

Use Cases for Change Management of Authoritative Data

Charter	Project Page	Terminology	References	Working Documents	Deliverables	Meeting Notes
-------------------------	------------------------------	-----------------------------	----------------------------	-----------------------------------	------------------------------	-------------------------------

- [Use Cases](#)
 - [Cache of Full Download](#)
 - [Cache of Labels in an Application](#)
 - [Activity Streams \(submitted by Jesse Lambertson\)](#)

Use Cases

Cache of Full Download

The application takes a full download of the authoritative data and stores it in a local cache. Access to the authoritative data is made through the cache.

Options for update:

- Re-download and cache the entire set of authoritative data. This does not identify what has changed.
 - Incremental downloads and update of the cache of just the data that has changed.
-

Cache of Labels in an Application

When the application makes a connection to authoritative data, the label is cached along with the URI to allow for quick display of the label in the UI.

Options for update:

- Periodic query of the authoritative data source to determine if the label has changed.
- Watch an incremental feed of changes to the data and update the cached data for any that are cached in the application.
- How to handle outliers that do not fit the expected types of changes?
- How to handle a change that no longer applies? Ex. When MeSH changes, even if the labels are similar, the old label is deprecated (gone) and a replacement label is created. In the past, identifiers were reused, which complicates the process.

On the provider side:

Entire Authoritative record in MARC records

- caching label and URIs in MARC records; want to keep up to date and in sync; caching entire authority record in sync

Receiving data on New Entities and Deleted Entities

Cache of data in Sinopia; What happens with conversions from BibFrame to MARC and vice versa.

- Data sync questions
- Workflow
- Heading toward data in triples instead of MARC
- Data integrity and data consistency

Local Data

- creating data locally until it can be included in the authoritative data
- in some cases, this is necessary
 - taking structured data from institutions and converting to RDF
 - creating URIs for manuscripts; want to link to other authoritative data when possible

- many are trying to move away from this

Question:

- What makes data authoritative?
 - When would the data be considered authoritative?
 - When is it considered local?
- Would it become authoritative data if an entity moves from local use to a shared location like ShareVDE?
- When additional local data is added to existing authoritative data and collected in a way that adds meaning, is this authoritative data?
 - Any work done outside of upstream, there needs to be agreement on what this means and how it will be applied upstream.

Activity Streams (submitted by [Jesse Lambertson](#))

Library of Congress is working on Activity Streams to possibly replace the ATOM feed they are currently using to notify of changes to the vocabularies.

Caveat: Activity Streams as we are talking about them here were originally the result of a W3 Social Web Working Group, which closed down its efforts in 2018. The idea was to track changes to feeds, sources, persons, etc, and to publish those changes in some way that can be used by machines and humans. In this case, we are looking at how an Activity Stream can be used to let users (machines and humans) know when updates (inserts, deletes, etc) to the vocabulary.

Activity Streams Examples: https://www.w3.org/wiki/Activity_Streams/Examples

- These examples reveal the significant complexity and variation within the tool we call 'Activity Streams'
- Many of the Attributes are the same from example to example, say the CORE Attributes, and some from the same term list.
- There are a number of extensions that have been implemented differently by the different institutions/corporations - not all of which have been continued in the same way.

Activity Streams Types: <https://www.w3.org/TR/activitystreams-vocabulary/#activity-types>

I (Jesse) reached out to the Getty, hearing they had implemented Activity Streams in their workflows. I got a mini-reply from them because they are busy.

[Received from them on 7 June 2021](#)

"Thanks for the message. We are content experts and do not deal with this, but I asked the tech team (one person) and here is the reply. I hope that's enough to help. Sorry, we have a tiny staff and pressing deadlines, thus cannot spare the time to discuss your specific questions. Good luck!"

We have implemented a Vocab activity stream in the Linked.Art version of the data and is based on the revision history in VCS but in a condensed form:

<https://data.getty.edu/vocab/activity-stream>

```
|| {"@context":"https://www.w3.org/ns/activitystreams","summary":"Getty Vocabularies Data (Production Environment)","type":"OrderedCollection","id":"https://data.getty.edu/vocab/activity-stream","totalItems":2913245,"first":{"id":"https://data.getty.edu/vocab/activity-stream/page/1","type":"OrderedCollectionPage"},"last":{"id":"https://data.getty.edu/vocab/activity-stream/page/29133","type":"OrderedCollectionPage"}} ||
```

"It is updated every month with the publish.

"Specifically, we have implemented the activity stream as an OrderedCollection as defined by the spec: <https://www.w3.org/TR/activitystreams-core/#collections>"

This realization shows us that there are real-life examples for implementing in libraries/cultural heritage institutions.

There will need to be some kind of variation for each workflow, platform, or the like, and in this case, we note the differences above, but we also know that the diversity and context differences for libraries and linked data will provide ample opportunities for variation. Does this mean that we need a fully defined and standalone ontology for implementing Activity Streams? Well, no, because we already have the example from the Getty above that uses standard elements from AS.

Plus, we have some information from Library of Congress ala Kevin Ford that says this (via e-mail 21 June 2021):

"I anticipate creating a custom activity type or two, but we've not done that yet."

*A number of the existing activity types are just fine for our purposes: Create/Add, Delete/Remove, and Update. However, for example, we want to publish an activity stream that contains minimally a list of resources whose authoritative label changed. Think along the lines of someone adding a death date to a name authority record. I think we will want to indicate that the *type* of change was something like "LabelChange" and for that we'd need a custom activity type because I don't think any of the existing ones will suit. But we've not yet attempted to figure what that 'type' will be or what its URI will be. Personally, I think that could be a useful outcome of this group."*

What this actual change or creation will need to be, is still to be realized, but whatever we come up with will follow the standard protocols as already delimited in the AS pages.

Does this mean we need an all new ontology? I am not sure yet, or if will be a derivation on a theme?

Time will tell
