

# Design - Import - Export

- Stakeholders
- Sprints
  - Sprint 1
- Sprinters
  - Sprint 3
- Sprinters
  - Sprint 4
- Sprinters

## Stakeholders

- Esmé Cowles
- Benjamin Armintor
- Michael Durbin
- Joshua Westgard
- Youn Noh
- Nick Ruest
- Michael J. Giarlo
- Jon Stroop
- Karen Estlund
- Jim Tuttle

## Sprints

### Sprint 1

## Sprinters

### Developers

- Esmé Cowles
- Benjamin Armintor
- Nick Ruest
- Michael Durbin

### Testing and Validation

- Michael Durbin
- Joshua Westgard
- Justin Simpson
- Youn Noh
- Yinlin Chen
- Nick Ruest
- Bethany Seeger

### Documentation

- Youn Noh
- Joshua Westgard
- Nick Ruest

### Sprint 3

## Sprinters

- Developers
  - Esmé Cowles
  - Nick Ruest
  - Jared Whiklo
  - Danny Bernstein ★

### Testing and Validation

- [Joshua Westgard](#)
- [Nick Ruest](#)
- [Kieran Etienne](#)
- [Adam Wead](#)
- [Justin Simpson](#)

#### Documentation

- [Joshua Westgard](#)
- [Nick Ruest](#)

## Sprint 4

### 2017-05 Import - Export Sprint 04 Meetings

## Sprinters

- Developers
  - [Danny Bernstein](#)
  - [Ben Pennell](#)
  - [Michael Durbin](#)
  - [Esmé Cowles](#)
  - [Joshua Westgard](#)
  - [Longshou Situ](#)
- Testing and Documentation
  - [Kieran Etienne](#)
  - [Harsha Ummerpillai](#)
  - [Youn Noh](#)
  - [Jim Coble](#)
  - [Bridget Almas](#)

#### Use cases

1. Transfer between Fedora and external preservation systems, such as APTTrust, MetaArchive, LOCKSS, DPN, Archivematica, etc
2. ~~Package~~ [Export] the content of a single Fedora container and all its descendant resources
3. Transfer between fedora instances or (more generally) from Fedora to an LDP archive
4. ~~load~~ [Import] the contents of a package into a specified container.
5. Round-tripping resources in Fedora in support of backup/restore
  - a. A start has been made on this in [FCREPO-1990](#);
  - b. The implementation referenced in the above ticket is not dead, though not actively being worked on at the moment; pull requests welcomed (though others may well wish to take it in a different direction).
  - c. A rebuilder that:
    - i. Is not solely dependent on a intact backup of the repository index
    - ii. Works off shredded serializations that can be supported with file preservation techniques
    - iii. Can recover as much as possible of a repository in the face of integrity issues (supports partial recovery)
    - iv. Supports gathering copies of the shreds (serializations) from multiple sources to recover a repository
6. Round-tripping resources in Fedora in support of Fedora repository version upgrades
7. ~~Batch loading arbitrary sets of resources from metadata spreadsheet and binaries (may well be difficult — or not worth it — to try to generalize such a feature).~~
8. Import or export containers or binaries using add, overwrite, or delete operations. Configure the data model and the source and the target for each resource that will be updated. Allow target containers to be non-empty before import and source containers to be non-empty after export. Maintain ordering, etc. Support versioning. Examples: add issues to a publication; add fragments to a manuscript; add data sets to a longitudinal study; add time-series images from telescopes; remove resources determined to be under copyright; release resources after restrictions on access have expired.
  - a. Perform multiple metadata-only exports, and then restore an earlier version from an export.

#### Use cases yet to be rolled into requirements

1. Import objects from an external system (such as Figshare, where a research data object might be prepared) into a Fedora preservation repository with either Hydra or Islandora on top. (Implies compliance with Hydra and/or Islandora object models)
2. To migrate from internal content to external content, export metadata only and then import it into another repository. The links to the new external content locations would be added afterwards.

#### Requirements

#### External Systems

1. **PHASE 2** Support import from and export to a TBD list of external systems.

- a. APTrust - University of Maryland ([Joshua Westgard](#))
- b. Archivemata - Artefactual Systems ([Justin Simpson](#))
- c. MetaArchive - Penn State ([Ben Goldman](#))
- d. Perseids - Tufts - [Bridget Almas](#)

## General

1. PHASE 1 Support transacting in RDF
2. PHASE 1 Support allowing the option to include Binaries
3. PHASE 1 Support references from exported resources to other exported resources
4. PHASE 2 Support transacting in [BagIt](#) bags
5. PHASE 1 Support import into a non-existing Fedora container
6. PHASE 2 Support import into an existing, empty Fedora container
7. PHASE 3 Support import into an existing, non-empty Fedora container with various policies: add, overwrite, delete, version, skip
8. PHASE 3 Support export of resource versions
9. PHASE 3 Support import of resource versions
10. PHASE 1 Support export of resource and its "members" based on the `ldp:contains` predicate
11. PHASE 2 Support export of resource and its "members" based on a user-provided membership predicate
12. Support recursive RDF insert/updates with [LDP Indirect Container](#) specified POST (and PUT / PATCH?) (ref: [FCREPO 2042](#))

## Round-tripping

Defined as: Export all or a subset of a Fedora repository and importing the export artifacts into a Fedora repository.

1. PHASE 3 Support preservation of dates during round-tripping
2. PHASE 3 Support preservation of version snapshots during round-tripping
3. PHASE 1 The URIs of the round-tripped resources must be the same as the original URIs
4. PHASE 3 Support lossless round-tripping. (ie, if you export a resource, delete that resource and import there is no difference from if you had never performed any of those operations).

## BagIt

1. PHASE 2 Single resource bags
2. PHASE 2 The structure and scope of accepted and produced BagIt bags must be configurable ([resource](#))
  - a. Clarification: *structure* relates to required and optional tagfiles in the bag
  - b. Clarification: *scope* relates to contents of the bag, e.g. single object or object and all members based on specific membership predicate
3. PHASE 3 Multi-resource bags
4. PHASE 3 Unambiguously support linking between resources within a bag, and from resources in the bag to resources outside the bag
  - a. e.g. for bagged resources A and B, if A contains statement `<A> myns:rel <B>`, then it is unambiguous that B is a resource in the bag. Suppose some archive ingests the bag and exposes its contents as web resources with URIs P and Q. If the archive preserves intra-bag links, resource P will have statement `<P> myns:rel <Q>`. Likewise, if A contains external link `<A> myns:rel2 <http://example.org/outside/the/bag>`, then an archive that preserves links will have `<P> myns:rel2 <http://example.org/outside/the/bag>`

## Verification Tool

1. PHASE 2 Verify same number of resources on disk as in fcrepo
2. PHASE 2 Verify same number of resources in fcrepo as on disk
3. PHASE 2 Verify same checksum for binaries
4. PHASE 2 Verify same triples for containers
5. PHASE 2 Record which resources have been verified (Include checksum for binary resources)
6. PHASE 2 Verify subset of repository resources
7. PHASE 3 Verify fcrepo to fcrepo
8. PHASE 3 Verify disk to disk
9. PHASE 3 Use generated config file as sole input

## Considerations

- Import/export performance as is possible under the assumption that this work is done via the REST interface

## Resources

- <https://tools.ietf.org/html/draft-kunze-bagit-08>
- <https://github.com/ruebot/bagit-profiles>
- <https://github.com/barmintor/bagit-ldp>
- <https://www.ietf.org/archive/id/draft-wilper-semantic-content-pkgs-00.txt>
- <http://dataconservancy.github.io/dc-packaging-spec/dc-packaging-spec-1.0.html> (explanation [below](#))
- <https://github.com/acdha/restful-bagit-server> (a resource-oriented RESTful HTTP API for exchanging bags)

- [Import - Export Sprint Resources](#)

#### Meetings

- [2016-07-22 - Import - Export Planning Meeting](#)
- [2016-07-29 - Import - Export Planning Meeting](#)
- [2016-08-05 - Import - Export Planning Meeting](#)
- [2016-10-31 - Import - Export Planning Meeting](#)
- [2016-11-07 - Import - Export Planning Meeting](#)
- [2016-11-14 - Import - Export Planning Meeting](#)
- [2016-08 Import - Export Sprint 01 Meetings](#)
- [2016-12 Import - Export Sprint 03 Meetings](#)