
Access: *Discovery*
OAI-PMH
SRW
SRU

Access: *Dissemination*
Project-Specific

Ingest, Archival Storage, Data Management



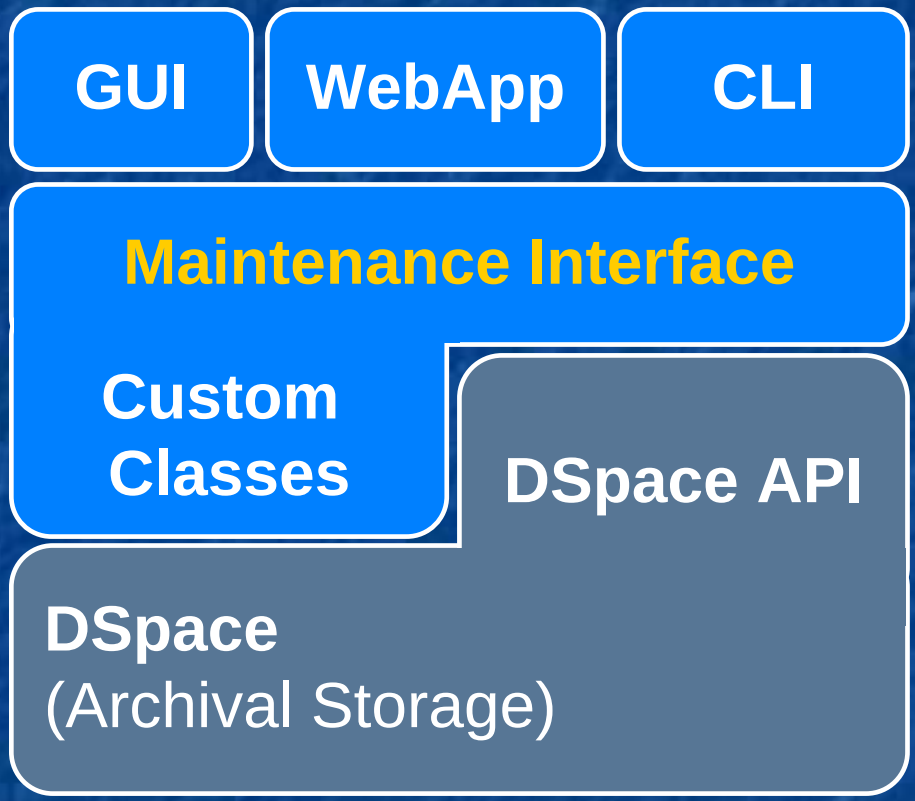
AIP Structure / Update / Rollback



Maintenance

- **Scope**:
  - enable **authorized users** to manage repository data and metadata
  - allow update operations to be **rolled back**
- Leverage existing components
  - DSpace APIs
  - Authentication Component
  - Authorization Component

Maintenance API



The **Maintenance Interface:**

- provides an archival-storage/data-management independent Java interface for PR Maintenance Operations
- can be invoked by a GUI, a WebApp, or through the **C**ommand **L**ine **I**nterface

NYU-DL PR Common Services

Common Services

Authentication
Project-Independent Scripts

Authorization
Project-Independent Scripts

Persistent Identifier Management
Project-Independent Scripts

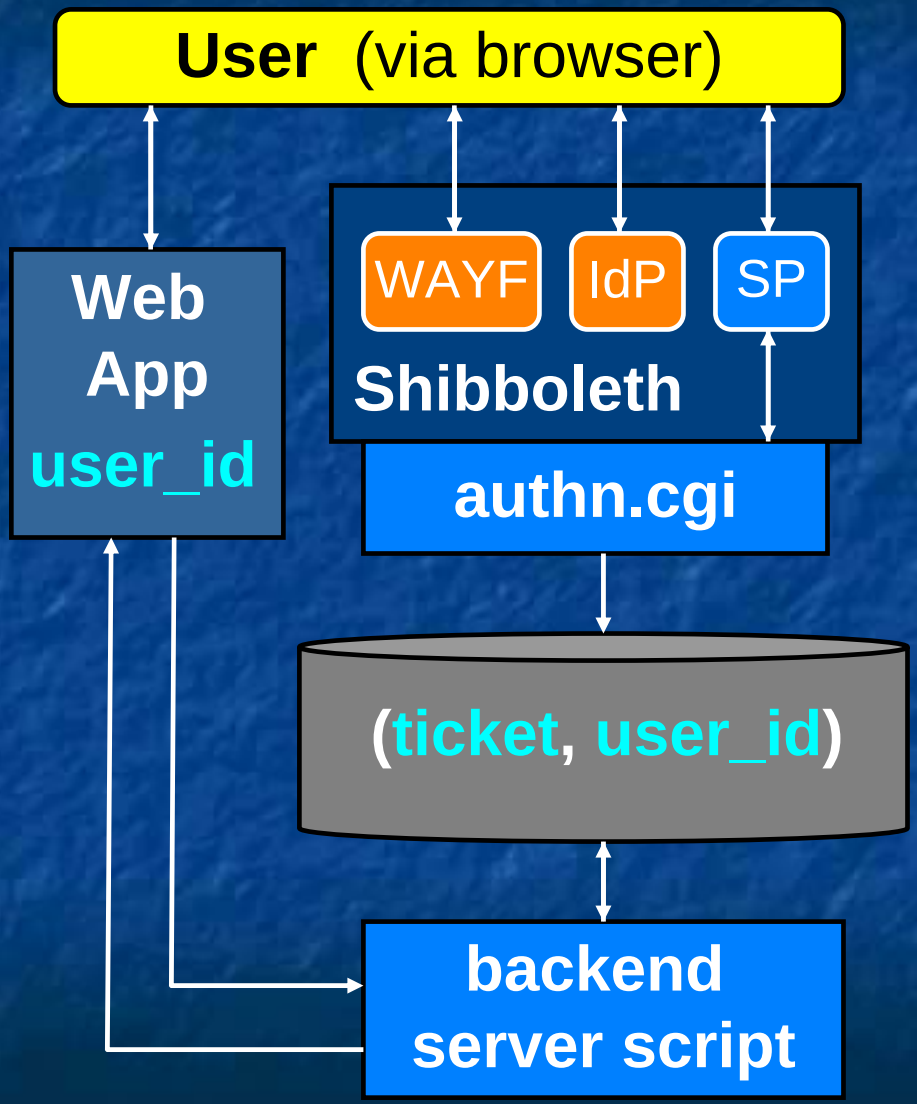
Persistent Identifier Resolution
CNRI Handle System® *

* HANDLE SYSTEM is a registered trademark of the Corporation for National Research Initiatives

Authentication Component

- Want to centralize Authentication functionality
  - implement once for all web applications, not once for each web application
- Built a **Shibboleth**-based Centralized Authentication Component
  - based on Yale Central Authentication Service[3] (CAS) design
  - leverages Database of Recorded American Music (DRAM) Shibboleth work at NYU

Authentication Component

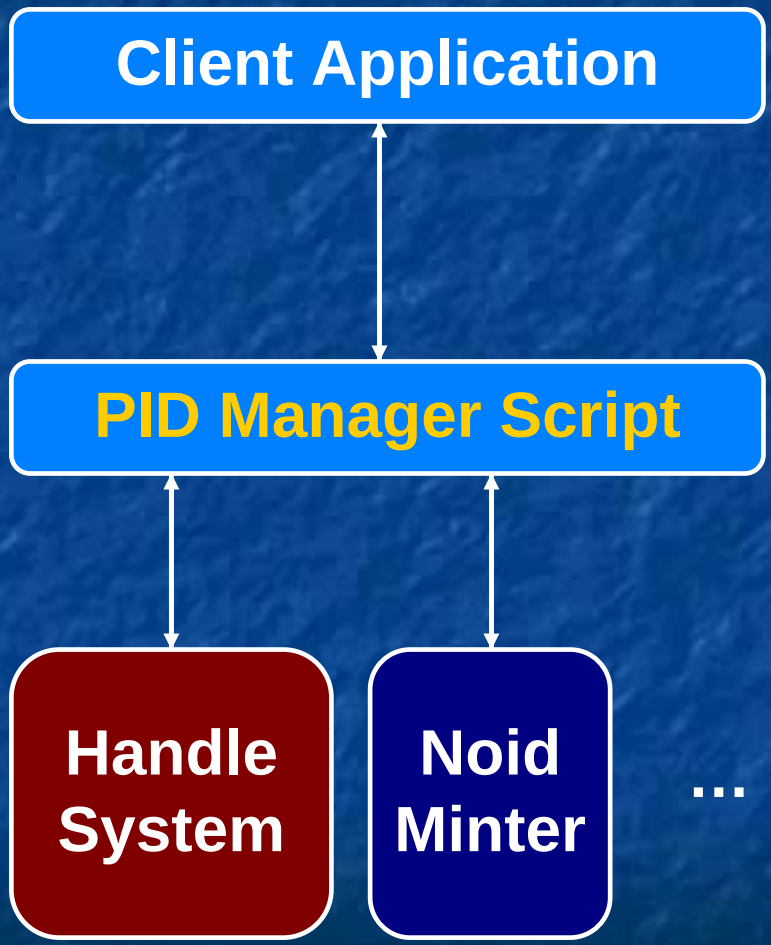


- Authentication Interaction:
  - Web App redirects to Shibboleth-protected script
  - User authenticates via Shibboleth infrastructure
  - authn script
    - extracts user_id
    - generates ticket
    - stores (ticket, user_id) pair
    - returns ticket to Web App
  - Web App exchanges ticket for user_id via XML-RPC
    - (ticket, user_id) pair deleted
  - web app uses authenticated user_id in application

Authorization Component

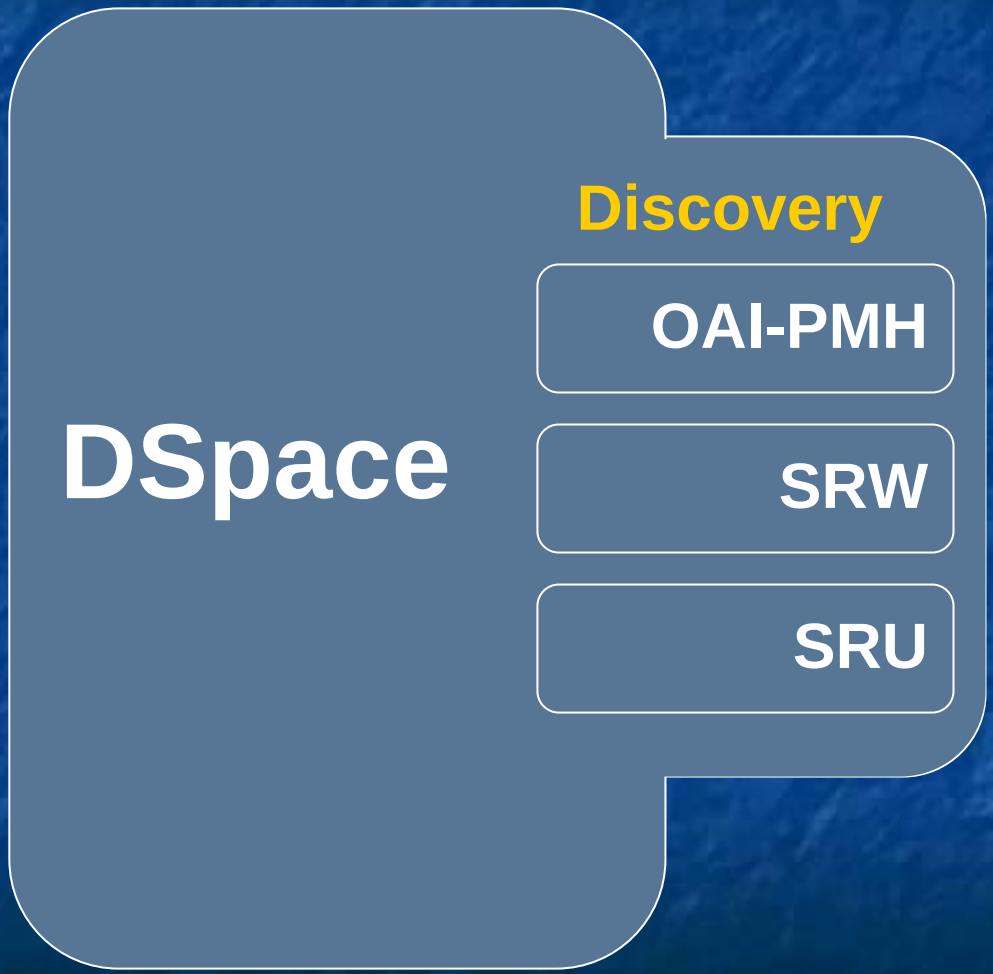
- Built centralized Authorization service
- Interaction:
  - client requests decision from Authorization via XML-RPC
  - Authorization component determines:
    - User's Role in Collection
    - Permitted Actions for Role
  - Authorization returns verdict

Persistent Identifier (PID) Manager



- Allocates and manages persistent identifiers
  - Handles, Noids, etc.
- Interaction:
  - Client requests an identifier operation via XML-RPC
    - e.g., create, update, etc.
  - Persistent Identifier Manager:
    - interacts with identifier-specific applications to satisfy request
    - returns status, data to Client

Access: Discovery

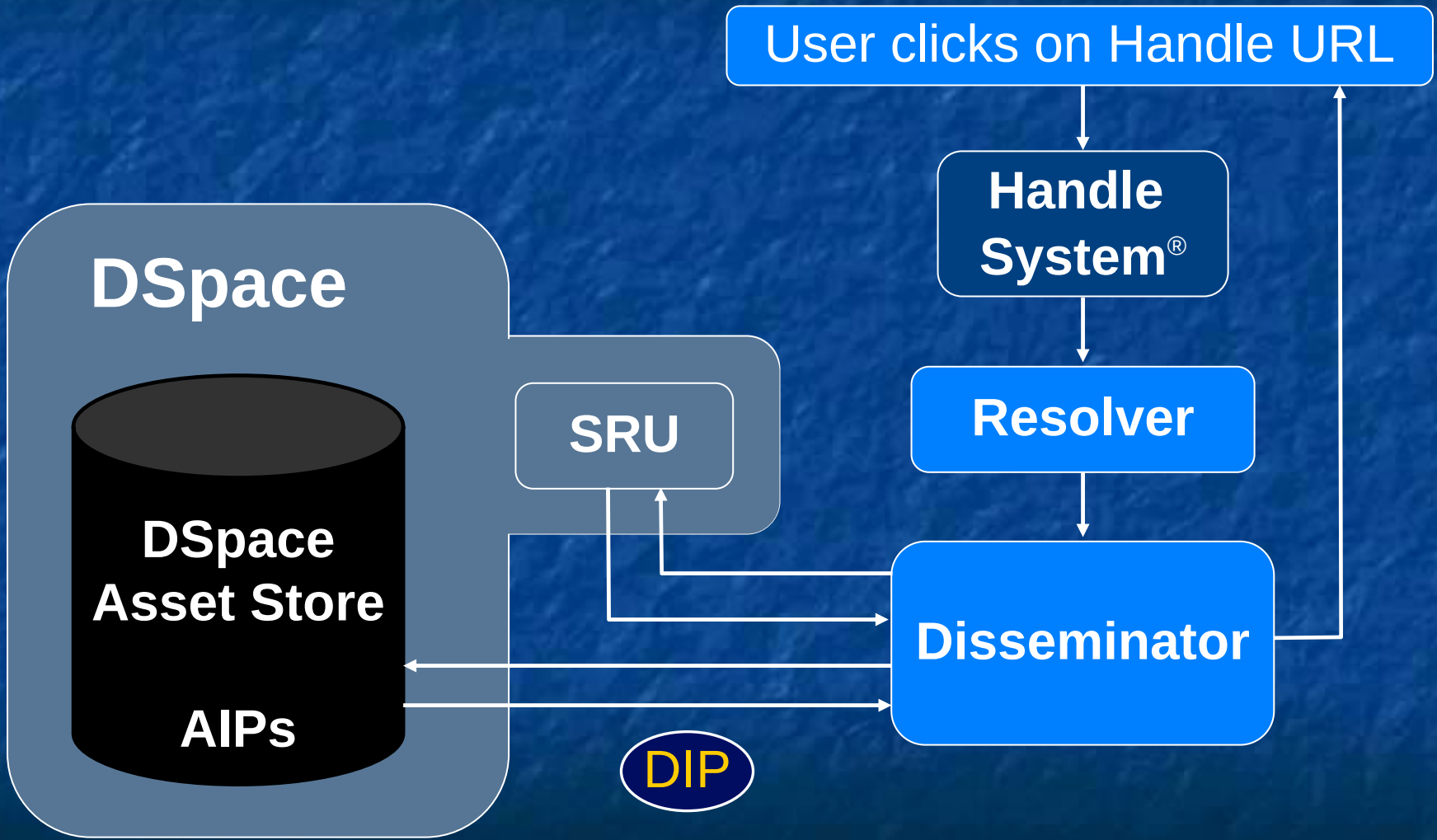


- **OAI-PMH** in standard DSpace distribution (OAICat from **OCLC**)
- **SRW/U:** DSpace add-on module available from **OCLC** (Ralph LeVan at **OCLC**)

Access: SRW/U Extension

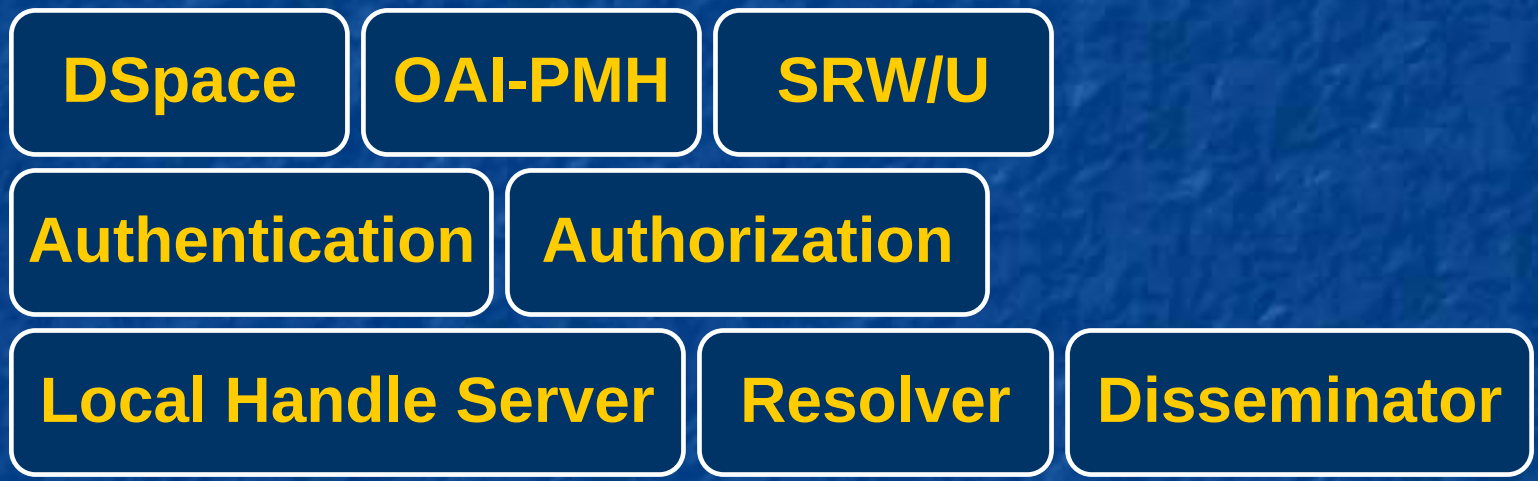
- Extended **SRW/U** to return real-time mapping using SRW/U protocol's "extra data" fields
  original filename → bitstream location
- **Disseminators** use mapping to access METS files, extract metadata, and build links to content
- Real-time map eliminates need to update AIP METS files when changing filesystems

Access: Dissemination



PR Status

- Built Proof-of-Concept PR from Core Components



- Tested:
  - DSpace Item Importer, SRU Discovery, Handle Resolution, Resolver Functionality, Dissemination, DSpace interaction with External Local Handle Server

PR Status & Future Plans

- Currently building remaining functionality

Ingest Scripts | **Maintenance API** | **PID Manager**

- Upgrading components for **Production**

Authorization | **Authentication** | **SRW/U** | **Resolver**

Disseminator

- Plan to load pilot data into the full system 1H'07

Acknowledgements:

- Andrew W. Mellon Foundation
- National Endowment for the Humanities
- Library of Congress

- Dean Carol Mandel
  - Dean of New York University Libraries
- Dr. David Ackerman
  - Executive Director .edu Services and Digital Library Program, NYU
- James Bullen
  - Head of the Digital Library Program, NYU

- Sayeed Choudhury, Tim DiLauro
  - Johns Hopkins University

Wrap up…

Questions?

Thank you for your time!

References

[1] Consultative Committee for Space Data Systems. 2002. *Reference Model for an Open Archival Information System*. Washington, D.C.: CCSDS Secretariat. Publication number CCSDS 650.0-B-1 BLUE BOOK. (PDF version of document downloaded 2006-01-09). page 2-8

[2] Ibid., page 4-1

[3] Yale University Technology and Planning. *ITS Central Authentication Service (version 1.0).* New Haven, CT: Yale University - Information Technology Services. http://www.yale.edu/tp/auth/cas10.html (accessed November 09, 2006).