

Preserving Digital Public Television

Part of the National Digital Information Infrastructure and
Preservation Program of the Library of Congress

Kara Van Malssen
Senior Research Fellow
New York University

Who We Are



**WNET/Thirteen
+ WGBH**

PBS

New York University

**Library of
Congress**

Thirteen/WNET and WGBH



Content and Production Expertise

- The two largest public television stations in the US
- Together produce the majority of national programs
- Both have preservation archives

PBS



Content and Network Design

- Distributes most of the national programming
- Creates and keeps broadcast masters of programs
- Developing file-based contribution and delivery system

New York University



Facilitation and Resources

- Leadership in designing digital libraries
- Experience in process for setting standards
- Expertise and assistance from MIAP faculty and students

Library of Congress



Funding, Leadership, and Networks

National Digital Information and Infrastructure Preservation Program (NDIIPP):

- Solutions for born digital content
- Most projects university-based
- Dealing with variety of topics: websites, geospatial data, etc.

Preserving Digital Public Television Project Goals

Identify at-risk **born digital public television** content

Explore appropriate file formats, wrappers, and metadata **standards**

Develop an **OAIS**-compliant repository

Lifecycle management (workflow automation, capture metadata at creation)

Create **selection guidelines** for completed programs and ancillary materials

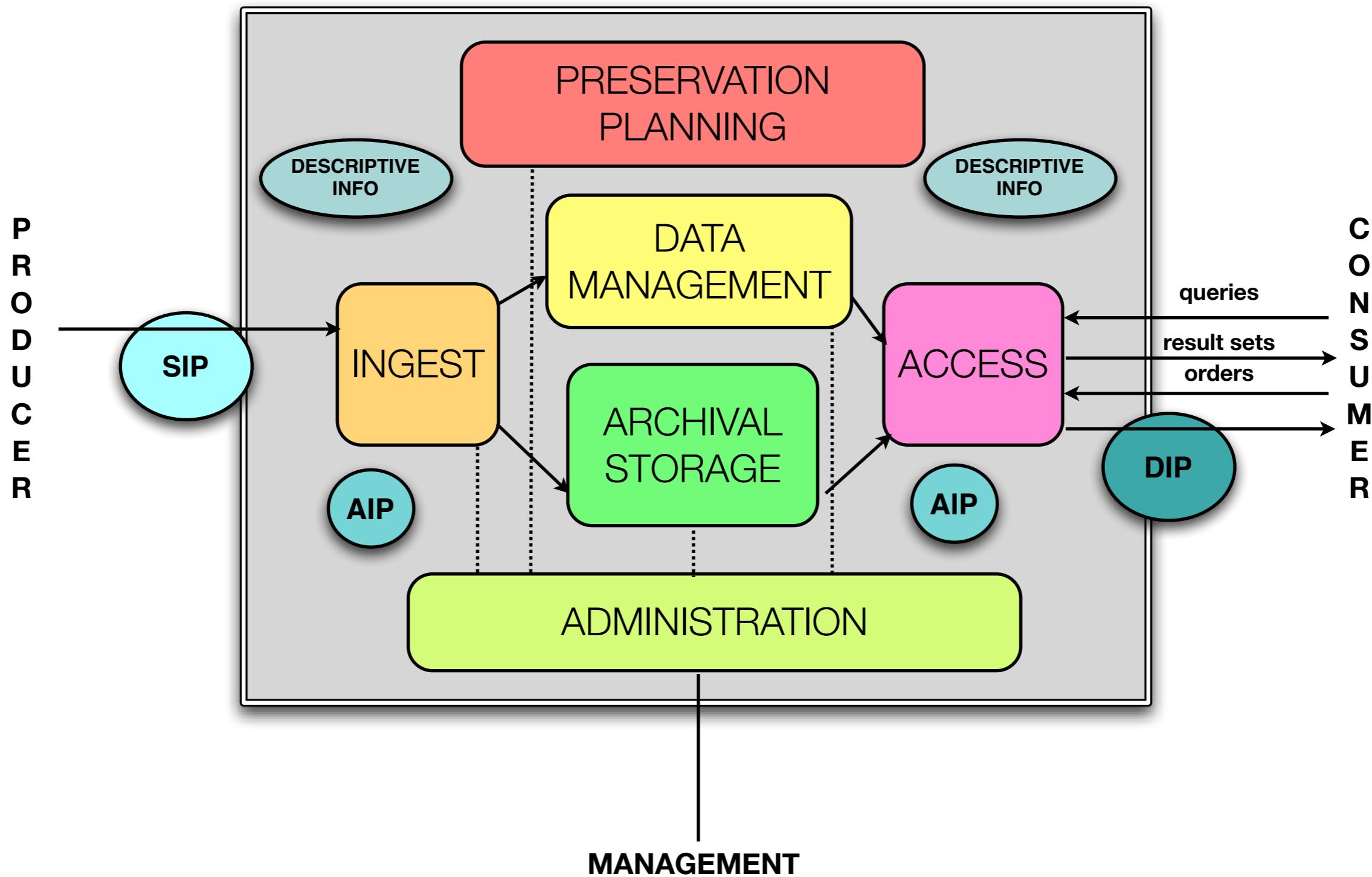
Develop a **sustainable** business model and identify **copyright** encumbrances

OAIS Reference Model

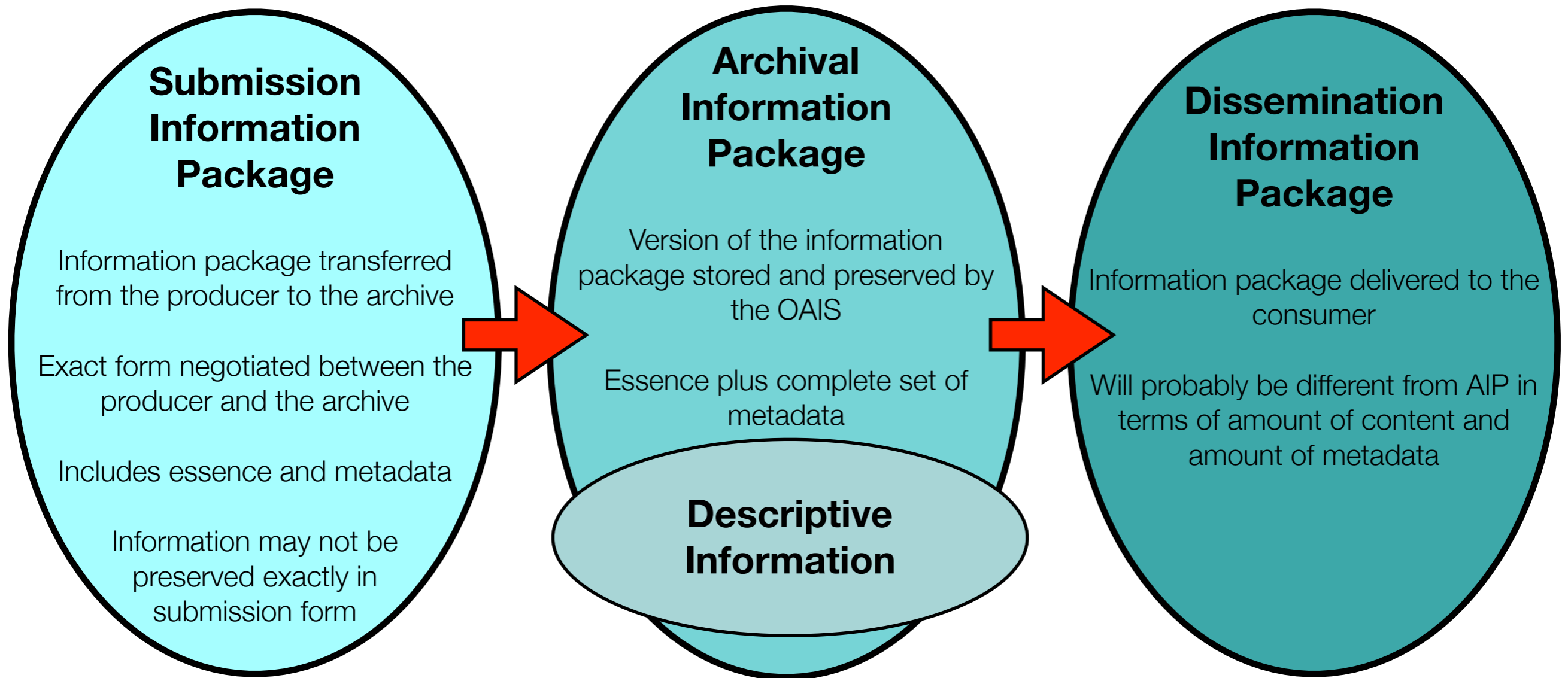
- Developed by the space science community, widely adopted in all digital preservation communities
- ISO standard
- Defines tasks and entities of a digital repository (but not how to implement them)
- Defines **Mandatory Responsibilities** 

Mandatory Responsibilities

- [**Negotiate** for and accept **appropriate information** from producers
- [Obtain sufficient **control** of the information to meet **long term preservation** objectives
- [Determine the **scope** of the archive's **user community**
- [Ensure that **preserved info** is **available** and **understandable** to user community
- [Follow **documented policies** and **procedures** for preservation and to enable dissemination of preserved information



OAIS Functional Model



OAIS Information Model


NYU Goals


Create a prototype repository for long term retention

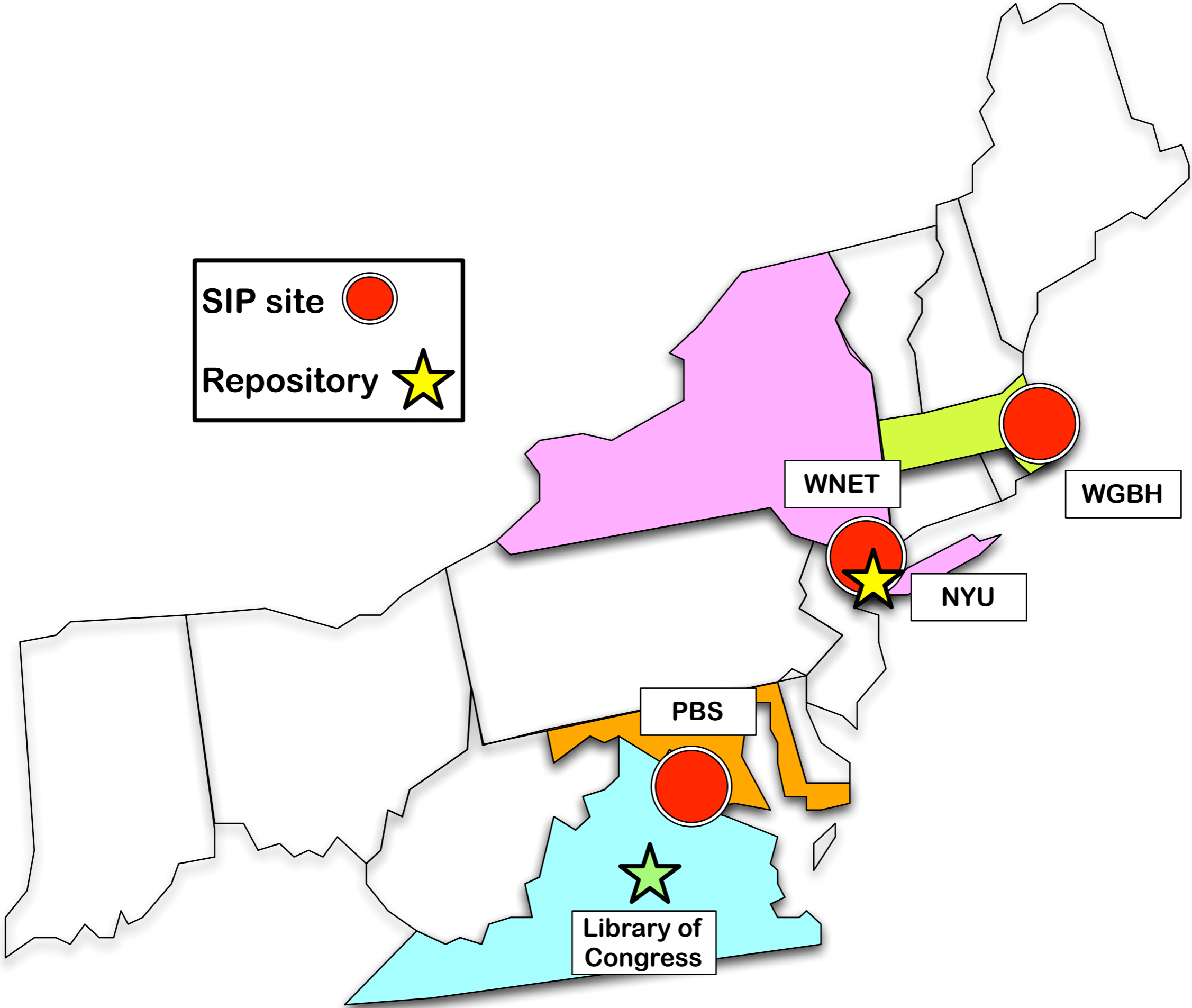
Aggregate content from partner stations + PBS for
sample programs

Use metadata that already exists (station archive databases,
DAMs, PBS database, traffic systems, file technical metadata)

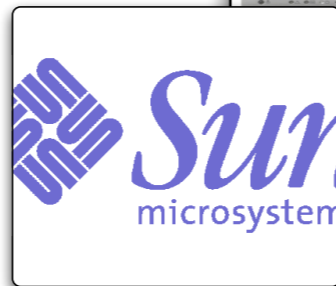
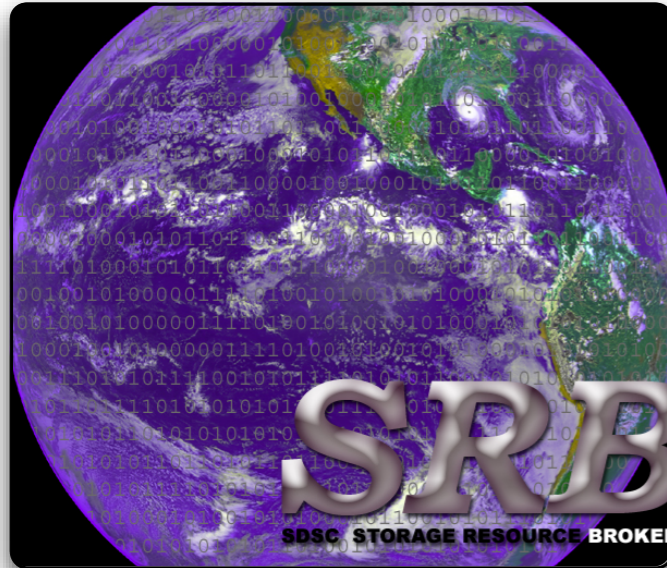
Transform data and package content, while
preserving relationships between items

SIP site 

Repository 



Repository Architecture (current implementation)



Storage on three dedicated "thumpers" (Sun X4500 Servers), 25 TB capacity each (RAID-Z)



Running on Solaris 10

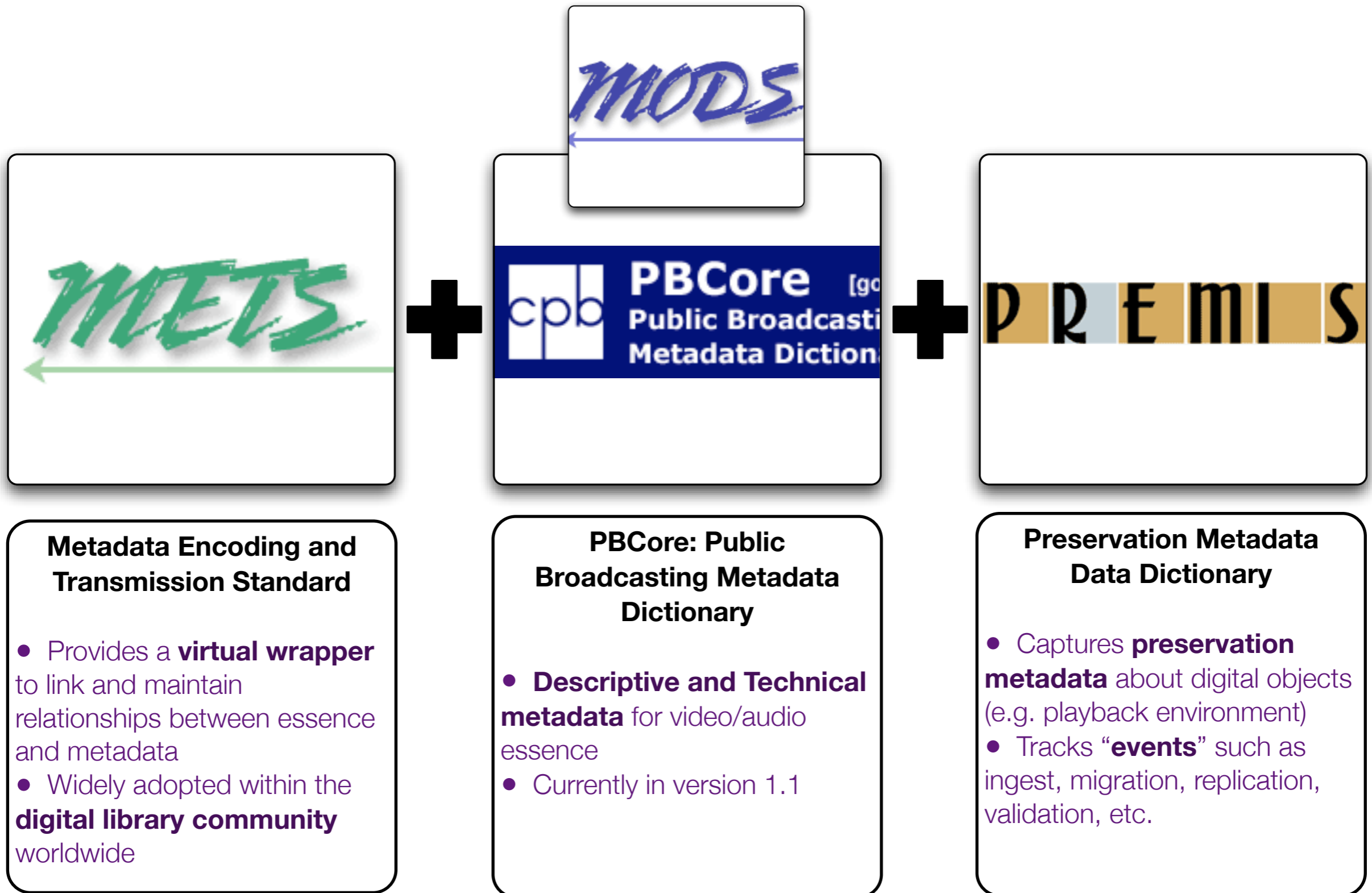
Storage Resource Broker

- Developed by San Diego Super Computer Center
- **Distributed system architecture**
- Supports shared collections that can be distributed across multiple organizations and heterogenous storage systems.



SRB MCAT database running on Oracle 10g instance

Metadata Model



PBCore: Classes

www.pbcore.org

Intellectual Content

Identifier, Title/Type, Subject, Genre, Description, Audience Rating...

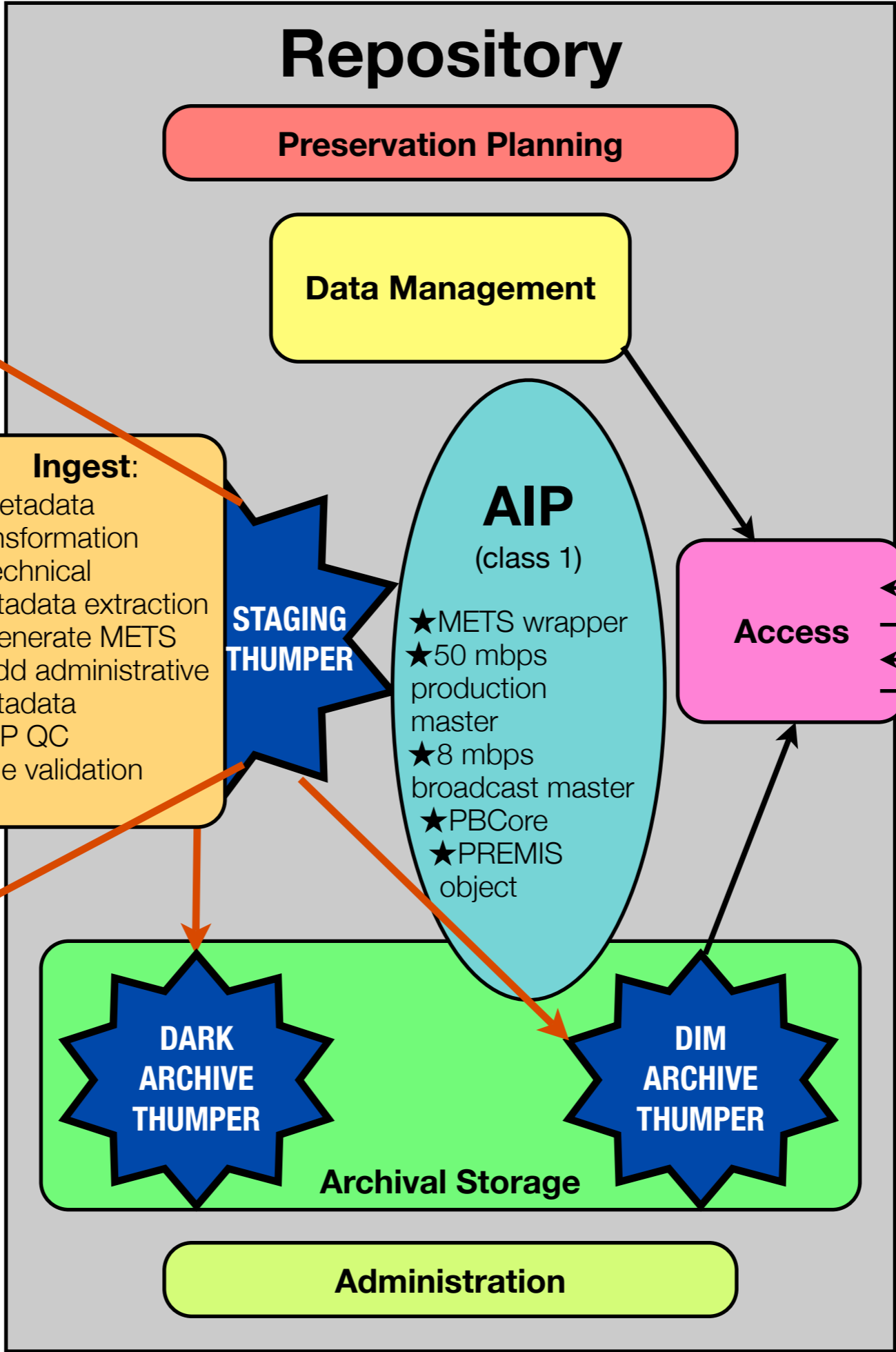
Intellectual Property

Creator/Role, Contributor/Role, Publisher/Role, Rights Summary

Instantiation

Format, Location, Encoding, File Size, Duration, Data Rate, Aspect Ratio, Colors, Tracks, Language, Alternative Modes, Dates...

Sample Workflow



WNET/WGBH:

- **production master** (MXF wrapped MPEG-2 50 mbps I-Frame only or QT wrapped DVC PRO 50)
- **database export** (XML from InMagic database or Artesia TEAMS DAM)
- **bonus materials** (transcripts, closed captioning, production credits)

PBS:

- **broadcast master** (QT wrapped MPEG-2 8 mbps)
- **PBCore export** (Based on PODS data)

Ingest:

- Metadata transformation
- Technical metadata extraction
- Generate METS
- Add administrative metadata
- SIP QC
- File validation

STAGING THUMPER

AIP (class 1)

- ★ METS wrapper
- ★ 50 mbps production master
- ★ 8 mbps broadcast master
- ★ PBCore
- ★ PREMIS object

Access

DARK ARCHIVE THUMPER

DIM ARCHIVE THUMPER

Archival Storage

Administration

Preservation Planning

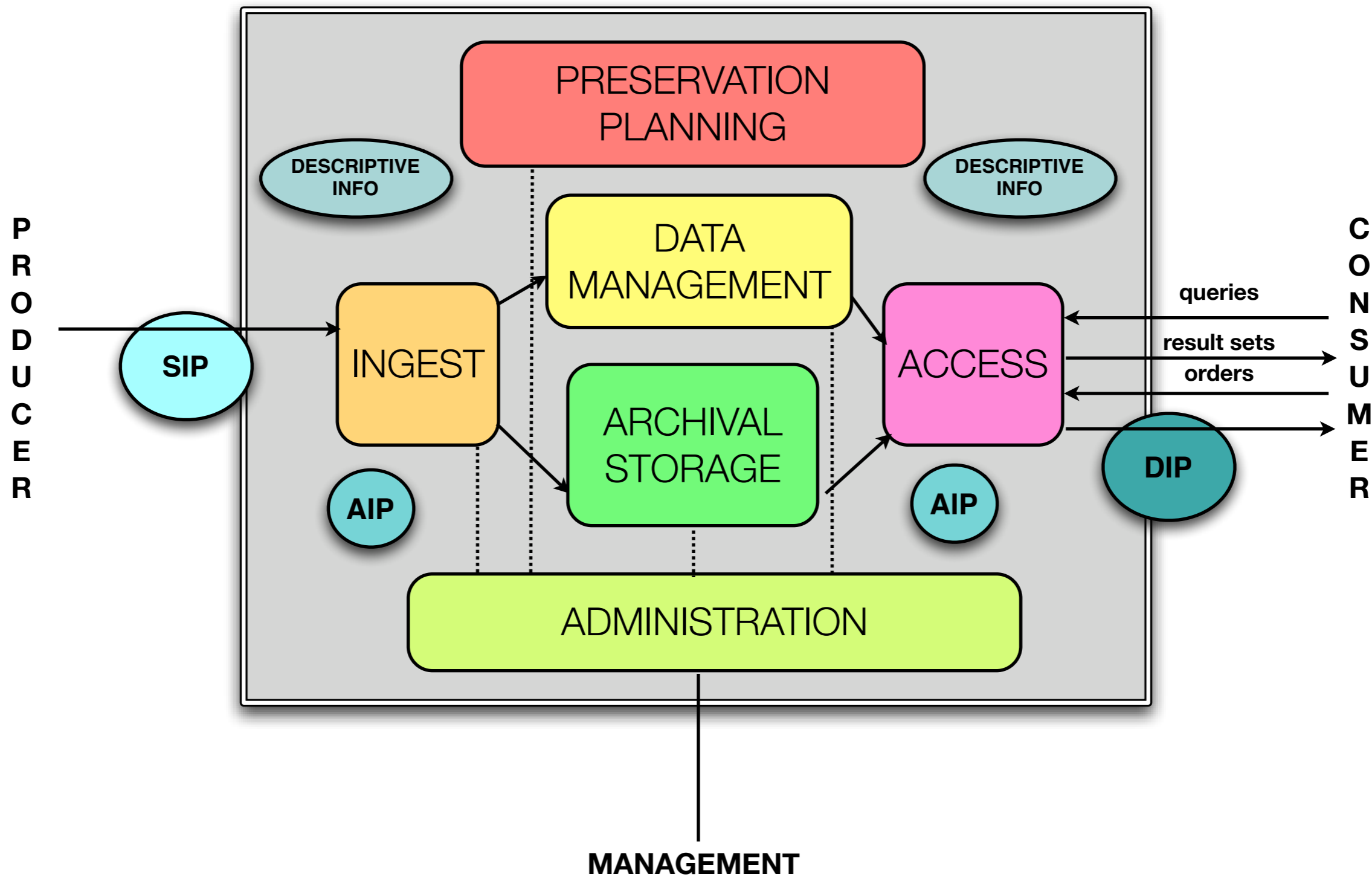
Data Management

DIP

- ★ Metadata only
- ★ Production Master + metadata only

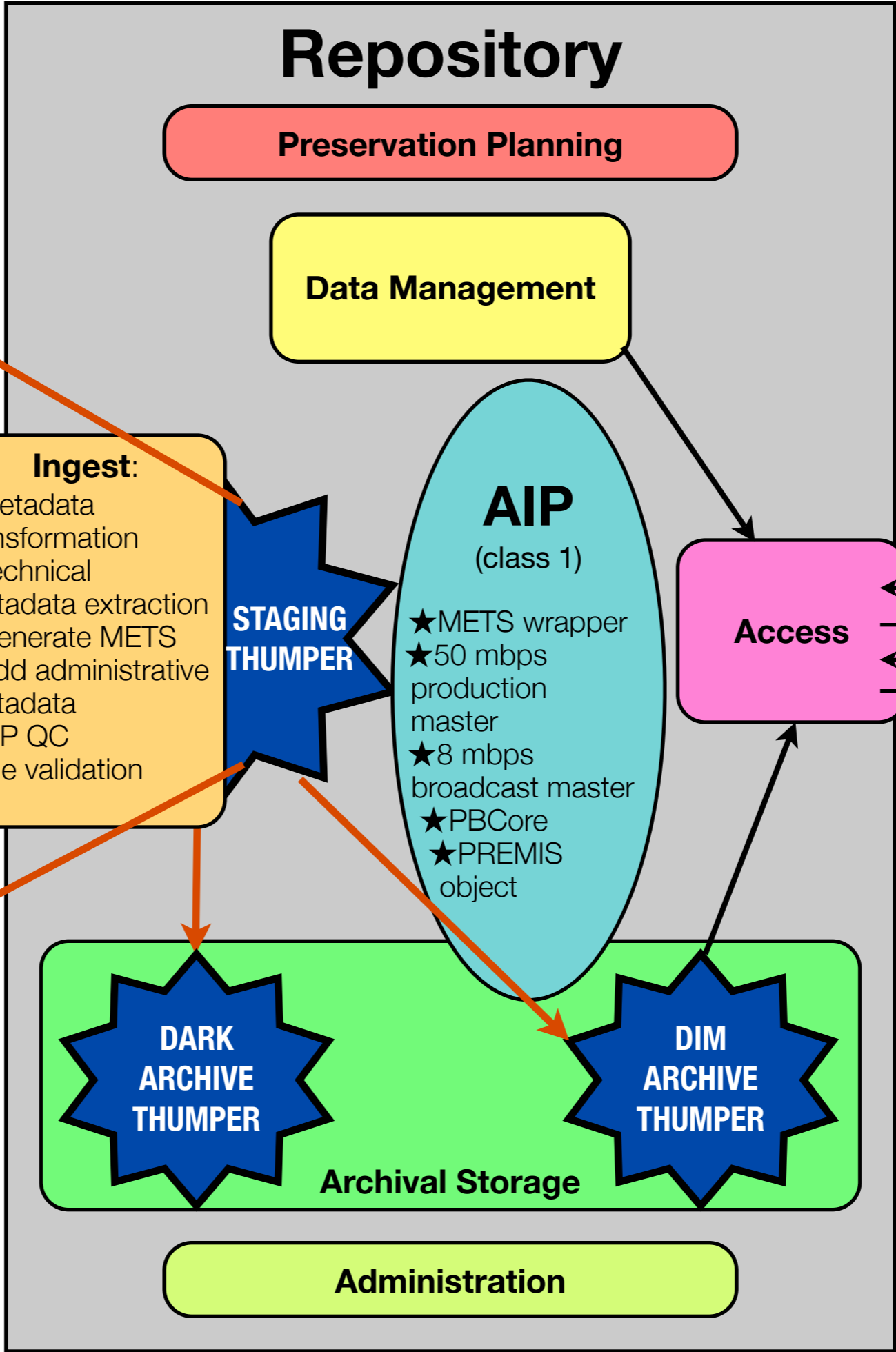
**WGBH
WNET
PBS**

- ★ Broadcast Master + metadata only
- ★ Entire package (without administrative metadata)



OAIS Functional Model

Sample Workflow



WNET/WGBH:

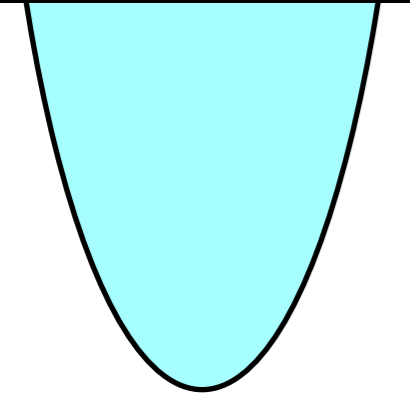
- **production master** (MXF wrapped MPEG-2 50 mbps I-Frame only or QT wrapped DVC PRO 50)
- **database export** (XML from InMagic database or Artesia TEAMS DAM)
- **bonus materials** (transcripts, closed captioning, production credits)



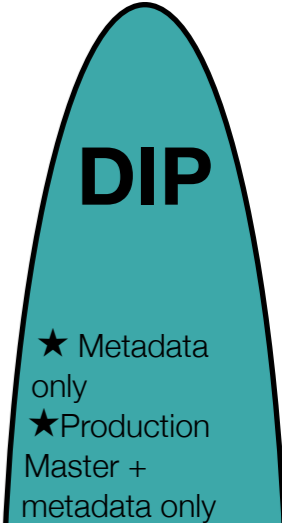
SIP

PBS:

- **broadcast master** (QT wrapped MPEG-2 8 mbps)
- **PBCore export** (Based on PODS data)



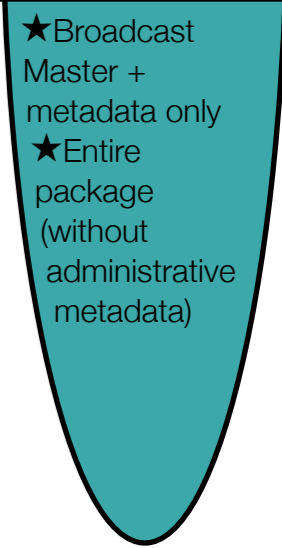
PBS:



DIP

- ★ Metadata only
- ★ Production Master + metadata only

**WGBH
WNET
PBS**



- ★ Broadcast Master + metadata only
- ★ Entire package (without administrative metadata)

Preservation Planning

Data Management

Ingest:

- Metadata transformation
- Technical metadata extraction
- Generate METS
- Add administrative metadata
- SIP QC
- File validation

STAGING THUMPER

AIP (class 1)

- ★ METS wrapper
- ★ 50 mbps production master
- ★ 8 mbps broadcast master
- ★ PBCore
- ★ PREMIS object

Access

Archival Storage

DARK ARCHIVE THUMPER

DIM ARCHIVE THUMPER

Administration

WNET/WGBH:

- **production master** (MXF wrapped MPEG-2 50 mbps I-Frame only or QT wrapped DVC PRO 50)
- **database export** (XML from InMagic database or Artesia TEAMS DAM)
- **bonus materials** (transcripts, closed captioning, production credits)

Data

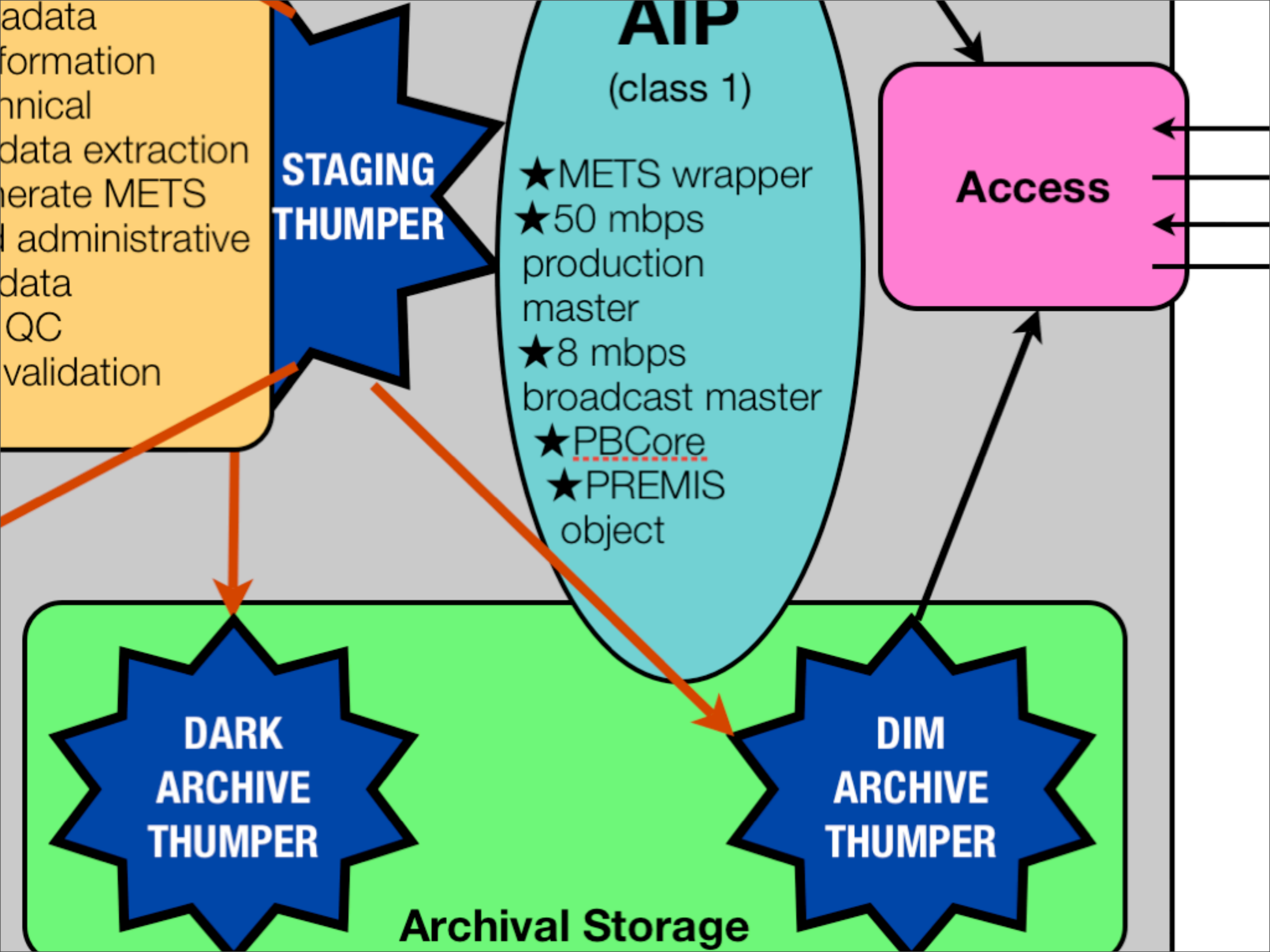
Ingest:

- Metadata transformation
- Technical metadata extraction
- Generate METS
- Add administrative metadata
- SIP QC
- File validation

STAGING THUMPER

PBS:

- **broadcast master** (QT wrapped MPEG-2 8 mbps)
- **PBCore export** (Based on PODS data)



Management

AIP
(class 1)

- ★ METS wrapper
- ★ 50 mbps production master
- ★ 8 mbps broadcast master
- ★ PBCore
- ★ PREMIS object

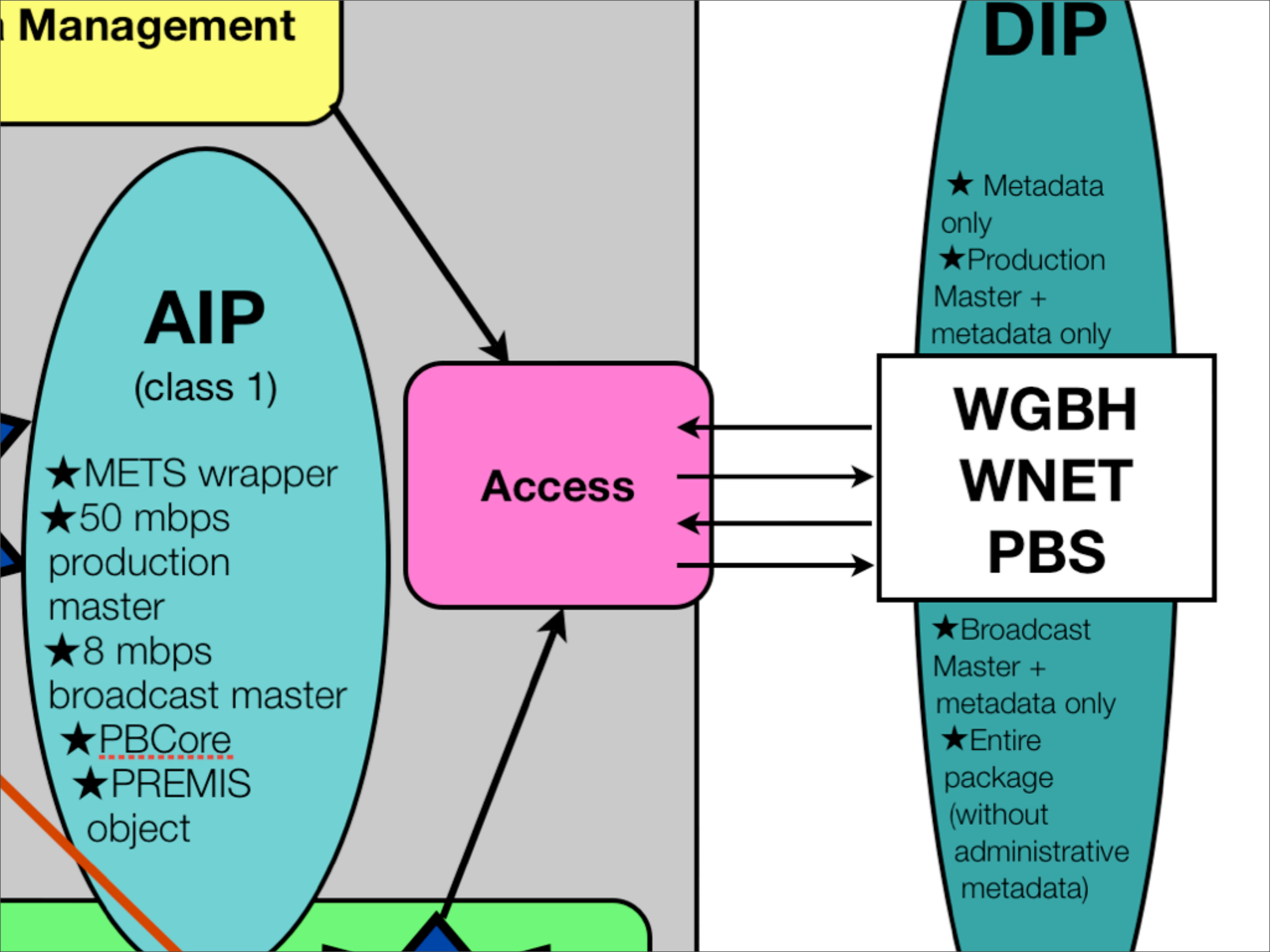
Access

DIP

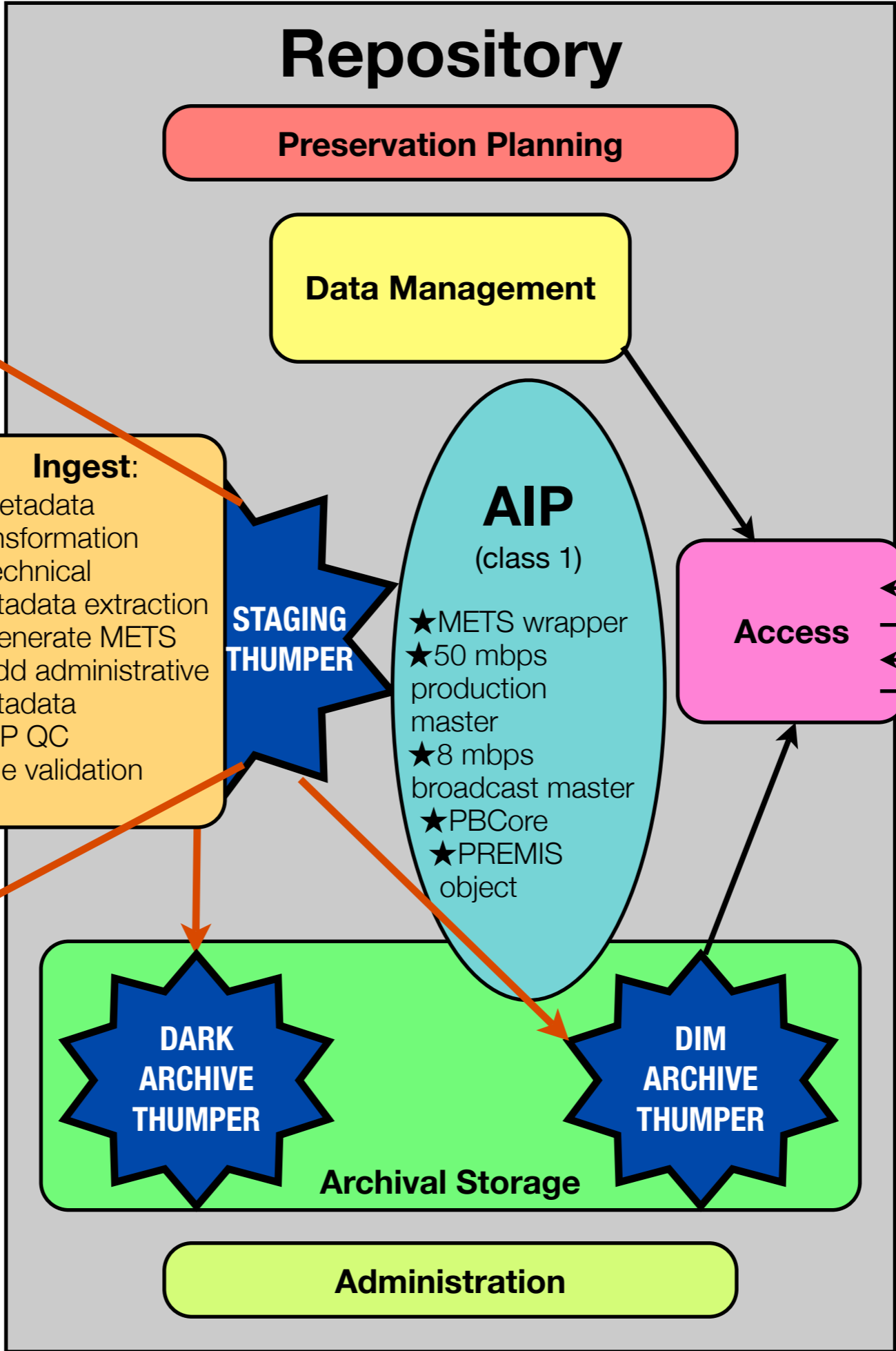
- ★ Metadata only
- ★ Production Master + metadata only

WGBH
WNET
PBS

- ★ Broadcast Master + metadata only
- ★ Entire package (without administrative metadata)



Sample Workflow



WNET/WGBH:

- **production master** (MXF wrapped MPEG-2 50 mbps I-Frame only or QT wrapped DVC PRO 50)
- **database export** (XML from InMagic database or Artesia TEAMS DAM)
- **bonus materials** (transcripts, closed captioning, production credits)

PBS:

- **broadcast master** (QT wrapped MPEG-2 8 mbps)
- **PBCore export** (Based on PODS data)

**WGBH
WNET
PBS**

DIP

- ★ Metadata only
- ★ Production Master + metadata only

- ★ Broadcast Master + metadata only
- ★ Entire package (without administrative metadata)

Ingest:

- Metadata transformation
- Technical metadata extraction
- Generate METS
- Add administrative metadata
- SIP QC
- File validation

**AIP
(class 1)**

- ★ METS wrapper
- ★ 50 mbps production master
- ★ 8 mbps broadcast master
- ★ PBCore
- ★ PREMIS object

Archival Storage

DARK ARCHIVE THUMPER **DIM ARCHIVE THUMPER**

Administration

Preservation Planning

Data Management

Access

SIP

Challenges

- **Variety of material** coming from **different locations** at **different times**
- Different file formats, databases
- **Selection** decisions 
- Remaining **scalable** (increased content, local and independent programming)

Episode Packages								
Package #	Version	Package Type	Duration	Audio	SAP	CC	DVI	Aspect
P153584-005	NAAT 002204	SD-Base	00:56:46	STEREO	No	Yes	Yes	Letterbox
P153585-005	NAAT 002204	HD-Base	00:56:46	SRND51	No	Yes	No	16:9
P153584-004	NAAT 002204	SD-Base	00:56:46	STEREO	No	Yes	Yes	Letterbox
P153585-004	NAAT 002204	HD-Base	00:56:46	SRND51	No	Yes	No	16:9
P196120-003	NAAT 002204	SD-Embedded Promo	00:56:46	STEREO	No	Yes	Yes	Letterbox
P196117-001	NAAT 002204B1	SD-Base	00:56:46	STEREO	No	No	Yes	4:3
P196117-002	NAAT 002204B2	SD-Base	00:56:46	STEREO	No	No	No	4:3
P153584-001	NAAT 002204C1	SD-Base	00:56:46	STEREO	No	Yes	Yes	4:3
P153584-002	NAAT 002204C2	SD-Base	00:56:46	STEREO	No	Yes	Yes	Letterbox
P153585-001	NAAT 002204H1	HD-Base	00:56:46	STEREO	No	Yes	No	16:9
P153585-002	NAAT 002204H2	HD-Base	00:56:46	STEREO	No	Yes	No	16:9
P196118-001	NAAT 002204I1	HD-Base	00:56:46	STEREO	No	No	Yes	16:9
P196118-002	NAAT 002204I2	HD-Base	00:56:46	STEREO	No	No	No	16:9
P196119-001	NAAT 002204O1	SD-Stacked	00:56:46	STEREO	No	No	No	4:3
P196120-001	NAAT 002204W1	SD-Embedded Promo	00:56:46	STEREO	No	Yes	Yes	Letterbox
P153584-003	NAAT 002204	SD-Base	00:56:46	STEREO	No	Yes	Yes	Letterbox
P153585-003	NAAT 002204	HD-Base	00:56:46	SRND51	No	Yes	Yes	16:9
P196120-002	NAAT 002204	SD-Embedded Promo	00:56:46	STEREO	No	Yes	Yes	Letterbox

Selection decisions

Shift in approach

VISION	CURRENTLY IMPLEMENTABLE
PBCore as part of SIP	Transform to PBCore in staging
All files MXF with metadata embedded	Variety of file formats, separate metadata
Ancillary production elements	Only completed programs
EDLs	Not useful without ancillary production elements
Promos for future researchers	No systematic way at this time
Disseminate segments	Full programs only
Closed captioning XML	Currently Unavailable

Success requires close collaboration with project partners

Thank you.

kara.vanmalssen@nyu.edu