

UW Madison PREMIS metadata strategy

Decisions for implementation: Fedora objects

Note: when these questions have been answered and vetted, this will be compiled into a profile/policy document.

- Our metadata standard will be PREMIS 3.0.
schemaLocation <https://digital.library.wisc.edu/1711.dl/XMLSchema-PREMIS-3.0>
- All Fedora objects will be PREMIS `intellectualEntity` objects; all content datastreams will be PREMIS `file` objects
- The PREMIS document for a given object will include both the `intellectualEntity` object and any `file` objects (datastreams) **directly** attached to the object: it will **not** include metadata for any child objects.
- Identifiers for PREMIS file objects will be based on the handle for the Fedora object, with the datastream ID (including version) appended as a fragment identifier.
- We will use UUID to create `rights` and `event` identifiers.
- We will link from `object` to `event` and `rights` elements.
- All PREMIS objects will log creation and deletion events; only file objects will have events over time (e.g., fixity checks, format migration), which will be recorded elsewhere.
- The `compositionLevel "0"` applies to every `file` object we have so far.
- We will not track Fedora internal datastreams or this PREMIS document itself, though they will be preserved along with all the other content that comprises the object.
- We will use `formatRegistry` when we're able to do so, e.g., using DROID. It may not be available in the first round of migration, in which case we should reprocess the objects when we can to add `formatRegistry`.
- A new PREMIS (`file`) object and creation event will be recorded for each new version (update) of a Fedora datastream.

Encoding Preservation Levels

We originally included a number of `premis:preservationLevel` elements in PREMIS, but quickly decided that we were going overboard with values that were likely to change over time and would be difficult to maintain. But some things, like uniqueness or preservation obligations, are likely to be stable over time.

It might be good to capture notions like e.g. "this is the only copy of this content, and we made a commitment to the Grad School to preserve it forever", or even "...the submitted file failed PDF validation, so we may not be able to fulfill our commitment at some point in the future". This will be useful in the future, in case diminished resources make it necessary to prioritize content for full preservation, or if we need to prioritize expensive operations like format migration. For initial preservation, we will not attempt to include this kind of supplementary information.

Definitions

preservationLevelValue: How important is it that we preserve this? applied to *intellectualEntity*, local values we will use are "High", "Medium", and "Low".

preservationLevelRole: What is the nature of our commitment to preserve? LOC values are "required" (we have to), "intent" (we want to), "capability" (we can).

preservationLevelRationale: Why are we preserving it? Local values are: "legal", "institutional", "contractual", or "at risk".

Data points and PREMIS Usage

Possible data points to collect:

- Status of the item that may affect preservation decisions about it (*preservationLevelRationale*):
 - this is the unique copy of record (born digital or reformatted from no-longer-existing source) (e.g., ETD)
 - this is born digital; other copies may exist
 - this is reformatted from an analog source known to be unique or otherwise at risk (including content reformatted as part of an analog preservation policy)
 - this is reformatted from an analog source whose risk low or unknown
- Commitment or obligation to preserve an object (*preservationLevelRole*, *preservationLevelRationale*):
 - commitment to Grad School to preserve ETDs
 - commitment to grant funders to preserve content
 - statutory requirement to preserve
 - contract with campus entity to preserve content (campus preservation service)

The first set of categories may be needlessly specific: for preservation purposes, we only need to know that the content is at high risk, without needing to specify the reason. A born-digital audio file may be at the same risk as one reformatted from a deteriorating analog source, in which case it is not necessary to specify its source (or lack of one). In fact, a born-digital file (e.g., a commercial recording) may have very low risk compared to one reformatted from an at-risk format. Therefore, a plain risk scale (high, medium, low) will be sufficient, and we will only assign the value "high" for our objects, when it is known to be the case for *any* reason.

We will not encode other elements, such as *preservationLevelType*, directly on Fedora objects, but rather register them as default policies or properties of specific repositories.

Decisions

1. **preservationLevel** will be specified for all Fedora objects that have been evaluated.

Therefore, lack of a *preservationLevel* indicates that an object has not been evaluated for its preservation value. Parent objects, because they contain crucial, irreplaceable metadata or

content, will typically have the same preservation level as their child objects (though they are not required to). The `preservationLevel` element should be applied to the PREMIS `intellectualEntity` object, and will be assumed to apply to its constituent files (datastreams) as well.

2. The PREMIS schema requires a value for `preservationLevelValue`. We will use the values "High", "Medium", and "Low". We will develop a rubric to use for determining these values.

3. We will use `preservationLevelRole` *only* when there is a requirement or intent to preserve, using the values from the [LoC ontology](#). We will *not* use the value `capability` in our PREMIS documents.

```
<preservationLevelRole
  authorityURI="http://id.loc.gov/vocabulary/preservation/preservationLe
velRole"
  valueURI="http://id.loc.gov/vocabulary/preservation/preservationLevelR
ole/req">
  requirement
</preservationLevelRole>
```

4. When the `preservationLevelRole` is `requirement`, it *must* be accompanied by a value in `preservationLevelRationale` describing the nature of the requirement. `preservationLevelRationale` *may* be used when `preservationLevelRole="intent"`, if there are suitable values in the ontology. Since the PREMIS schema does not provide `@authorityURI` or `@valueURI` for `preservationLevelRationale`, we have developed [our own ontology](#), and use a URI as the content of the element.

5. The *date of the creation of these elements* will be encoded in `preservationLevelDateAssigned`.

6. We will not use `preservationLevelRole` or `preservationLevelRationale` for other circumstances, e.g. for content that seems to be at low risk or for which we have no specific commitment to preserve. However, we will still use `preservationLevel` with `preservationLevelValue "Low"` so we know the object has been evaluated.

Example

For an ETD:

```
<preservationLevel>
  <preservationLevelValue>High</preservationLevelValue>
  <preservationLevelRole
    authorityURI="http://id.loc.gov/vocabulary/preservation/pres
ervationLevelRole"
    valueURI="http://id.loc.gov/vocabulary/preservation/p
reservationLevelRole/req">
    requirement
```

```
</preservationLevelRole>
<preservationLevelRationale>
  https://digital.library.wisc.edu/1711.dl/vocabulary/preservation/presLevelRationale#institutional
</preservationLevelRationale>
<preservationLevelRationale>
  https://digital.library.wisc.edu/1711.dl/vocabulary/preservation/presLevelRationale#atRisk
</preservationLevelRationale>
<preservationLevelDateAssigned>2021-05-
13</preservationLevelDateAssigned>
</preservationLevel>
```

For an object we do not intend to preserve:

```
<preservationLevel>
  <preservationLevelValue>Low</preservationLevelValue>
</preservationLevel>
```