

# Adding OAI-ORE Support to Repository Platforms

*Alexey Maslov, Adam Mikeal, Scott Phillips,  
John Leggett, Mark McFarland*

**Texas Digital Library**  
OR'09

# Overview

- *Texas Digital Library  
Use Case for OAI-ORE*
- *Mapping ORE model to  
DSpace architecture*
- *Implementation*
- *Results and  
Implications*







# Texas Digital Library

- State-wide initiative
- Eighteen members
  - Public/Private
  - Small/Medium/Large



# Electronic Theses and Dissertations

- Federated Collection
- Built on top of DSpace/Manakin

Previous Page	Now showing items 101-120 of 6733	Next Page
[+] Advances in cationic graft polymerization lithography Meiring, Heather Faye, Doctor of Philosophy (The University of Texas at Austin, 2008-08-28)		
[+] Advance the DNA computing Qiu, Zhiqun Frank, PHD (Texas A&M University, August 2003)		
[+] Advantages and disadvantages of microporous membranes in a hollow fiber bioreactor for space applications Ruiz Careri, Maria Noel, Master of Science in Civil Engineering (Texas Tech University, 2005-07-26)		
[+] The Advantages Of Implementing Software Engineering Process Models Preuninger, Ricky Don, M.S. (University of Texas at Arlington, April 2006)		
[+] Adventures in main group chemistry: from molecules to materials Findlater, Michael, Doctor of Philosophy (The University of Texas at Austin, 2008-08-29)		
[+] Adverse camber: a ballet for questionable ensemble		

# Current Federation Method

- Performed via scripted ingest process
- New batch every semester
- Manual corrections to existing content

# Replacement Requirements

- Perform maintenance automatically
- Detect changes in existing content
- Support interchange of metadata and content

# Harvesting Solution

- Use the Open Archives Initiative Protocol for Metadata Harvesting
- Member institutions as data providers
- TDL Federated Repository as a service provider

# OAI-PMH, advantages

- Ubiquitous
- Supports selective harvesting
- Tracks changes
- Can be automated



# OAI-PMH, obstacles

- No existing harvesting solution for DSpace
- Supports harvesting of metadata specifically

# Disseminating content

- How do you disseminate content through a metadata harvesting protocol?
  - Wrap it in a packaging format
  - Include the metadata
  - Encode the references to the files
  - Harvest the package

# METS, advantages

- Metadata Encoding and Transmission Standard
- Maintained by the Library of Congress
- Mature standard
- Widely adopted

# Packaging, disadvantages

- Complete packaging format
- Open to interpretation
- Ambiguities at the OAI-PMH layer

# OAI-ORE

*“Open Archives Initiative Object Reuse and Exchange defines standards for the description and exchange of aggregations of Web resources.”*

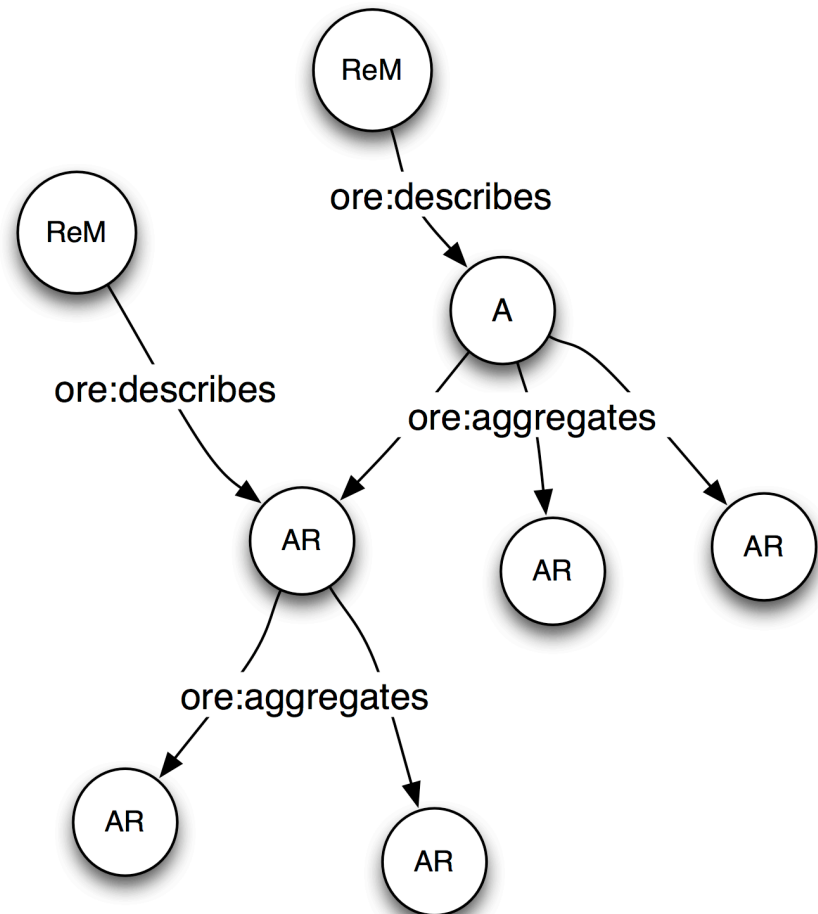
- Specialized
- Simple

# Mapping DSpace to OAI-ORE

- *ORE Abstract Data Model*
- *DSpace architecture*
- *The Mapping*

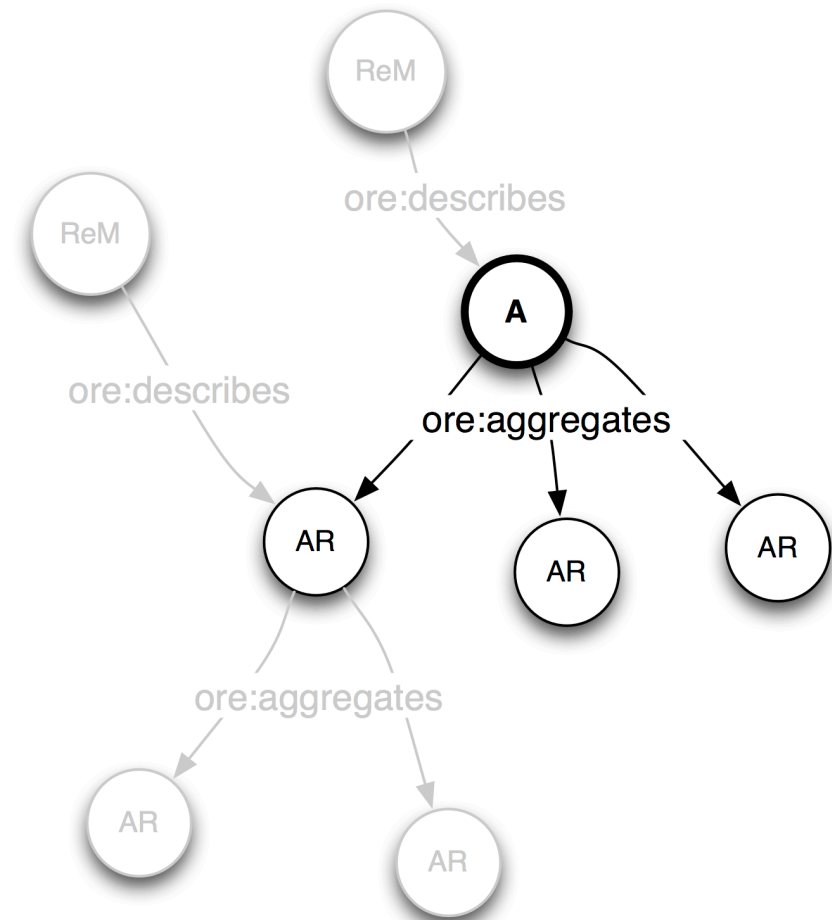
# ORE Data Model

- Aggregations
- Aggregated Resources
- Resource Maps



# Aggregation (A)

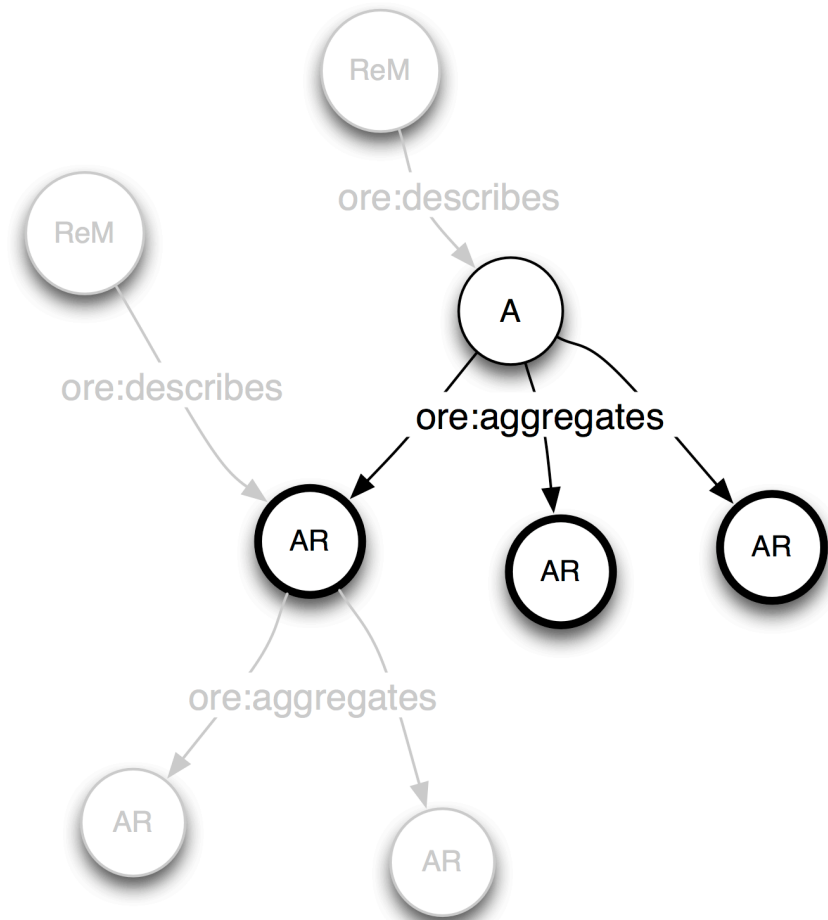
- Describes a set of resources
- Conceptual construct





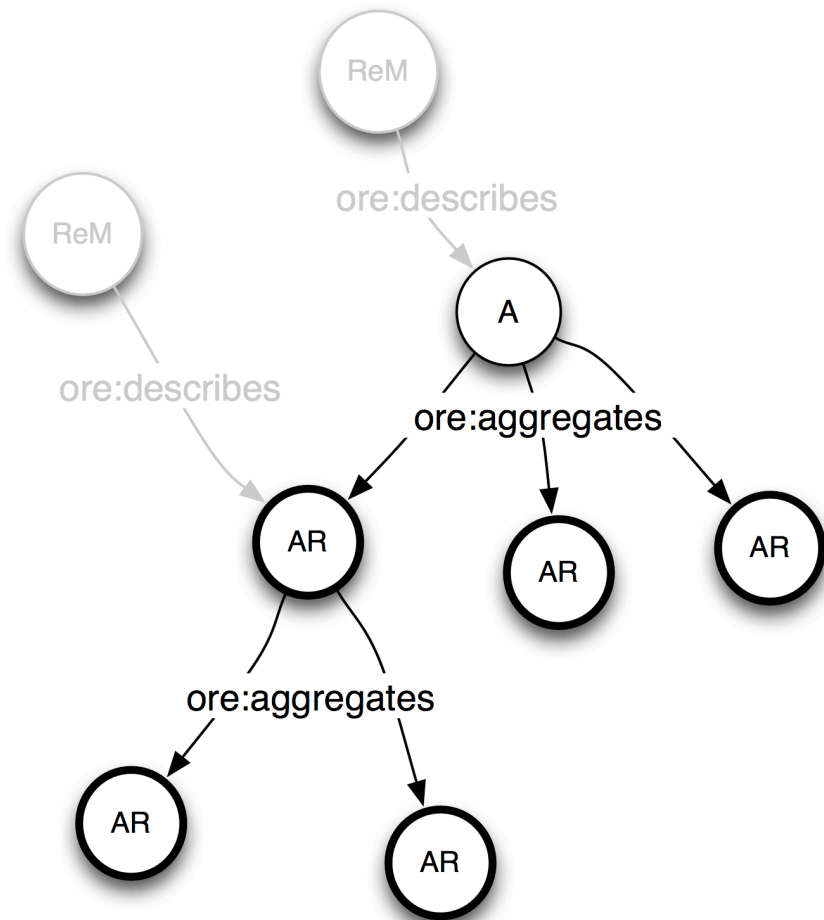
# Aggregated Resource (AR)

- Object of interest
- Part of an aggregation
- Can itself be an aggregation



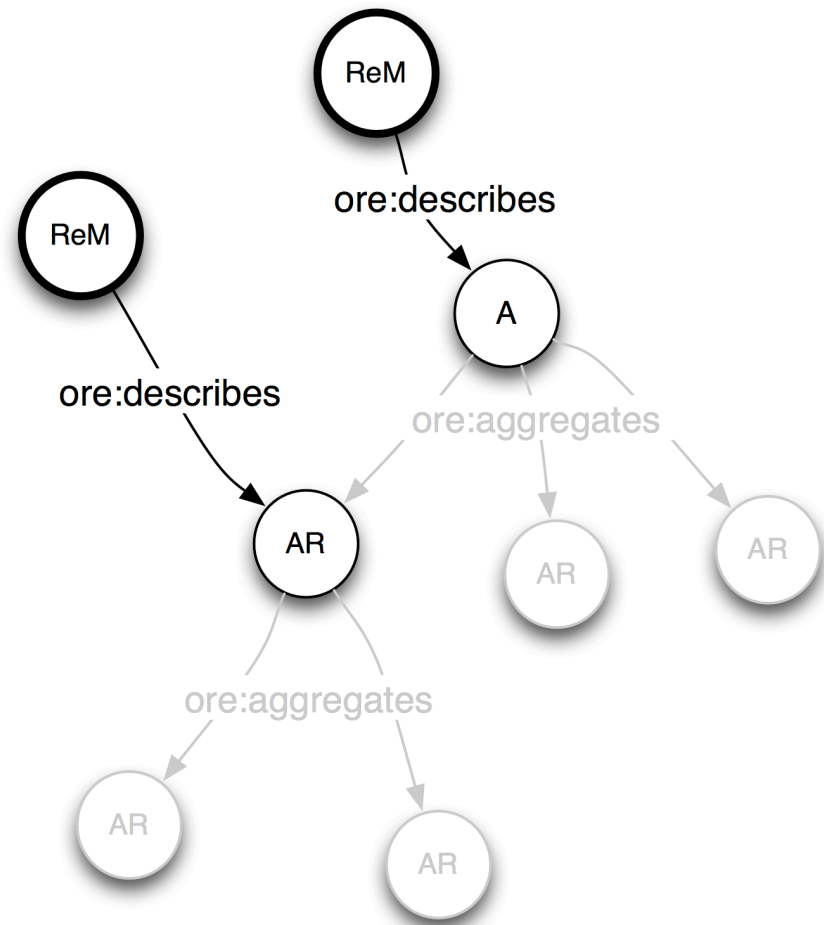
# Aggregated Resource (AR)

- Object of interest
- Part of an aggregation
- Can itself be an aggregation



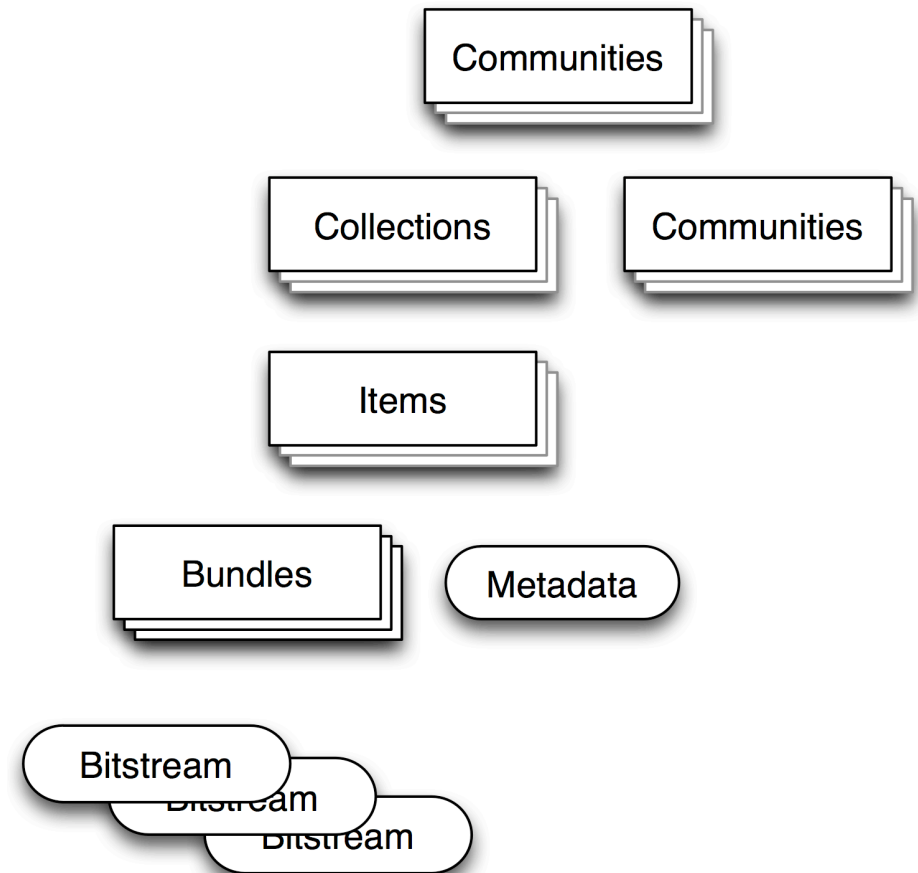
# Resource Map (ReM)

- Describes an aggregation
- Enumerates its aggregated resources
- Can be serialized in RDF or Atom XML



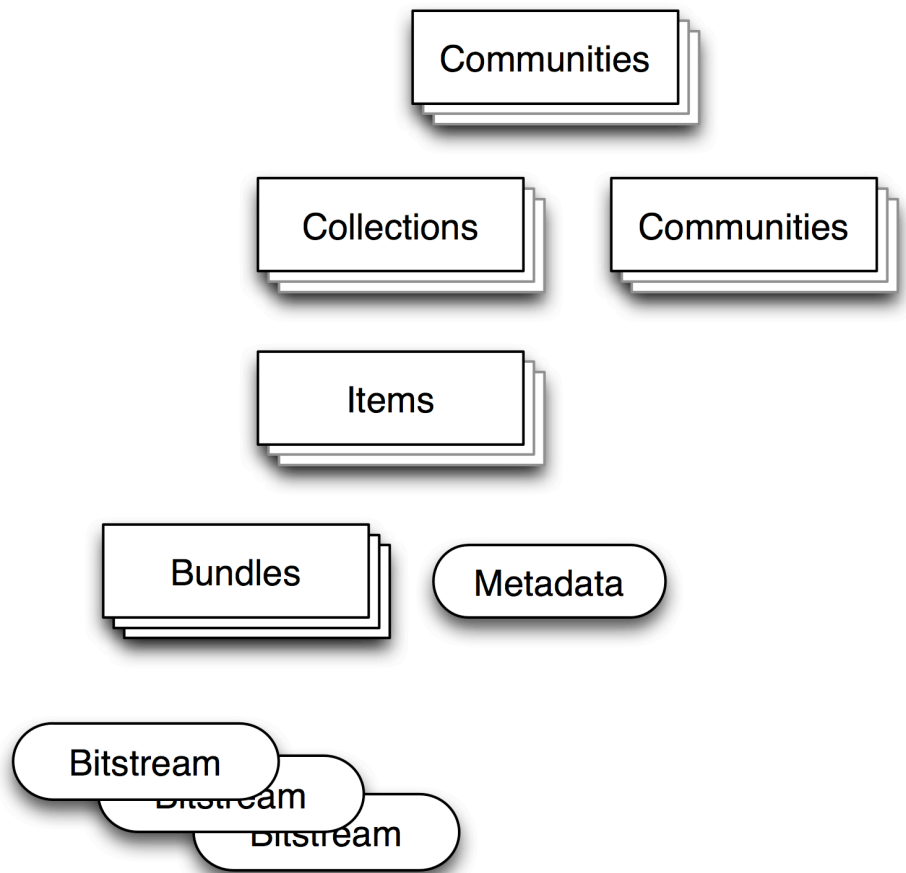
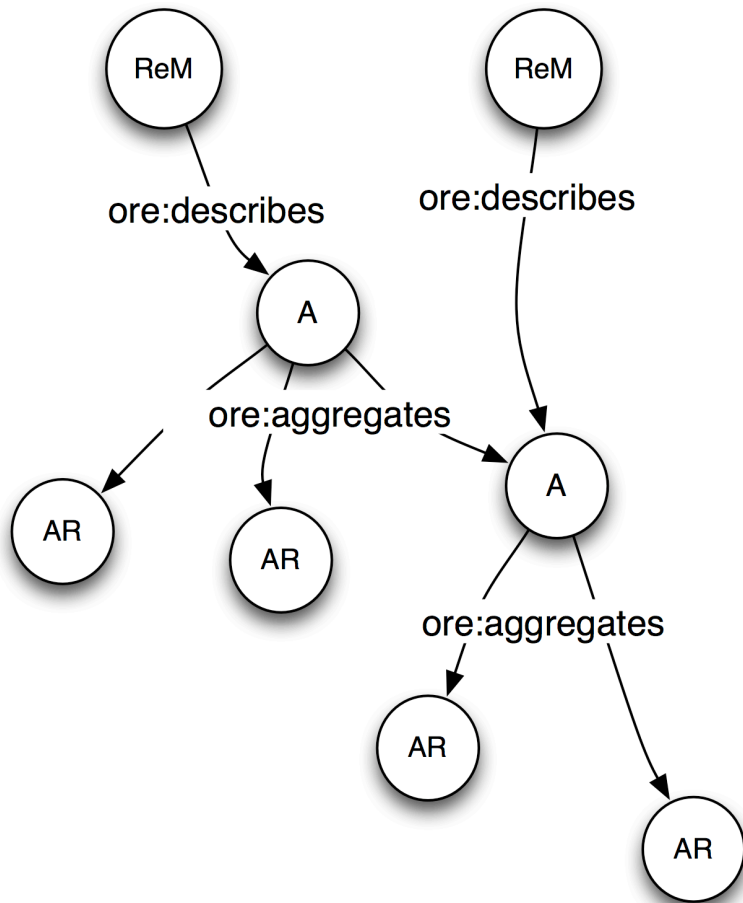
# DSpace Model v1.x

- Communities
- Collections
- Items
- Bundles
- Bitstreams

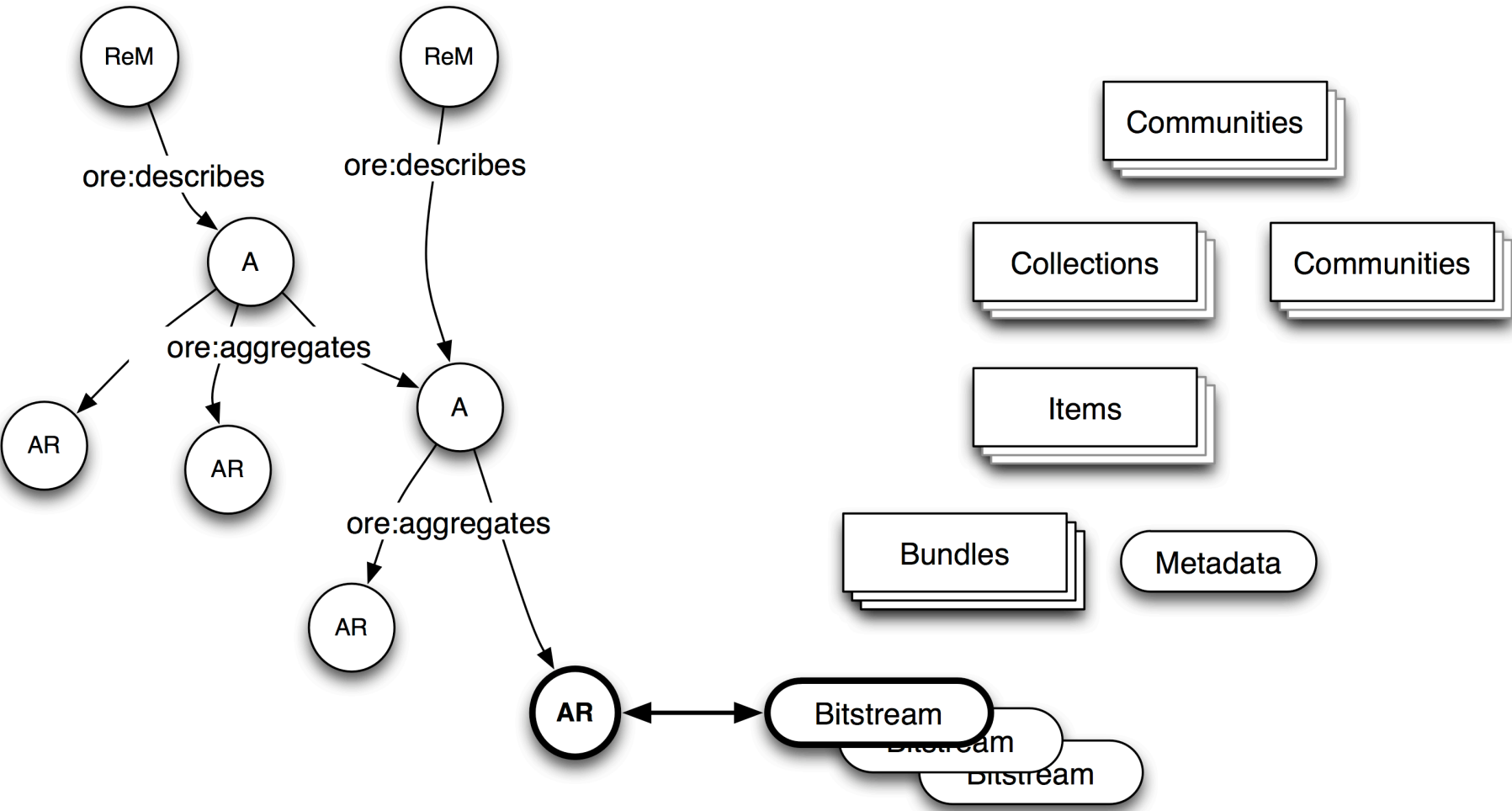


# ORE

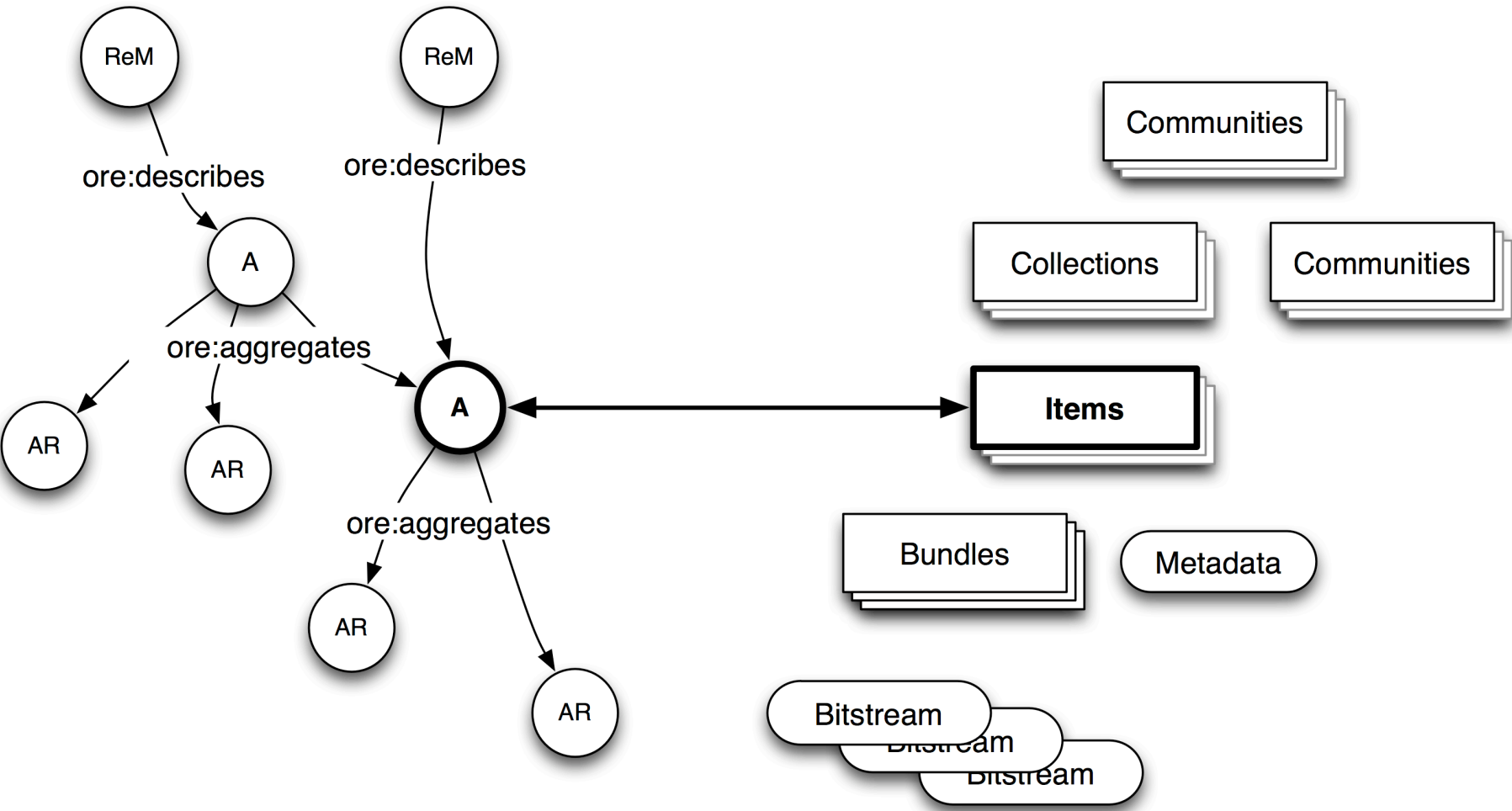
# DSpace



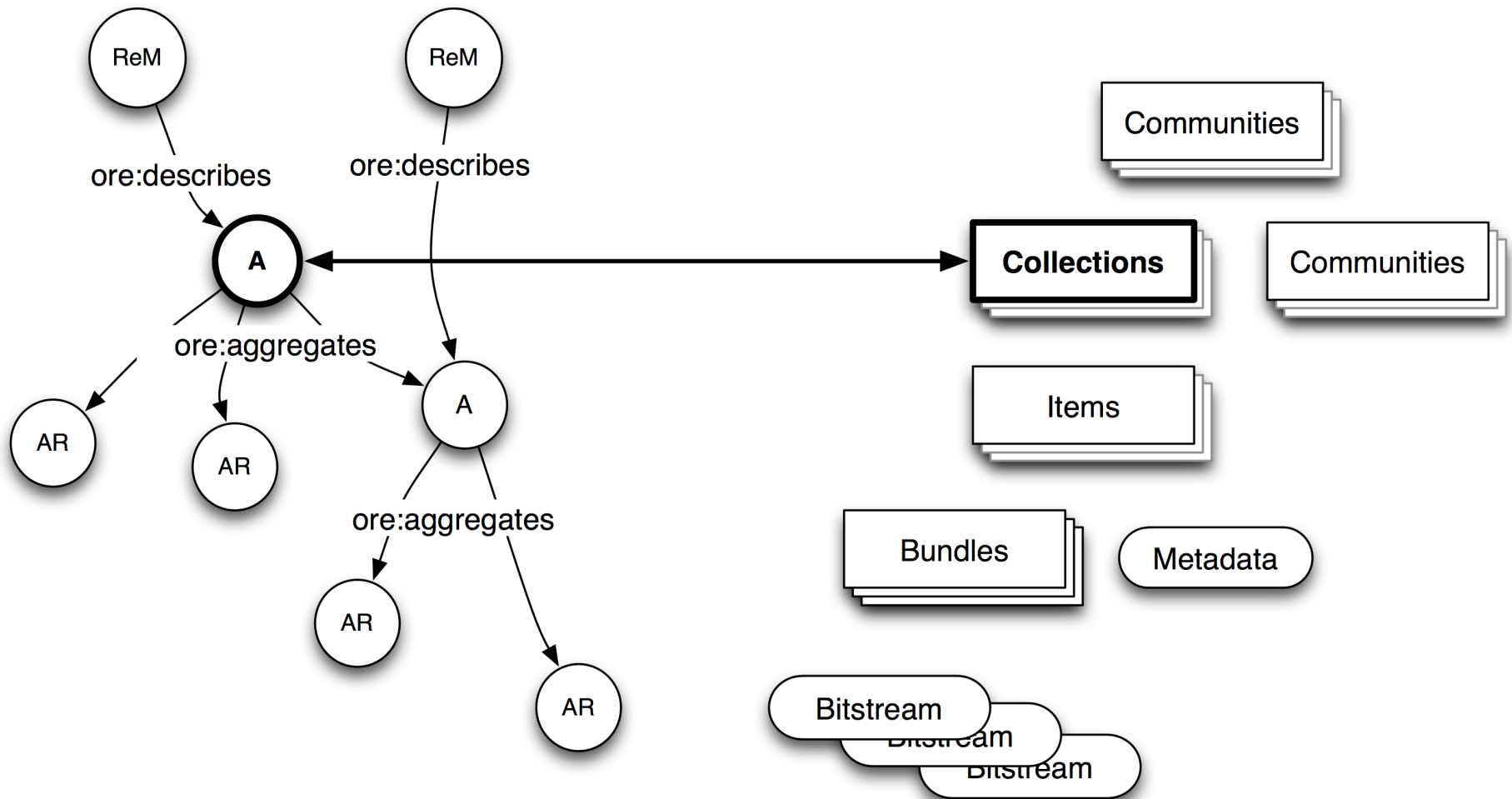
# Mapping



# Mapping

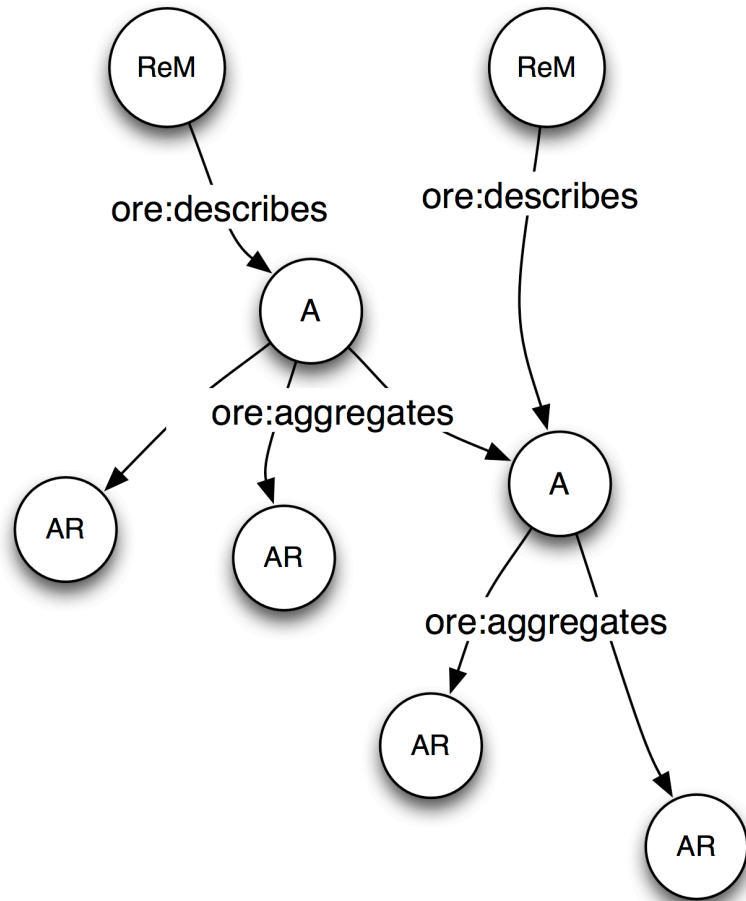


# Mapping





# Bundles?



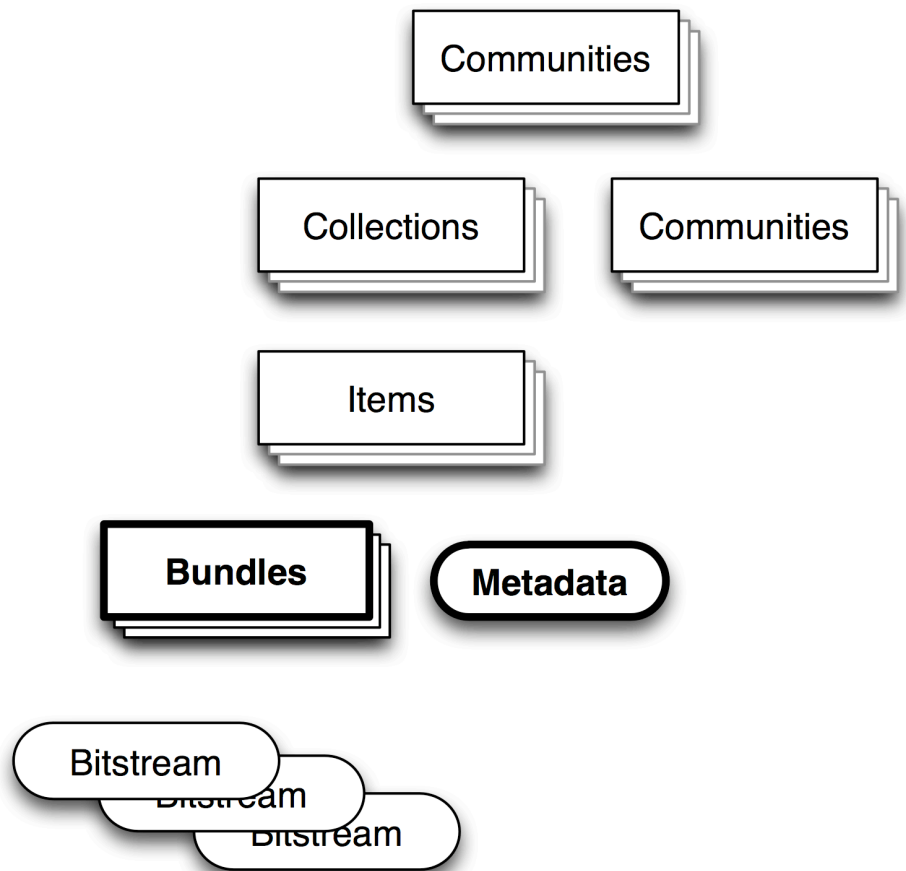
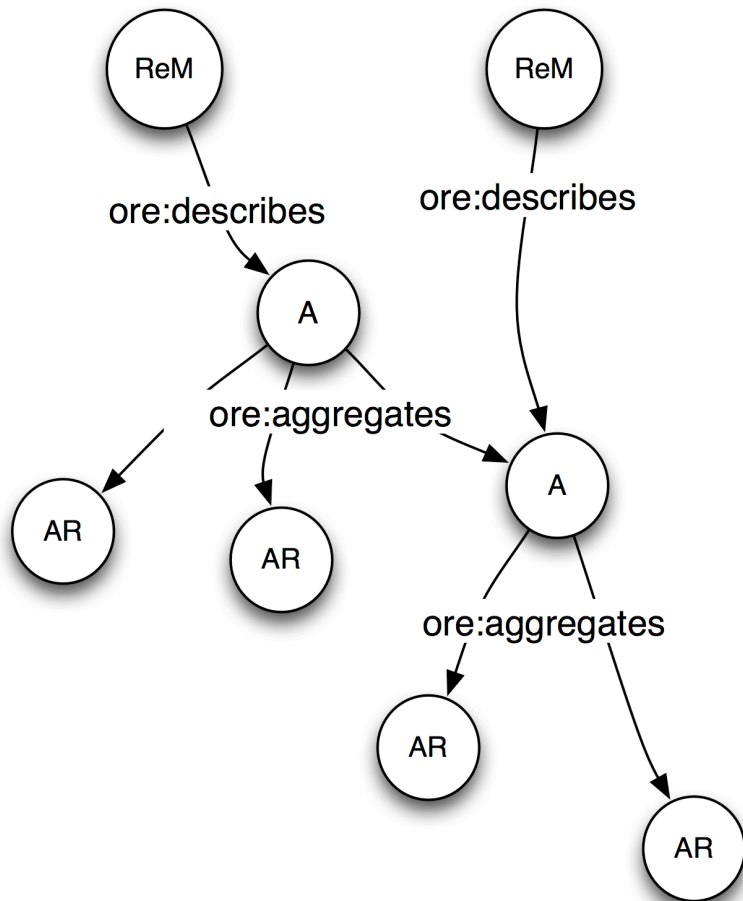
# Bundles, Potential Options

- Bundles as Aggregations of Bitstreams
- Bundles as filters for Aggregated Resources
- Bundles as DSpace-specific metadata

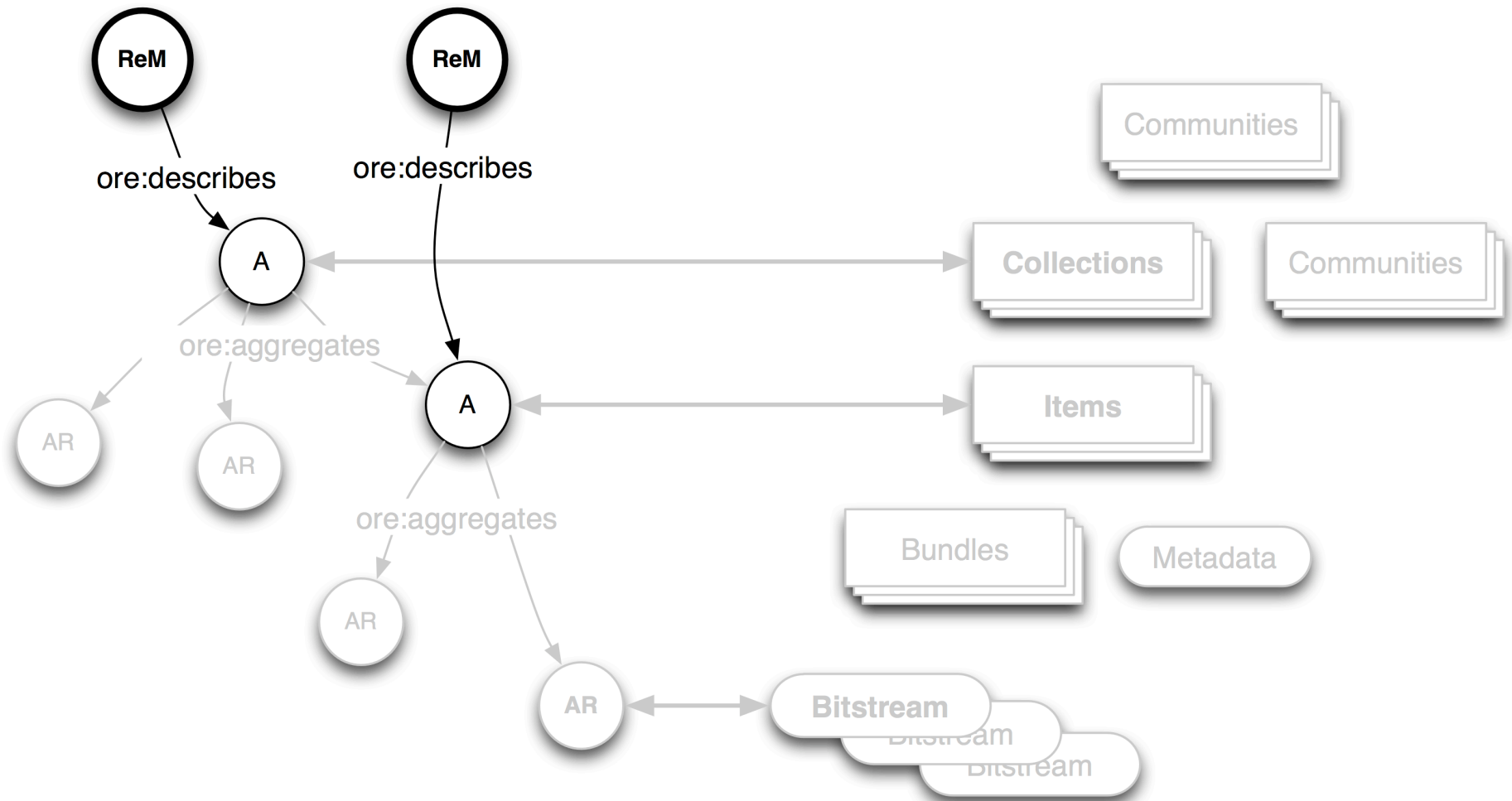
# Bundles, Observations

- By default, specialized for internal tasks
- Extendible for any use
- Obscured from the end user

# DSpace Bundles



# Serialization in Atom



# Implementation

- *ORE Dissemination*
- *ORE Harvesting*
- *Automation*

# Interfacing with DSpace

- Web UI
- LNI and SWORD
- Ingest and export scripts
- **Crosswalks**
  - **Ingestion**
  - **Dissemination**

# ORE Dissemination Crosswalk

- Requires:
  - A DSpace Item
- Produces:
  - Atom-serialized ORE ReM



# ORE Dissemination via OAI-PMH

- Dissemination crosswalk produces ORE ReMs from DSpace Items
- OAI-PMH data provider disseminates them

# ORE Harvesting

- Item-level ORE ReM interpreter
- Collection-level OAI-PMH harvester
- Repository level harvest scheduler

# ORE Ingestion Crosswalk

- Requires:
  - A DSpace Item
  - Atom-serialized ORE ReM
- Produces:
  - A DSpace Item with Bitstreams created from AR's

# OAI-PMH Harvester

- Queries remote OAI-PMH providers
- Processes responses as individual records
- Implemented at Collection level

# Collection Settings

- Source of collection's content
- OAI-PMH provider information
- Harvesting Level

DSpace Home → Collections → Harvesting

## Notice

Harvesting settings are valid.

### Edit Collection: ETD time test 4.15 (UT-TAMU)

- [Edit Metadata](#)
- [Assign Roles](#)
- [Content Source](#)

**Content source:**

This is a standard DSpace collection

This collection harvests its content from an external source

**Harvested Collection Location**

**OAI Provider:**   
The url of the target repository's OAI provider service

**OAI Set id:**   
The persistent identifier used by the OAI provider to designate the target collection

**Metadata Format:**

**Harvesting Options**

**Content being harvested:**

Harvest metadata only.

Harvest metadata and references to bitstreams (requires ORE support).

Harvest metadata and bitstreams (requires ORE support).

Advance

All of DS

- > Com
- > By Is
- > Aut
- > Title
- > Subj

Logout

Profile

Submissi

Access (

- > Peop
- > Gros
- > Aut

Registri

- > Met
- > Forn

Items

Withdraw

Control I

# Collection Source

DSpace Home → Collections → Harvesting

## Notice

Harvesting settings are valid.

### Edit Collection: ETD time test 4.15 (UT-TAMU)

- [Edit Metadata](#)
- [Assign Roles](#)
- [Content Source](#)

**Content source:**

This is a standard DSpace collection

This collection harvests its content from an external source

**Harvested Collection Location**

**OAI Provider:**

**Content source:**

This is a standard DSpace collection

This collection harvests its content from an external source

**Harvesting Options**

**Content being harvested:**

Harvest metadata only.

Harvest metadata and references to bitstreams (requires ORE support).

Harvest metadata and bitstreams (requires ORE support).

[Test Settings](#)

Advanced Search

All of DSpace

- > [Com](#)
- > [By Is](#)
- > [Aut](#)
- > [Titl](#)
- > [Subj](#)

[Logout](#)

[Profile](#)

[Submissi](#)

Access Control

- > [Peop](#)
- > [Grou](#)
- > [Aut](#)

Registri

- > [Met:](#)
- > [Forn](#)

Items

[Withdraw](#)

[Control I](#)

# OAI-PMH Settings

DSpace Home → Collections → Harvesting

## Harvested Collection Location

OAI Provider:

The url of the target repository's OAI provider service

OAI Set id:

The persistent identifier used by the OAI provider to designate the target collection

Metadata  
Format:

Advance

All of DS

- > Com
- > By Is
- > Auth
- > Title
- > Subj

Logout

Profile

Submissi

Access C

- > Peop
- > Gros
- > Auth

Registri

- > Meti
- > Forn

Items

Withdraw

Control I

# Harvest Level

[DSpace Home](#) → [Collections](#) → [Harvesting](#)

## Notice

Harvesting settings are valid.

## Edit Collection: ETD time test 4.15 (UT-TAMU)

- [Edit Metadata](#)
- [Assign Roles](#)
- [Content Source](#)

### Harvesting Options

**Content being harvested:**

- Harvest metadata only.
- Harvest metadata and references to bitstreams (requires ORE support).
- Harvest metadata and bitstreams (requires ORE support).

### Harvesting Options

**Content being harvested:**

- Harvest metadata only.
- Harvest metadata and references to bitstreams (requires ORE support).
- Harvest metadata and bitstreams (requires ORE support).

[Save](#) [Return](#)

[Advance](#)

All of DS

> [Com](#)  
> [By Is](#)  
> [Aut](#)  
> [Titl](#)  
> [Subj](#)

[Logout](#)

[Profile](#)

[Submissi](#)

Access (

> [Peop](#)  
> [Grou](#)  
> [Aut](#)

Registri

> [Met](#)  
> [Forn](#)

[Items](#)

[Withdraw](#)

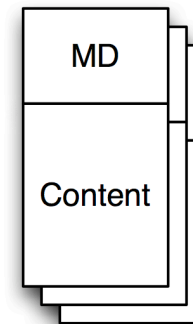
[Control I](#)



# Harvesting a Collection

Local collection  
(OAI-PMH harvester)

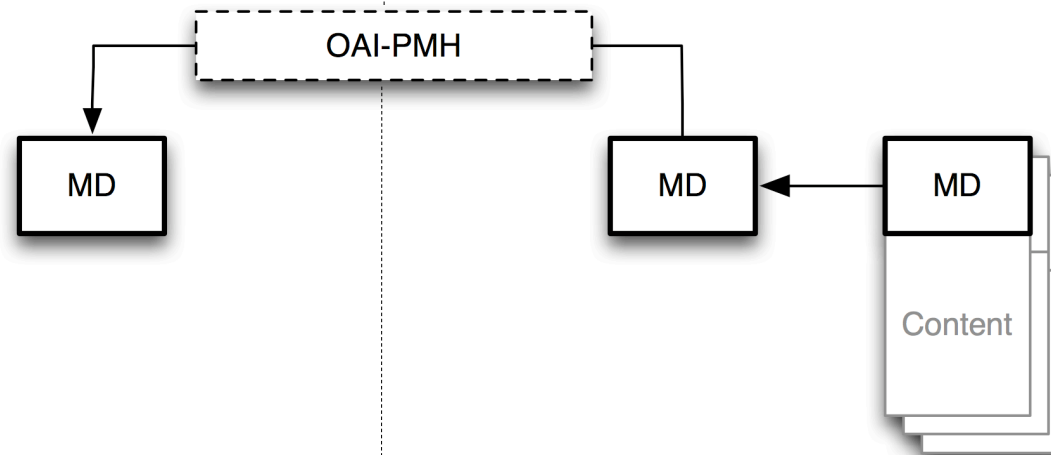
Remote collection  
(OAI-PMH provider)



# Harvest Metadata

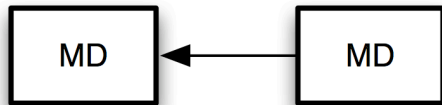
Local collection  
(OAI-PMH harvester)

Remote collection  
(OAI-PMH provider)

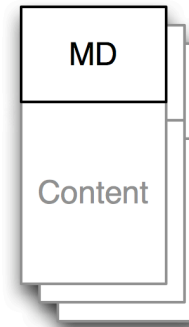


# Metadata Replicated

Local collection  
(OAI-PMH harvester)



Remote collection  
(OAI-PMH provider)

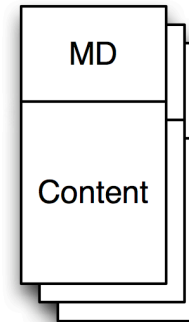


# Case 1: Metadata Only

Local collection  
(OAI-PMH harvester)



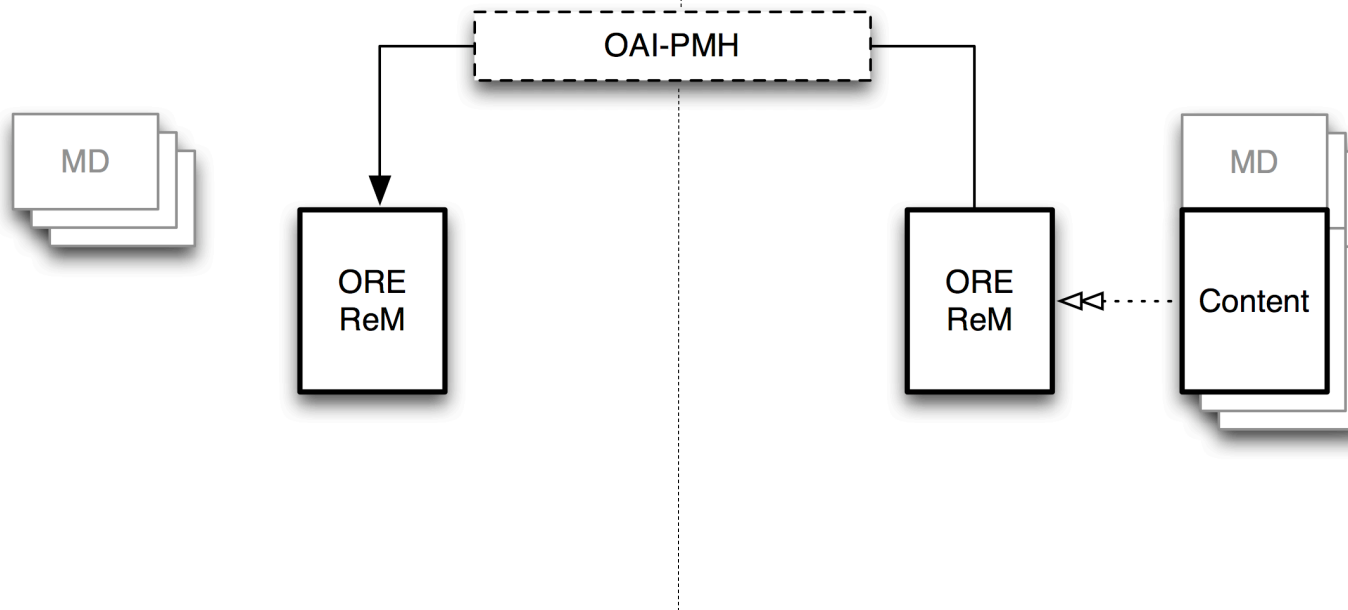
Remote collection  
(OAI-PMH provider)



# Harvest ORE ReMs

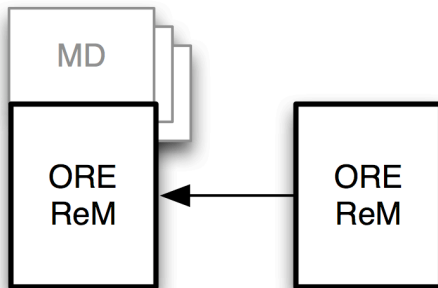
Local collection  
(OAI-PMH harvester)

Remote collection  
(OAI-PMH provider)

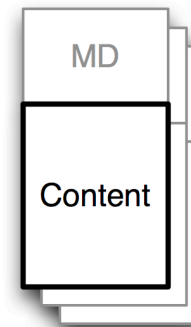


# Case 2: Metadata + Content Ref's

Local collection  
(OAI-PMH harvester)

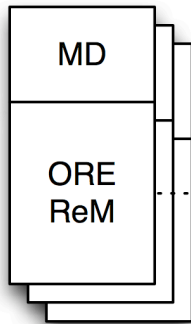


Remote collection  
(OAI-PMH provider)

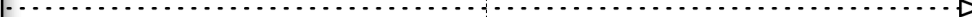
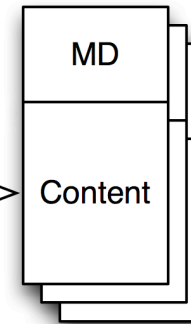


# Case 2: Metadata + Content Ref's

Local collection  
(OAI-PMH harvester)



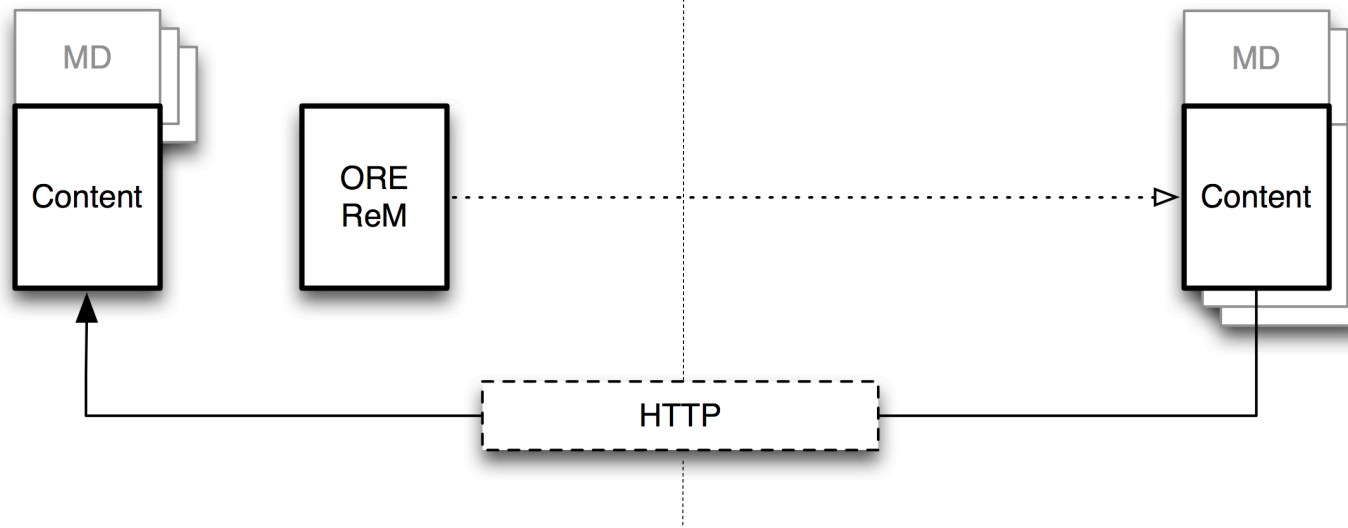
Remote collection  
(OAI-PMH provider)



# Case 3: Metadata + Content

Local collection  
(OAI-PMH harvester)

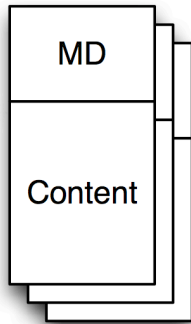
Remote collection  
(OAI-PMH provider)



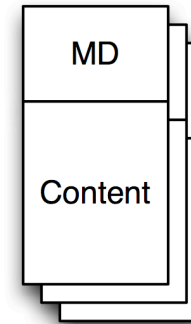


# Case 3: Metadata + Content

Local collection  
(OAI-PMH harvester)



Remote collection  
(OAI-PMH provider)



# Harvest Scheduling System

- Monitors harvested collections
- Starts harvests at regular intervals
- Alerts administrators of errors

# Results

- *The Primary Use Case*
- *TDL in General*
- *The Greater Web Community*

# Harvesting using PMH+ORE

- Federated ETD collection currently in pre-production at TDL
- Addresses primary requirements
  - Performs maintenance automatically
  - Detects changes in existing content
  - Supports interchange of metadata and content

# Other Possibilities

- Specialized DSpace instances
- Flexible repository architecture
- Interoperability with other repository systems

# Current Priorities

- Live deployment at TDL
- Release to the open source community
- Integration into DSpace 1.6



INSTITUTE *of*  
**Museum** and **Library**  
SERVICES

*National Leadership Grant #LG-05-07-0095-07*

Questions?