PREMIS PREservation Metadata Implementation Strategies

Pop-Up PREMIS Implementation Fair at iPres 2018

2018-09-24

Karin Bredenberg
Chair PREMIS Editorial Committee
National Archives of Sweden

PREMIS PREservation Metadata Implementation Strategies

WELCOME!

Agenda

- What have we been up to in the EC?
- Questions to the EC
- User presentations:
  - Aaron Elkiss
    - HathiTrust's implementation of PREMIS
    - Dark Blue
  - Mark Jordan
    - Concurrent migration from PREMIS 2 to PREMIS 3 and Islandora to CLAW
- Input time!
  - Brief results from the newly conducted survey
  - Discussion time

UPDATES FROM THE EC

ONTOLOGY

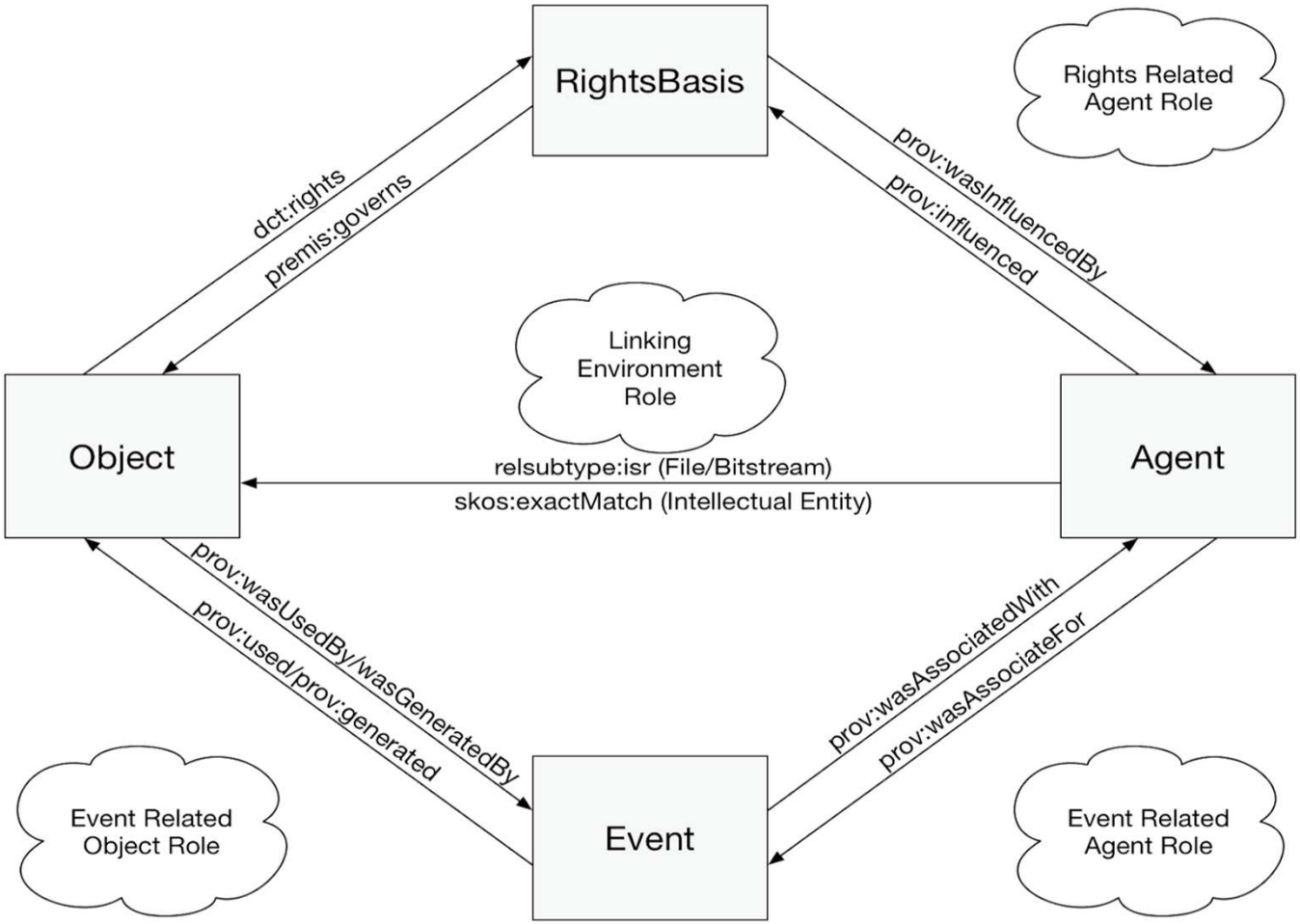
Revision of PREMIS ontology work

- PREMIS OWL Ontology Revision Working Group
  - Previous version based on version 2.2 and issued in June 2013
  - Revision based on PREMIS Data Dictionary Version 3.0 with its data model changes
  - Substantial remodeling of ontology taking into account current LD best practices
    - Draft released Dec. 2017 followed by comment period; final Sept. 2018
- Use cases
  - Facilitate interoperability between repositories and registries publishing or exchanging metadata about digital objects
  - Exchange digital preservation events from a preservation repository with other systems
  - Use in Linked Data/RDF-based repositories (e.g. SPAR at BnF, Fedora 4)
  - Enhance other ontologies and application profiles with rich preservation metadata

Principles

- Reuse existing elements from other ontologies or make relationships to them (e.g. skos:closeMatch)
  - Thorough review of applicable ontologies, e.g. PROV-O, ODRL, DC, PCDM
- Use Linked Data aware controlled vocabularies for enumerated lists, e.g. http://id.loc.gov/preservationdescriptions
  - Make relationships between vocabularies and PREMIS ontology
- Stay faithful to PREMIS Data Dictionary and model as much as possible
- Supporting documentation to make implementation easier (especially user guidelines)

PREMIS ontology data model: the big picture



Main changes to id.loc.gov preservation vocabularies

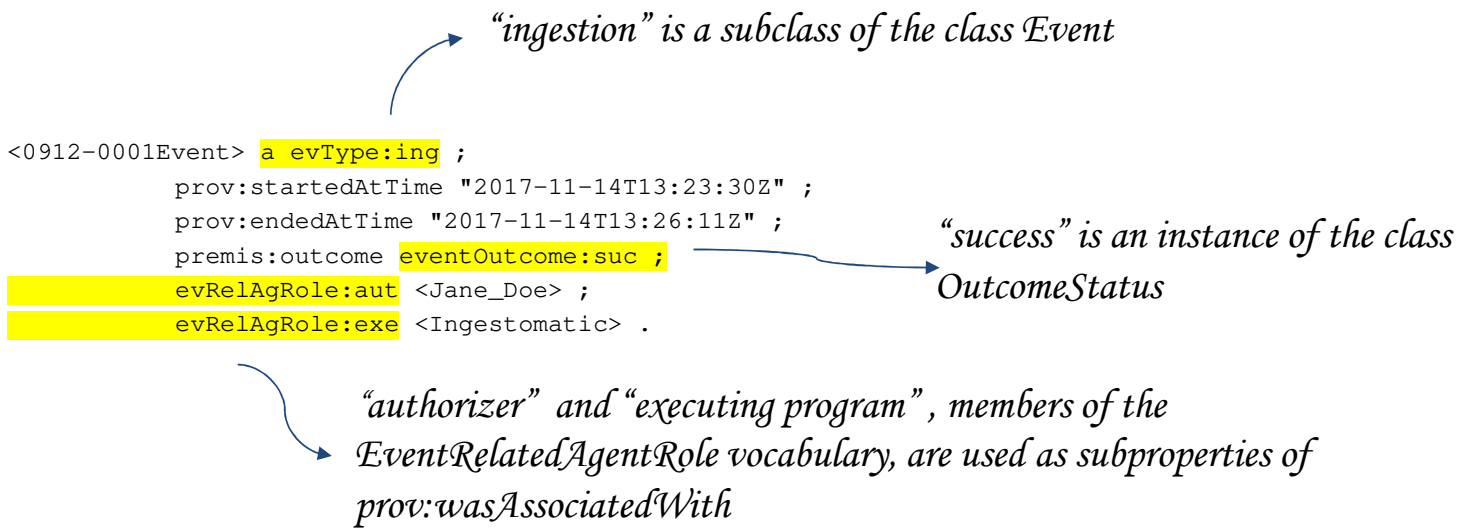
- Members of vocabulary may be subclass/subproperty of class/property in PREMIS OWL ontology, e.g.
  - eventType members are subclasses of premis:Event
  - relationshipSubType members are subproperties of premis:relationship
- Members of vocabulary may be individuals of a class, e.g.
  - environmentCharacteristic members are instances of premis:EnvironmentCharacteristic
- actionsGranted, environmentPurpose, inhibitorTarget merged into eventType vocabulary
  - Declared as individuals of premis:Action
  - Event types declared as subclasses of premis:Event

Using LOC preservation vocabularies

```
@prefix dct: <http://purl.org/dc/terms/> .
@prefix dce: <http://purl.org/dc/elements/1.1/> .
@prefix foaf: <http://xmlns.com/foaf/spec/> .
@prefix premis: <http://id.loc.gov/vocabulary/preservation> .
@prefix prov: <http://w3.org/ns/prov#> .
@prefix rdf: <http://www.w3.org/TR/rdf-schema/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix crypHashFunc: <http://id.loc.gov/vocabulary/preservation/cryptographicHashFunctions> .
@prefix evOutcome: <http://id.loc.gov/vocabulary/eventOutcome> .
@prefix evRelAgRole: <http://id.loc.gov/vocabulary/preservation/eventRelatedAgentRole> .
@prefix evType: <http://id.loc.gov/vocabulary/preservation/eventType> .
@prefix presLevType:  <http://id.loc.gov/vocabulary/preservation/preservationLevelType> .
@prefix presLevRole:  <http://id.loc.gov/vocabulary/preservation/preservationLevelRole> .
@prefix relSubType:  <http://id.loc.gov/vocabulary/preservation/relationshipSubType> .

<0912-0001Event> a evType:ing ;
            prov:startedAtTime "2017-11-14T13:23:30Z" ;
            prov:endedAtTime "2017-11-14T13:26:11Z" ;
            premis:outcome evOutcome:suc .
```

LOC vocabulary terms: subclasses, instances and subproperties



Future work on ontology

- Complete revision of preservation vocabularies
- Expand preservation vocabularies
  - Add new members as needed by the community
  - Investigate security classification actions for eventType (actions granted)
- Investigate using SHACL or ShEx for validation and to facilitate implementation
- PREMIS Editorial Committee to look at future PREMIS DD changes arising from ontology development

PREMIS PREservation Metadata Implementation Strategies

CONTROLLED VOCABULARIES

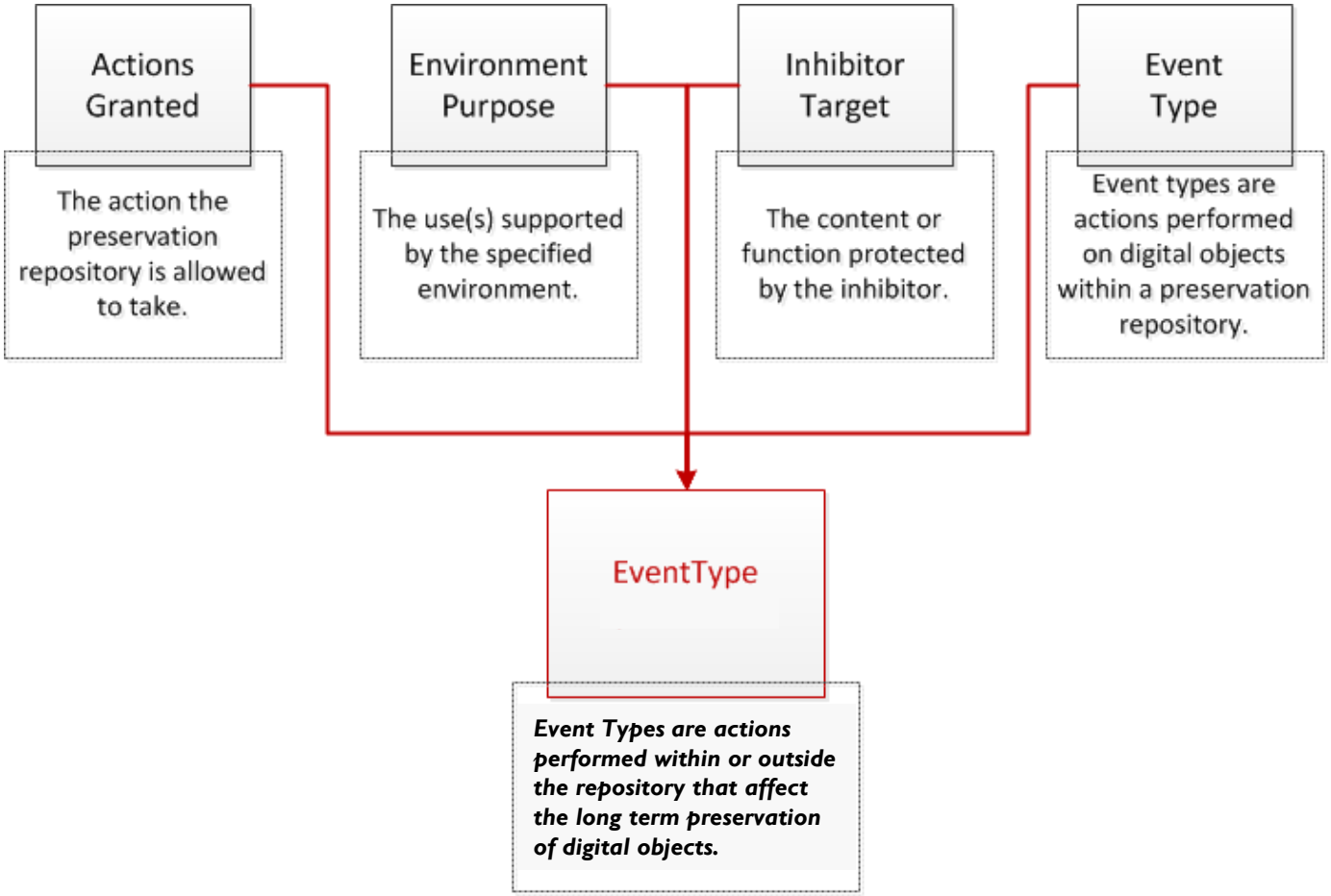
Current work on id.loc.gov

- PREMIS ontology work looking at vocabs.
- Recommendation to join vocabularies that were “action” or ”function” based.
- Some of the discussed vocabs also needed a tidy up.

As was

	Actions Granted	Environment Purpose	Inhibitor Target	Event Type
SCOPE	The action the preservation repository is allowed to take.	The use(s) supported by the specified environment.	The content or function protected by the inhibitor.	Event types are actions performed on digital objects within a preservation repository.
E.G.	delete disseminate migrate modify replicate use	compile create edit interpret render	all content play function print function	filename change fixity check forensic feature analysis format identification imaging information package creation

End result


Actions Granted
The action the preservation repository is allowed to take.
Environment Purpose
The use(s) supported by the specified environment.
Inhibitor Target
The content or function protected by the inhibitor.
Event Type
Event types are actions performed on digital objects within a preservation repository.
EventType
Event Types are actions performed within or outside the repository that affect the long term preservation of digital objects.

Rationale

- The rationalised list is function based
- No need for the vocab to distinguish between permission and prohibition (ActionsGranted InhibitorTarget)
- Vocab does not need to be relevant for *every semantic unit* it is related to.
- Tidy up and add to vocab.

This controlled vocabulary is now relevant for the following semantic units of the PREMIS Data Dictionary:

1.5.6.2 inhibitorTarget
1.13.5 relatedEnvironmentPurpose
2.2 eventType
4.1.7.1 act

URIs for old terms will point to new terms

QUESTIONS TO THE EC

PREMIS PREservation Metadata Implementation Strategies

USER PRESENTATIONS

PREMIS PREservation Metadata Implementation Strategies

IMPACT OF PREMIS SURVEY

What/Why/Who/etc..

After the dust settled on v3, we wanted to begin to look at the future.

We wanted to understand the current **impact** of PREMIS in the digital preservation community (and beyond) in order to help shape its **future**.

What we wanted to know

- How widely is PREMIS used?
- How is it being used to support digital preservation programmes?
- What can be done to improve PREMIS?
- Can we improve services
- How do we expand / reimagine the community
- Why is the listserv quiet? Have we done our job, or are there barriers to participation?

High level (draft!) results

- 65 respondents to date
- 84% maintain preservation MD
- 64% use some aspect of PREMIS

High level (draft!) results

- 7% use it for physical collection items
- 60% using v2, and 40% using v3 (1 org using v1)
- Most planning to move to v3

Q19: Why haven't you implemented PREMIS?

ANSWER CHOICES	RESPONSES	
It doesn't meet our needs	0.00%	0
It isn't implemented in our toolset	35.00%	7
It is too complicated	15.00%	3
Conformance requirements are too high	10.00%	2
We don't understand it well enough	10.00%	2
We aren't familiar with it	5.00%	1
Our digital preservation metadata solution predates PREMIS.	5.00%	1
Other:	0.00%	0
Other (please specify)	20.00%	4
TOTAL		20

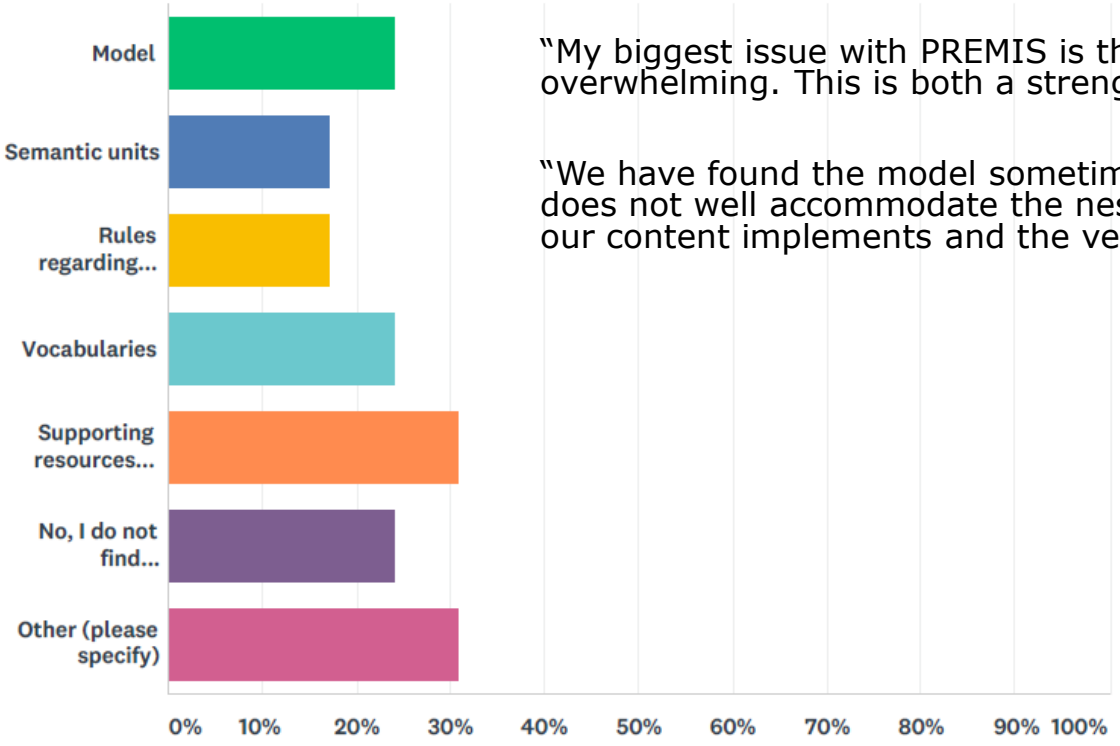
"We plan to adopt PREMIS, but cannot at this time. Currently building an access-focused, preservation-ready digital repository"

"We have so much to do that we can't implement more than the simplest vocabularies for our metadata."

"It has been seen as too verbose due to the size of the standard/misunderstanding that all parts of the model must be implemented.
`

Q21 Do you find limitations with any of the following features of PREMIS?

Answered: 29 Skipped: 36



"My biggest issue with PREMIS is that it is so overwhelming. This is both a strength and weakness."

"We have found the model sometimes limiting as it does not well accommodate the nested relationships our content implements and the versioning we do."

Q22. What support from the PREMIS Editorial Committee would help you in adopting PREMIS more easily

- More tutorials
- Specific examples of implementation
- Webinars

What we would like to know

- Your reflections on PREMIS
- What needs to be changed or updated to best meet your requirements?
- What can the Editorial Committee do to help you with making decisions about the implementation of PREMIS?
- What can the Editorial Committee do to aid implementation?