

AipBackupRestoreDiscussion

- 1 [AIP Backup & Restore Discussion](#)
 - 1.1 [Notes on Development](#)
 - 1.1.1 [What code has really changed \(as of DSpace 1.7\)?](#)
 - 1.2 [To-Do List – What remains to be done!](#)
 - 1.2.1 [Testing Special Cases during Restore/Replace](#)
 - 1.2.1.1 [Item Restoration/Replacement](#)
 - 1.2.1.2 [Collection Restoration/Replacement](#)
 - 1.2.1.3 [Community Restoration/Replacement](#)
 - 1.2.2 [Admin UI work](#)
 - 1.2.3 [Notes on AIP ingest speed & improving it](#)
 - 1.3 [Discussion / Use Cases](#)
 - 1.4 [Questions / Comments?](#)

The official AIP Backup & Restore Documentation is available at: [AIP Backup and Restore](#). This page is just for discussion around the existing feature and how to make it better.

AIP Backup & Restore Discussion

This page serves as a Discussion page for the [AIP Backup and Restore](#) feature, first released in DSpace 1.7.0.

Below you'll find more information pertaining to decisions made when developing this feature, etc.

Feel free to add your own notes!

Notes on Development

What code has really changed (as of DSpace 1.7)?

The majority of the code changes are in two main areas:

1. [org.dspace.content.packager.*](#) - Packager classes
 - `PackageIngestor` interface - Now ingests 'java.io.File' objects instead of `InputStreams` (to better support recursive imports of Communities/Collections)
 - `PackageDisseminator` interface - Now exports 'java.io.File' objects instead of `OutputStreams` (to better support recursive exports of Communities/Collections)
 - `DSpaceAIPDisseminator` - Disseminates/Exports AIP(s)
 - `DSpaceAIPIngestor` - Ingests exported AIP(s)
 - Changes were also made to refactor / enhance the `AbstractMETSDisseminator`, `AbstractMETSIgestor`, and `METSManifest` classes
2. [org.dspace.content.crosswalk.*](#)
 - `AIPDIMCrosswalk` - Crosswalks DIM metadata for AIPs
 - `AIPTechMDCrosswalk` - Crosswalks METS TechMD sections for AIPs
 - There were also changes to the `MODSDisseminationCrosswalk` and `XSLTDisseminationCrosswalk` to support creating "Site" AIPs

For More Information

 For a full list of code changes (including patches) see: [AipCoreAPIChanges](#)

Warning For Developers

 Because of the changes to the `PackageIngestor` and `PackageDisseminator` interfaces, if you've created any local Packagers at your institution, those will need to be refactored.

To-Do List – What remains to be done!

Testing Special Cases during Restore/Replace

The below special cases need further testing, especially when performing a "Restore" or "Replace". Mostly, these are just notes for Tim (and other developers), to ensure that all these various "edge" cases can be restored properly (or perhaps not restored properly, if the decision is made that it needs not be restored).

As each special case is implemented, we can check off the item in the below list. Special cases which have been fully tested & implemented are marked with a . Feel free to add more special cases to this listing, if we missed anything.

Anything not marked with a  either is not working or has not yet been fully tested! If you test it and it works, let us know, so we can check it off the list!

Item Restoration/Replacement

Special Cases

- ✓ Restore existing Deposit License from AIP – i.e. do not add a new license (or change the license) during restore/replace
- ✓ Restore existing CC License(s)
- ✓ Restore item mappings to multiple collections (for items which are mapped to several collections)
- ✓ Restore withdrawal state
- ✓ Restore embargo state
- ✓ Restore permissions & roles (user/group permissions) on Items, Bundles & Bitstreams
- ✓ Restoring metadata in a custom Metadata Field (i.e. non-default "dc" field)
- ✓ Restoring metadata in a custom Metadata Schema (i.e. not "dc").
 - Note: Schema must be created manually, but after that, the fields will be auto-created and auto-restored.
- ✓ Restore item having no bitstreams and/or no bundles.
- Options to restore just metadata or just particular bitstreams/bundles?
 - ✓ Exists on export, but not yet on import.
- *Will not restore items which have not made it into the "archived" state.* In other words, at this time, there are no plans to restore items which are still in an approval workflow (WorkflowItems) or items which are unfinished submissions (Workspaceltems). Workspaceltems and WorkflowItems are never exported as AIPs.

Collection Restoration/Replacement

Special Cases

- ✓ Restore permissions & roles (user/group permissions) on Collections
 - ✓ Restore Workflow approval groups
- ✓ Restore Collection-specific license
- ✓ Restore Collection's Item Template?
- Restore Collection's content source info? (e.g. OAI-Harvesting Collections versus normal Collections)

Community Restoration/Replacement

Special Cases

- ✓ Restore permissions & roles (user/group permissions) on Communities

Admin UI work

As part of the [CurationTaskProposal](#) (led by Richard Rodgers & MIT), a new Curation Framework is in the works. This Curation Framework will have a Command Line interface initially. However, the goal for 1.7, is to also have Administrative UI tools which are able to kick off various "curation tools". Among these curation tools will be the ability to export/import AIPs via the Admin UI.

Notes on AIP ingest speed & improving it

Some very basic ingestion speed tests were performed on a set of 26 AIPs (which represented a Community containing a Collection containing 24 Items). These tests found that, by default, the parsing/ingest settings are currently **not** optimized for speed.

Here are the basic (non-scientific) results

- Default Settings (NO Validation): took about 11 seconds to ingest all 26 AIPs
- Turning validate On (`-o validate=true` flag) (but using external Schemas): took about 1 minute, 12 seconds to ingest all 26 AIPs
- Locally cached all schemas (with validation turned on): took about 12 seconds to ingest all 26 AIPs
 - You can locally cache all schemas by using the `met.sxsd.*` settings in `dspace.cfg`. See [AIP Backup and Restore#AIP Configurations To Improve Ingestion Speed while Validating](#) in the official [AIP Backup and Restore](#) documentation.

Discussion / Use Cases

Please add your own potential use cases or discussion topics

- [DuraCloud DSpace Interaction Notes](#) - Notes/Discussion on how DSpace and DuraCloud may need to interact more directly. This page is **specific** to DuraCloud Use Cases.
- [AIP Export Implementation Notes](#) - Notes/Discussion on this specific AIP Backup/Restore Implementation (not specific to DuraCloud).
- [MIT Use Cases](#) - Notes on defining common operations in a replication system.

Questions / Comments?

Questions or comments – either add them inline above, or contact [Tim Donohue](#)