



## Integrating DSpace with DuraCloud

DuraSpace Webinar: 30 Nov 2011

Tim Donohue  
DuraSpace



Licensed under Creative Commons Attribution-Share Alike 3.0 Unported License (CC BY-SA 3.0)  
To request other use: [info@duracloud.org](mailto:info@duracloud.org)



# Basis for DSpace Integration

## 1. DSpace AIP Backup & Restore (1.7 +)

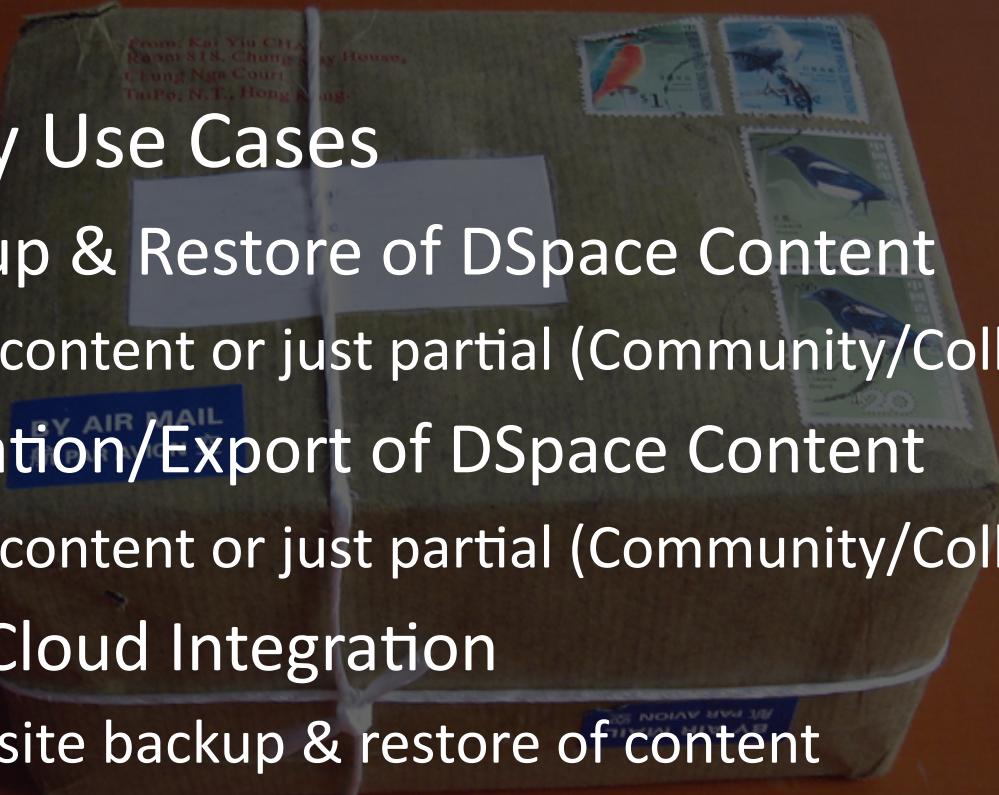
(Initial DuraCloud use case: Backup & Restore)

## 2. DSpace Curation Task System (1.7 +)

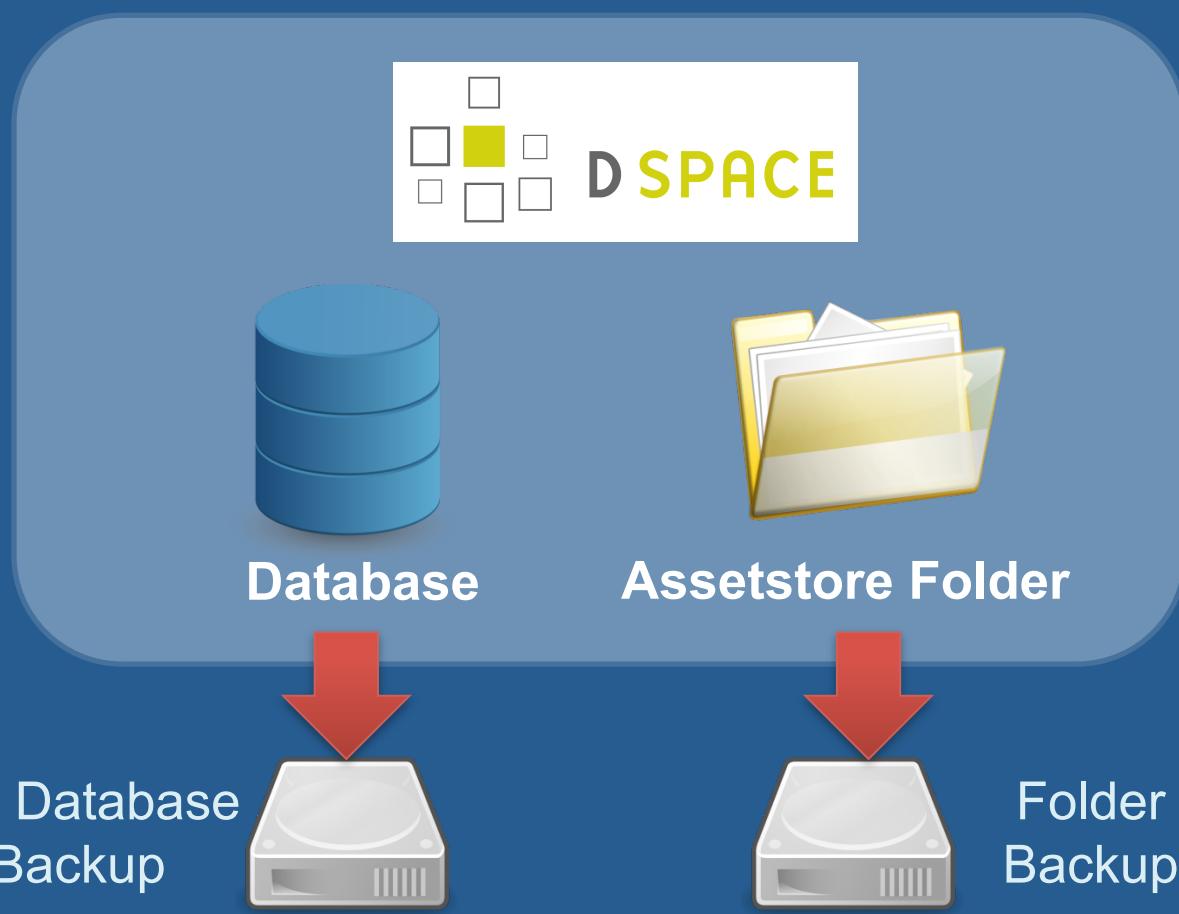
## 3. DSpace Replication Task Suite (1.8 add-on)

# Intro to Archival Info Pkgs (1.7+)

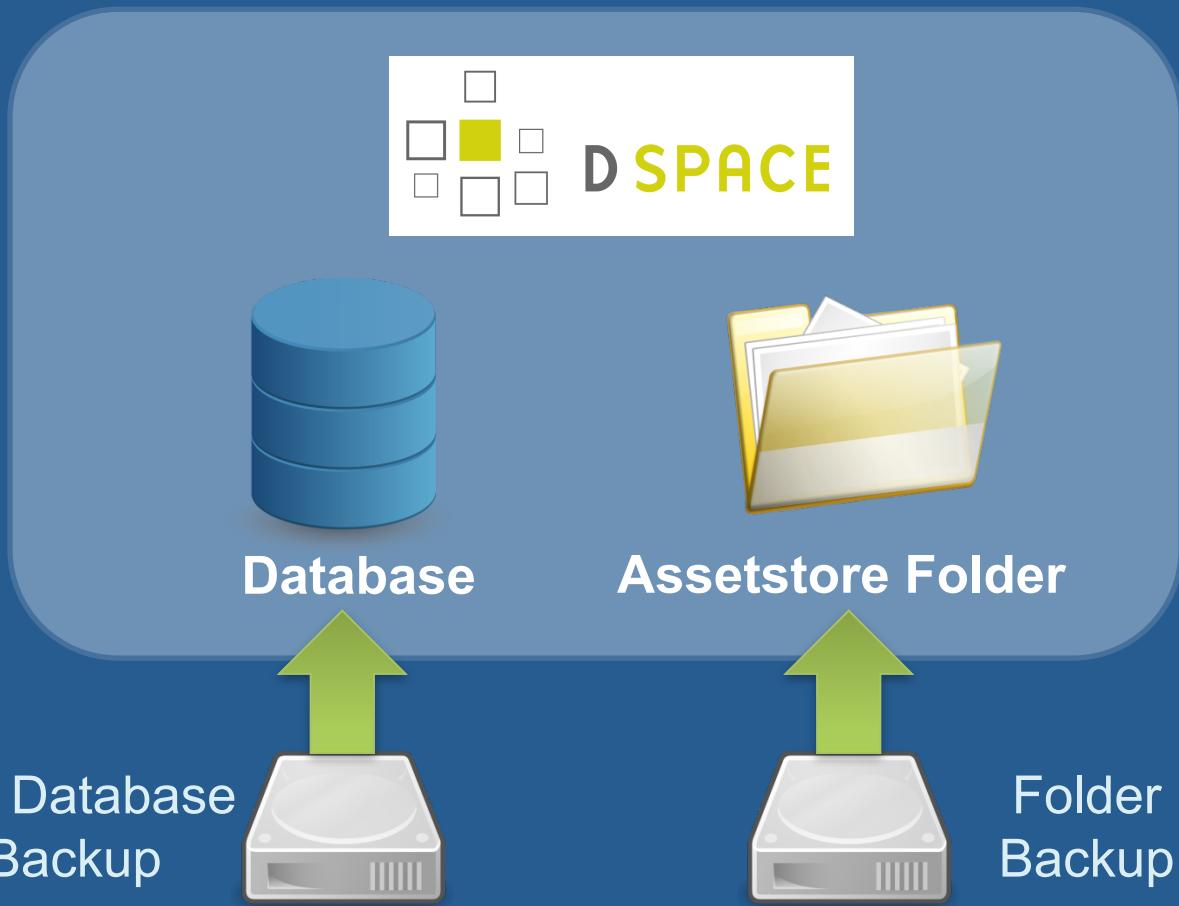
- Primary Use Cases
  - Backup & Restore of DSpace Content
    - All content or just partial (Community/Collection/Item)
  - Migration/Export of DSpace Content
    - All content or just partial (Community/Collection/Item)
  - DuraCloud Integration
    - Offsite backup & restore of content



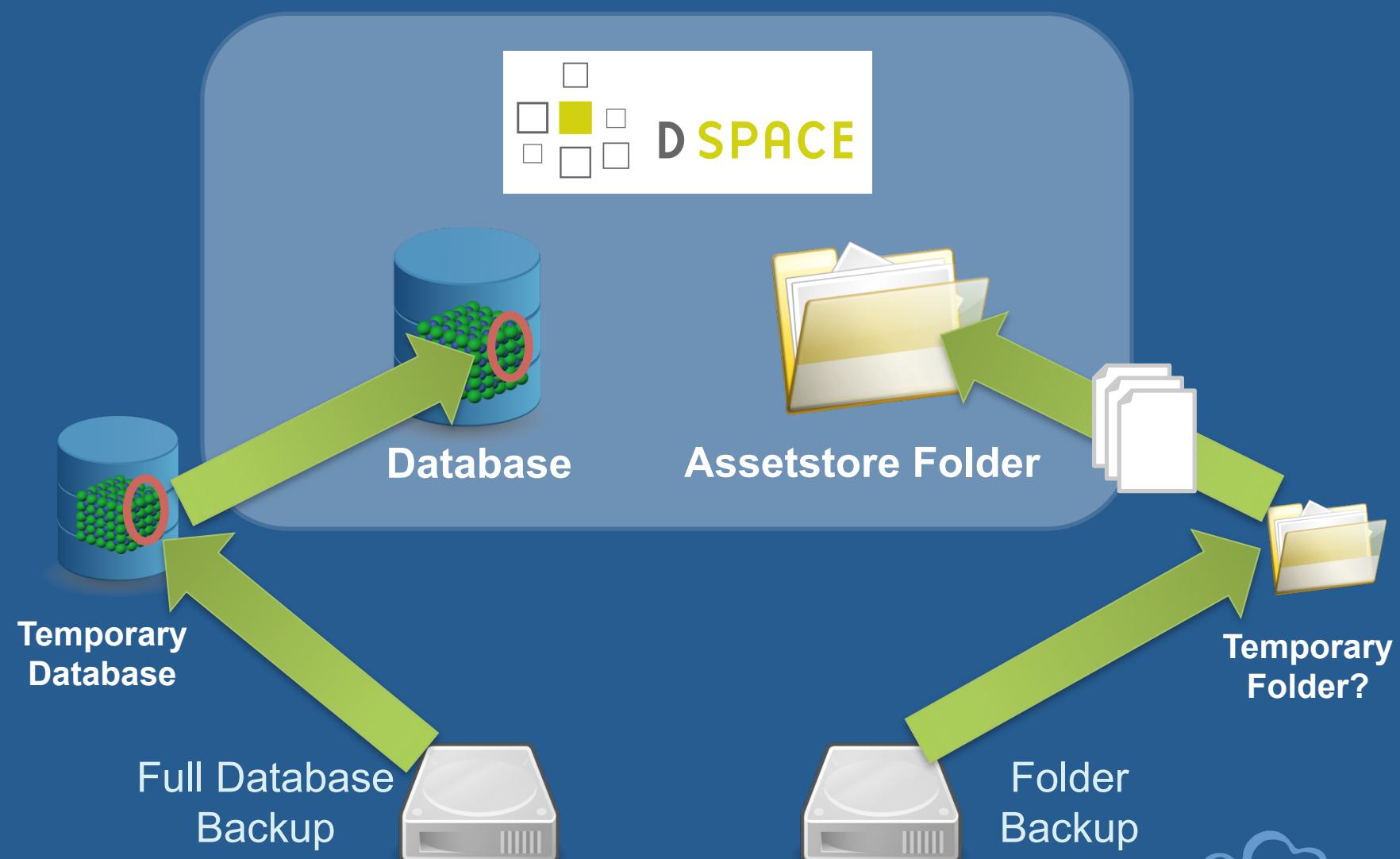
# How to Backup DSpace (pre-1.7)



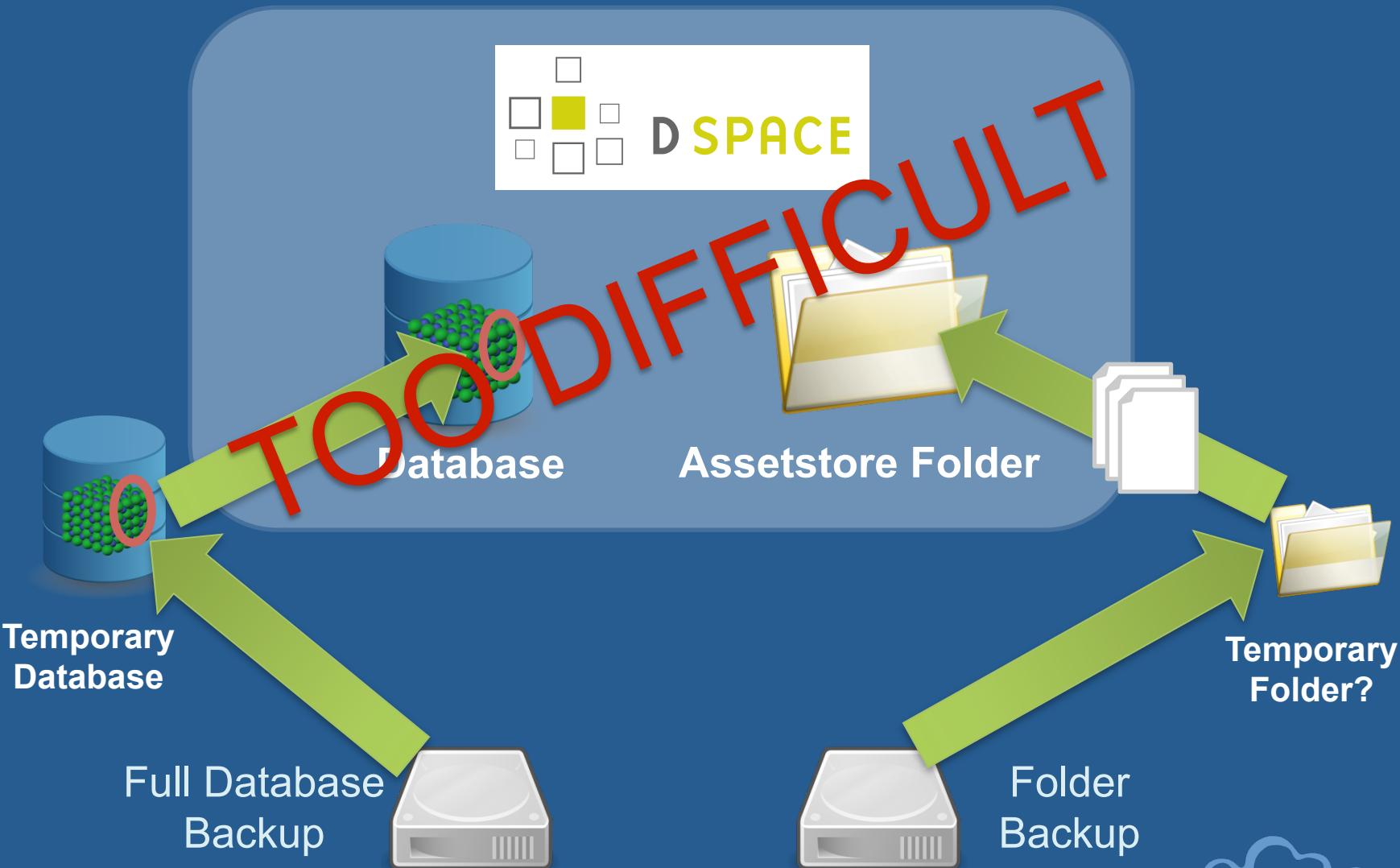
# How to Restore All (pre-1.7)



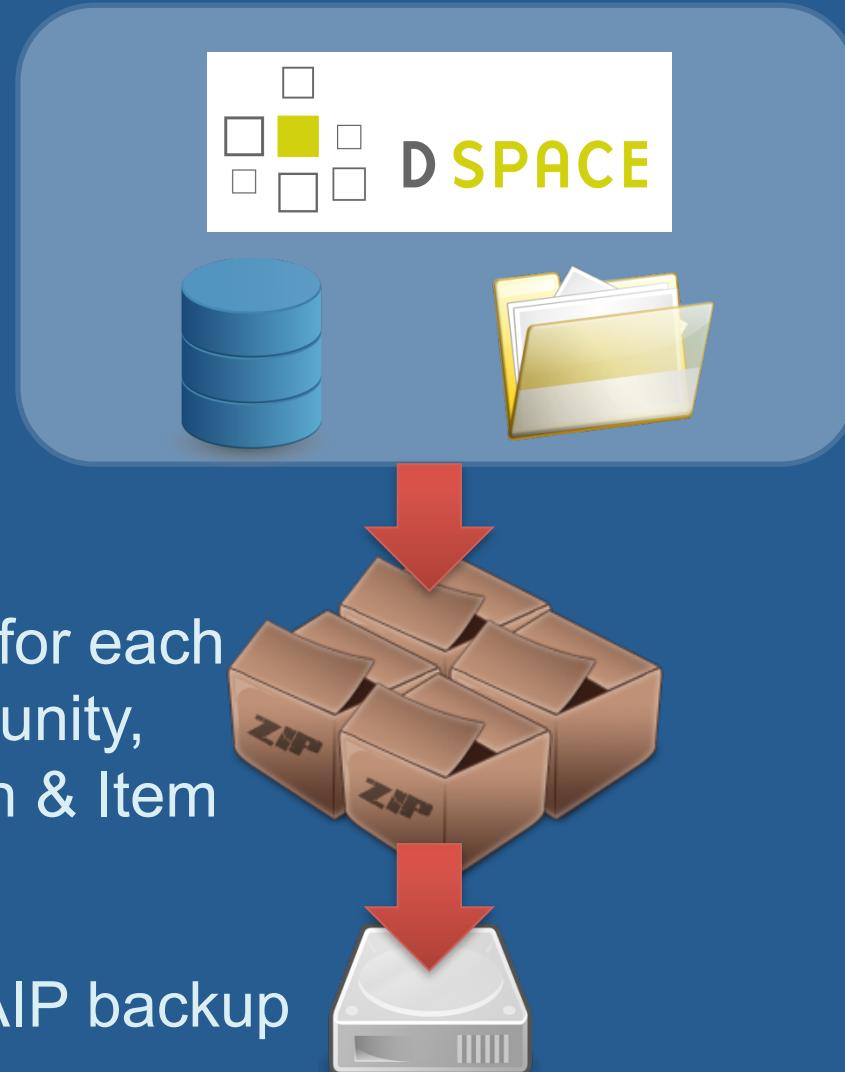
# How to Restore a Collection (pre-1.7)



# How to Restore a Collection (pre-1.7)



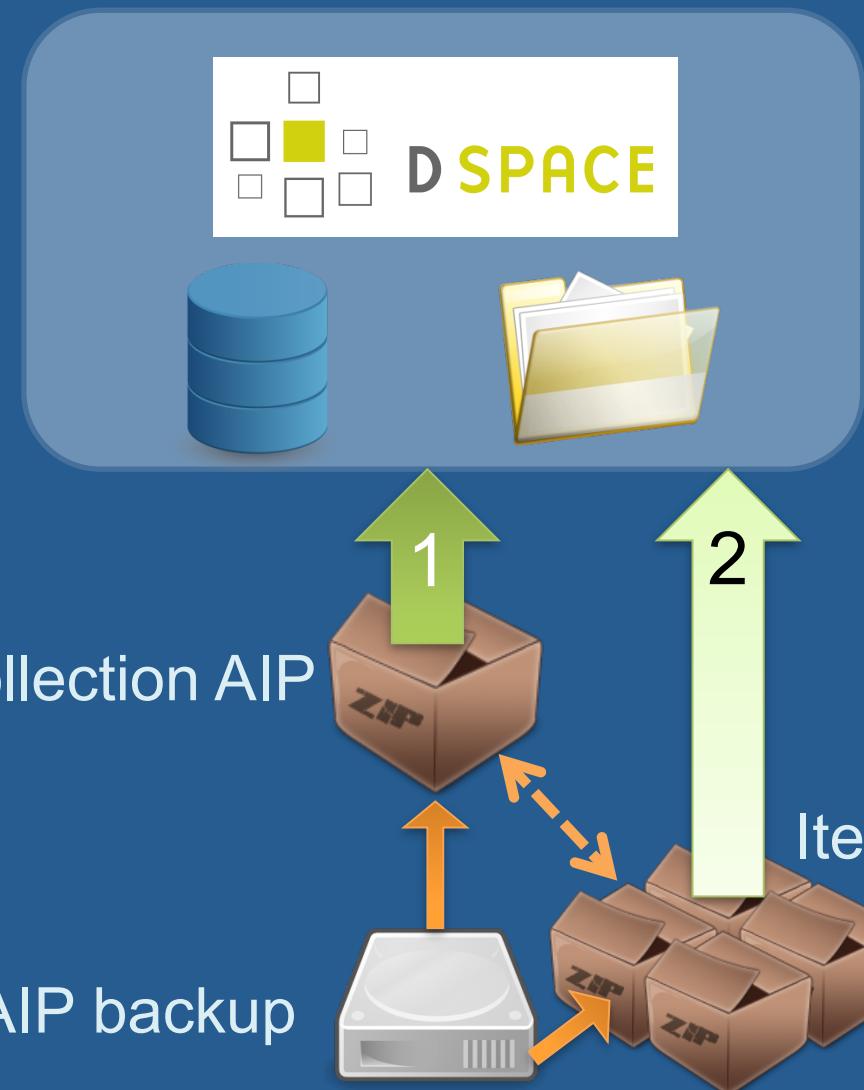
# Backup via Archival Info Pkgs



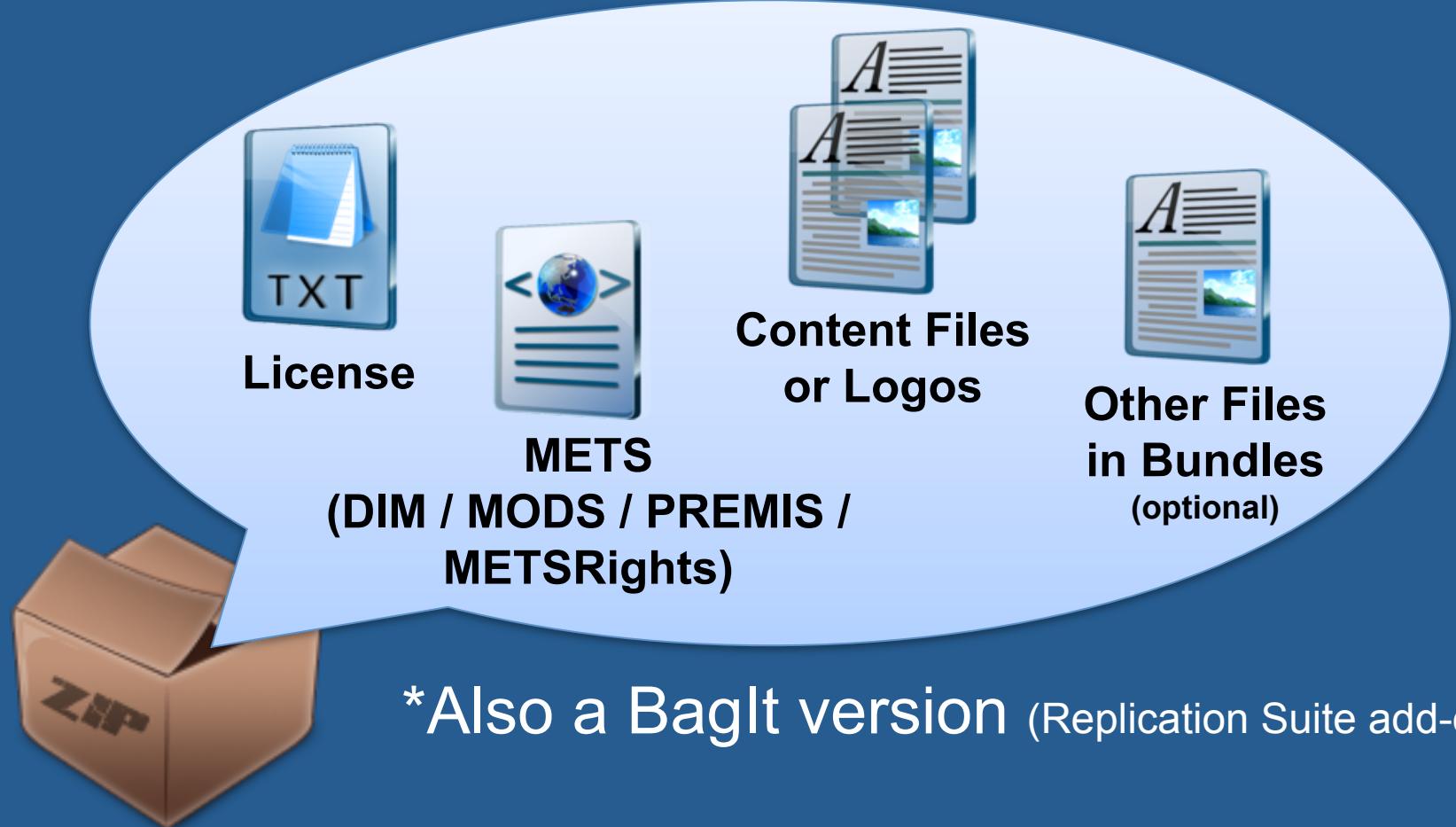
# Restore All via Archival Info Pkgs



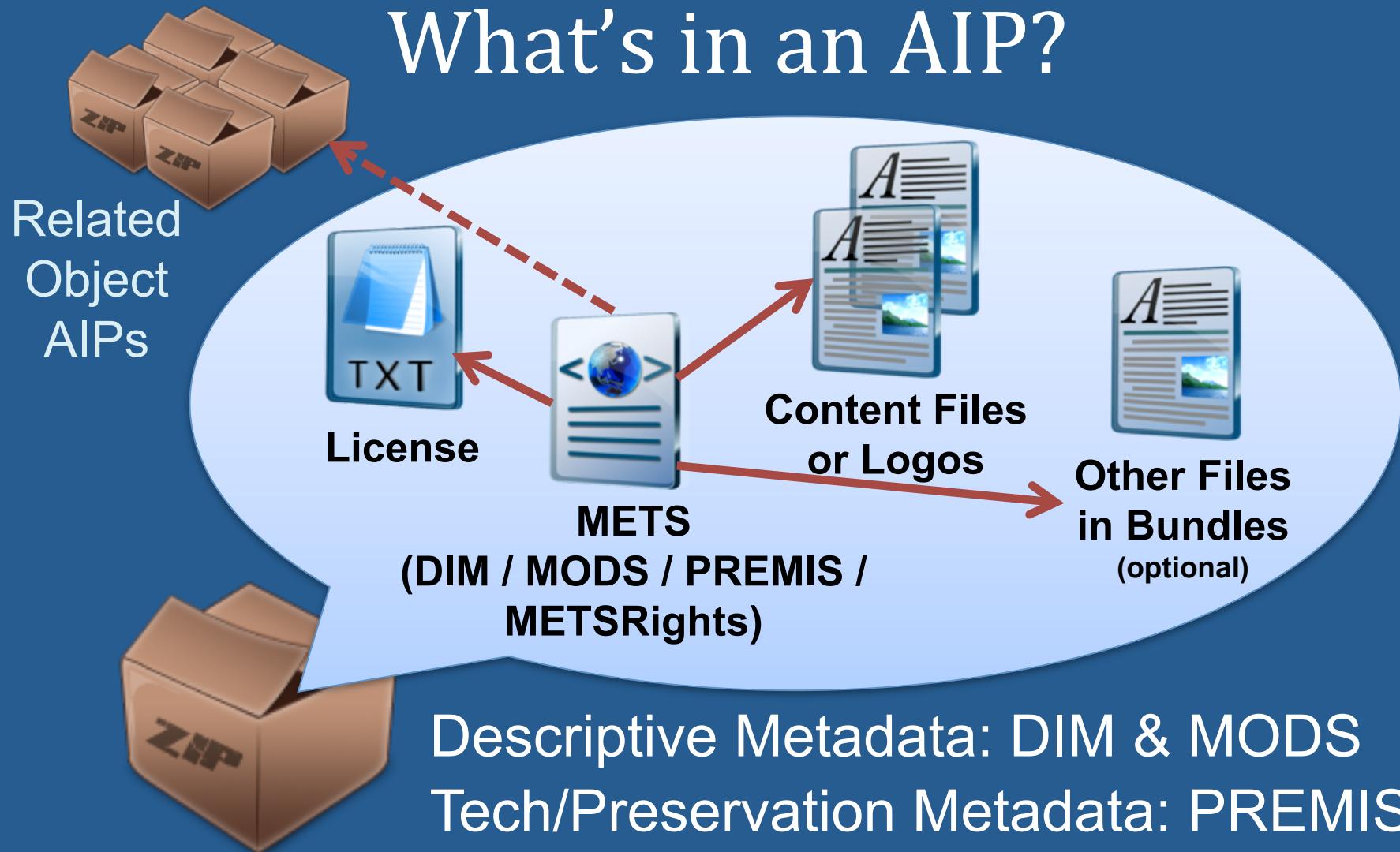
# Restore a Collection via AIPs



# What's in an AIP?



# What's in an AIP?



# The “Site” AIP



Top-Level  
Community  
AIPs



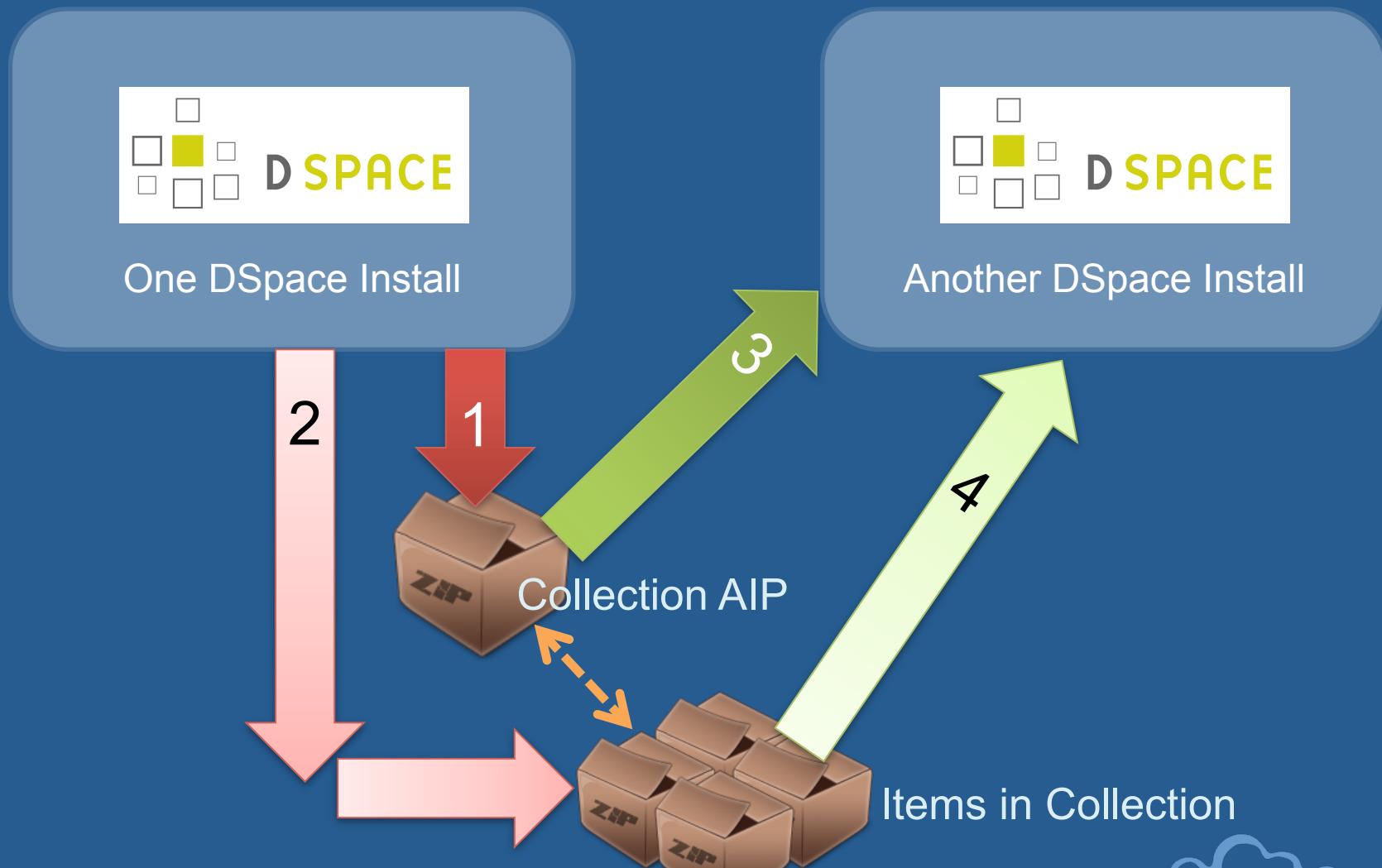
**METS**  
**(DIM / MODS / PREMIS /**  
**METSRights)**

Special AIP for site-wide info/metadata:  
(e.g. Group Memberships, EPeople)

# What can AIPs restore?

- ✓ Restore All In-Archive Content (Files + Metadata)
- ✓ Restore All People & Groups
- ✓ Restore All Permissions / Access Rights
- ✓ Restore Community / Collection Logos, Metadata, Rights & Item Templates
- ✓ Restore Community / Collection / Item Hierarchy
- ✗ Restore In-Process / Incomplete Items
- ✗ Restore Collection OAI-PMH/ORE Harvest Settings
- ✗ Restore all configuration files (dspace.cfg, etc.)

# AIP Use Case: Migrate Content



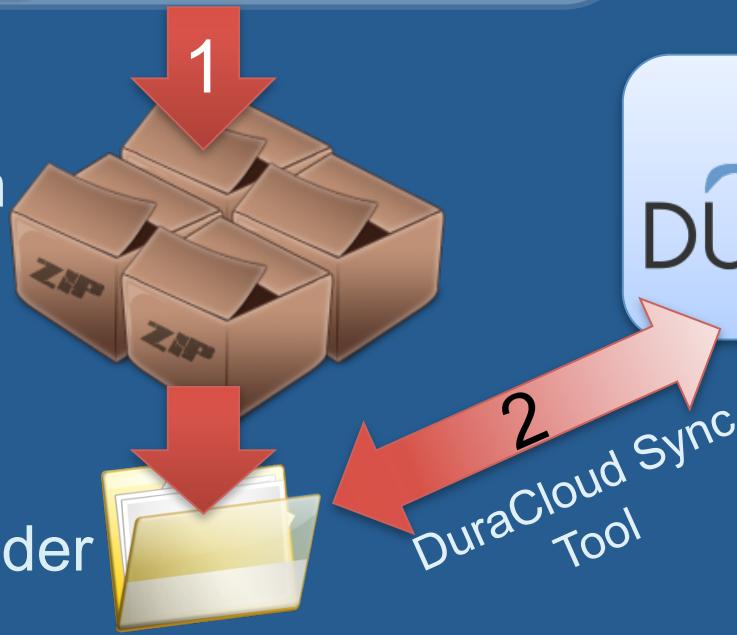
# “Manual” AIP backup to DuraCloud



This two step route is required for DSpace 1.7.x

Package for each Community, Collection & Item

Local “Watch” Folder



# “Manual” AIP restore from DuraCloud



This two step route is required for DSpace 1.7.x

Package for each Community, Collection & Item

Local Folder



1  
DuraCloud Retrieval Tool





New: DSpace Replication Suite add-on

# Intro to DSpace Curation System

- Enables a basic ‘microservices’ approach to curating DSpace objects
- Anyone can build a task & share it
- Currently tasks must be written in Java
  - Experimental support for non-Java tasks in 1.8
- “Frees” admin tasks from Command Line
  - Can now run from Admin UI or CLI
  - Can also “queue” tasks for later processing

# DSpace 1.8.0 WARNING

- In DSpace 1.8.0, ANY curation task run across your entire DSpace site will report a “NullPointerException” error
- THIS DOES NOT MEAN YOUR TASK FAILED!
- Check dspace logs to see if task did succeed
- This bug will be fixed in 1.8.1
  - <https://jira.duraspace.org/browse/DS-1077>

# DSpace Replication Suite Add-on

- A set of curation tasks geared towards ‘replicating’ (backup/restore/audit) content
- Compatible with DSpace 1.8.1 or above
  - Not recommended on 1.8.0 (see previous slide)
- “Wraps” DSpace AIP Backup & Restore tool
- Provides configurable AIP storage plugins for filesystem or DuraCloud
- Provides optional BagIt packaged AIPs

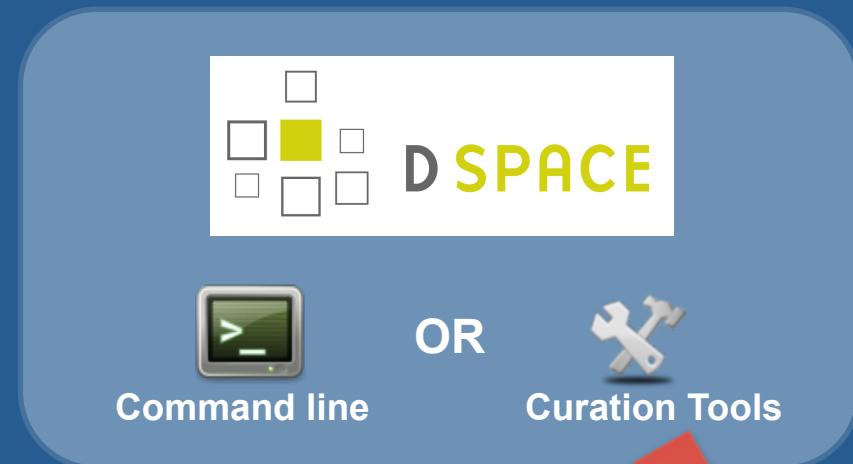
# The Suite of Tasks

- ✓ Transmit AIP(s)
- ✓ Verify AIP(s) exist
- ✓ Fetch AIP(s)
- ✓ Audit against AIP(s)
- ✓ Remove AIP(s)
- ✓ Restore Missing Object(s) from AIP(s)
- ✓ Replace Existing Object(s) from AIP(s)
- ✓ Read Odometer (I/O upload/download stats)
- ✓ Estimate storage size of AIP(s) (rough estimate)

All tasks can be configured to store AIPs in an existing  
DuraCloud account.

Local/Mounted filesystem storage is also supported.

# Backup AIP to DuraCloud



Package for each  
Community,  
Collection & Item

Local Temp Folder  
(Temporary Cache)

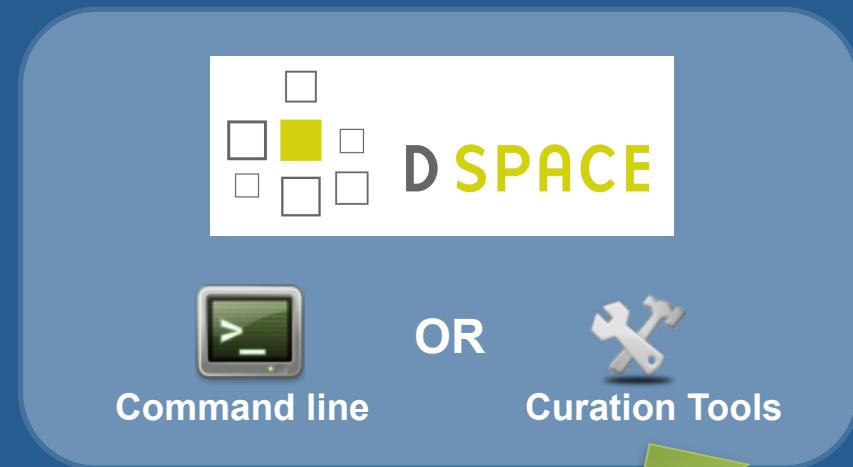


## Replication Task Suite:

- One step process:  
Generate and Upload AIP
- Via UI or CLI



# Restore AIP from DuraCloud



Package for each  
Community,  
Collection & Item

Local Temp Folder  
(Temporary Cache)

## Replication Task Suite:

- One step process:  
    Retrieve and Restore AIP
- Via UI or CLI
- Also 'auditing' tools





# Known Limitations

- 🚫 Cannot yet take advantage of DuraCloud streaming capabilities (AIPs are zip files)
- 🚫 Cannot yet take advantage of DuraCloud transformation services (AIPs are zip files)

# “Early Access” Release

- Early December (likely by Mon, Dec 5)
- This release is for early adopters to try out add-on & provide feedback
- Download from Replication Task Suite page:  
<https://wiki.duraspace.org/display/DSPACE/ReplicationTaskSuite>
- Install & Configuration instructions also on above page

# In Large Thanks to...

- **MIT** : Richard Rodgers & Wendy Bossons
  - Developed Curation Task Framework
  - Developed initial Replication Suite tasks (BagIt versions)
- **@mire** : Mark Diggory
  - Gave feedback throughout early development

# For More Information

- Replication Task Suite:
  - <https://wiki.duraspace.org/display/DSPACE/ReplicationTaskSuite>
- AIP Backup & Restore:
  - <https://wiki.duraspace.org/display/DSDOC18/AIP+Backup+and+Restore>
- Curation Task System:
  - <https://wiki.duraspace.org/display/DSDOC18/Curation+System>

# Photo/Icon Acknowledgments

- Package : <http://www.flickr.com/photos/halfbisqued/2353845688/>
- Harddrive & Terminal icons: [http://tango.freedesktop.org/Tango\\_Desktop\\_Project](http://tango.freedesktop.org/Tango_Desktop_Project)
- Folder icon: <http://www.openclipart.org/detail/13740>
- Database icon: <http://www.openclipart.org/detail/68413>
- Zip Pkg icon: <http://veryicon.com/icons/system/capital-icon-suite-mac/zip-10.html>
- File icons: <http://veryicon.com/icons/system/rhor-v2-part-3/>
- Checkmark & Delete icons: <http://veryicon.com/icons/system/on-stage/>
- Tools Icon: <http://veryicon.com/icons/system/azullustre/>