2016-07-22 - Import - Export Planning Meeting

Time/Place

- Time: 3:00pm Eastern Time US
- Dial-in Number: (712) 775-7035
  - Participant Code: 479307#
  - International numbers: Conference Call Information
  - Web Access: https://www.freeconferencecallhd.com/wp-content/themes/responsive/flashphone/flash-phone.php

Attendees

- Andrew Woods
- Nick Ruest
- Michael J. Giarlo
- Joshua Westgard ⭐
- Youn Noh
- Karen Estlund (present?)
- Esmé Cowles
- David Wilcox
- Jon Stroop
- Benjamin Armintor (present?)
- Mark Jordan
- David Chandek-Stark (present?)
- Elliot Metsger (present?)
- Aaron Birkland
- Justin Simpson
- A. Soroka
- Michael Durbin
- Unknown User (acoburn)
- Audrey Templeton
- Don Richards

Related

- Design - Import - Export

Agenda

1. Introduction and topic summary
   a. Establish common understanding of function of Import/Export
2. Use case discussion
3. Review initial requirements
   a. Two flavours of import/export effort; RDF and binaries vs BagIt Bags
4. Confirm commitments
   a. Stakeholders
      i. Uses cases
      ii. Requirements
   b. Developers
   c. Testing and validation
   d. Documentation
5. Workplan and timelines
   a. Sprint scheduling

Minutes

1. Roll call of attendees (~16 persons were on the call).
2. Goals of this meeting:
   a. Shared understanding of what we would like to achieve (specifically in 2 week sprint) – probably it is going to be a subset of what falls into the scope of import/export;
   b. Look at the calendar to plan when the feature sprint might be scheduled.
3. Andrew Woods commented on the two flavors of import/export:
   a. Raw RDF serialized in some way on disk, plus the binaries or files associated with those objects;
   b. Bagit bags that could be imported/exported (round tripped) in/out of a Fedora repository.
4. Andrew Woods is hosting this initial call; Nick Ruest will lead the effort/sprint going forward
5. Nick Ruest led a walk through of the use cases and requirements on the design page:
   a. Use Case #1: Transfer between Fedora and external preservation systems:
      i. there was general agreement that round-tripping was implied in 'transfer between';

ii. discussion of whether content should be assumed to be in Bagit bags (often yes, but not universally so, as for example in LOCKSS);
        iii. after some discussion, suggestion was made (and there was general agreement) to leave the use case as is and let the particulars be refined later.
    b. Use Case #2: Export the content of a single Fedora container and all of its descendants:
        i. the basic idea is to have the ability to define a subset of resources (by default it seems logical that this be via LDP:contains relationships);
        ii. hypothetically one might wish to be able to configure the property that determines containment (for example according to PCDM relationships).
    c. Use Case #3: Transfer between fedora instances or (more generally) from Fedora to another LDP instance (e.g. Apache Marmotta):
        i. this use case was intended to help flesh out whether the transfer was always going to be in/out of the same repository, or whether it could be more generally conceived as a means of transferring between LDP repositories (i.e. not only for backup /restore or deposit to a dark archive, but also useful in the context of a transfer of stewardship from one institution to another).
    d. Use Case #4: Import into a specified, existing container:
        i. one problem here is that there could be very different ideas about what to do if there are conflicts during the import;
        ii. lossless round-tripping is not currently possible (not least because creation dates cannot be set), though it was noted that the barriers to lossless round-tripping are rooted in policy decisions, and those policies could be changed;
        iii. the issue of versioning was raised: the versioning in Fedora 3 would allow a new version to be created even if a resource already exists – it would become a new datastream version of the resource; it was noted that while versioning exists in F4, it is not the same as the datastream-based versioning of F3;
        iv. the issue of adding resources to an existing export was raised; an additional use case will be created that captures this scenario.
    e. Use Case #5: Round-tripping resources in Fedora in support of backup/restore.
    f. Use Case #6: Round-tripping resources in Fedora in support of Fedora repository version upgrades.
6. Due to a lack of time, use cases 5-6 and the list of specific requirements were not discussed.
7. It was agreed that a second planning meeting would be held in one week at the same time.
8. Andrew Woods raised one final consideration for planning this work: namely, that a development sprint is planned for September 19-23 at Penn State, and one of the items on the agenda for this sprint is import/export. This sprint would thus be a good opportunity to fix bugs or add additional features to an initial implementation, assuming it is possible to carry forward the planning, scheduling, and completion of an initial development sprint in the interim.
9. The question of the language used for the implementation was discussed. It was agreed that that's an implementation decision that should be driven by the developers who are available to work on the sprint, as well as considering the long-term maintainability of the resulting code.