

Introduction to the PREMIS OWL ontology

Draft ontology for version 3.0 of the PREMIS Data
Dictionary

Webinar presentation, February 22, 2018

Outline of Webinar

- History of PREMIS OWL ontology
- Use cases for preservation metadata as Linked Data
- Principles followed
- An overview of the data model: top-level entities and relationships
- Reuse of other ontologies
- Integration with preservation controlled vocabularies
- Examples
- Community review

PREMIS ontology work

- PREMIS OWL Ontology
 - Initial draft done by Sam Coppens as PhD thesis in Belgium
 - Working group of PREMIS-EC reviewed and revised
 - Faithful to version 2.2 of Data Dictionary semantic units
 - Issued in June 2013
- Current revision
 - Based on PREMIS Data Dictionary Version 3.0 with its data model changes
 - Working group of PREMIS-EC formed
 - Substantial remodeling of ontology taking into account current LD best practices
 - Initial development of goals and principles by Rob Sanderson/Esmé Cowles
 - Analysis of other relevant ontologies, e.g. PROV-O, ODRL, PCDM, DC
 - Draft released Dec. 2017; comment period until late March 2018

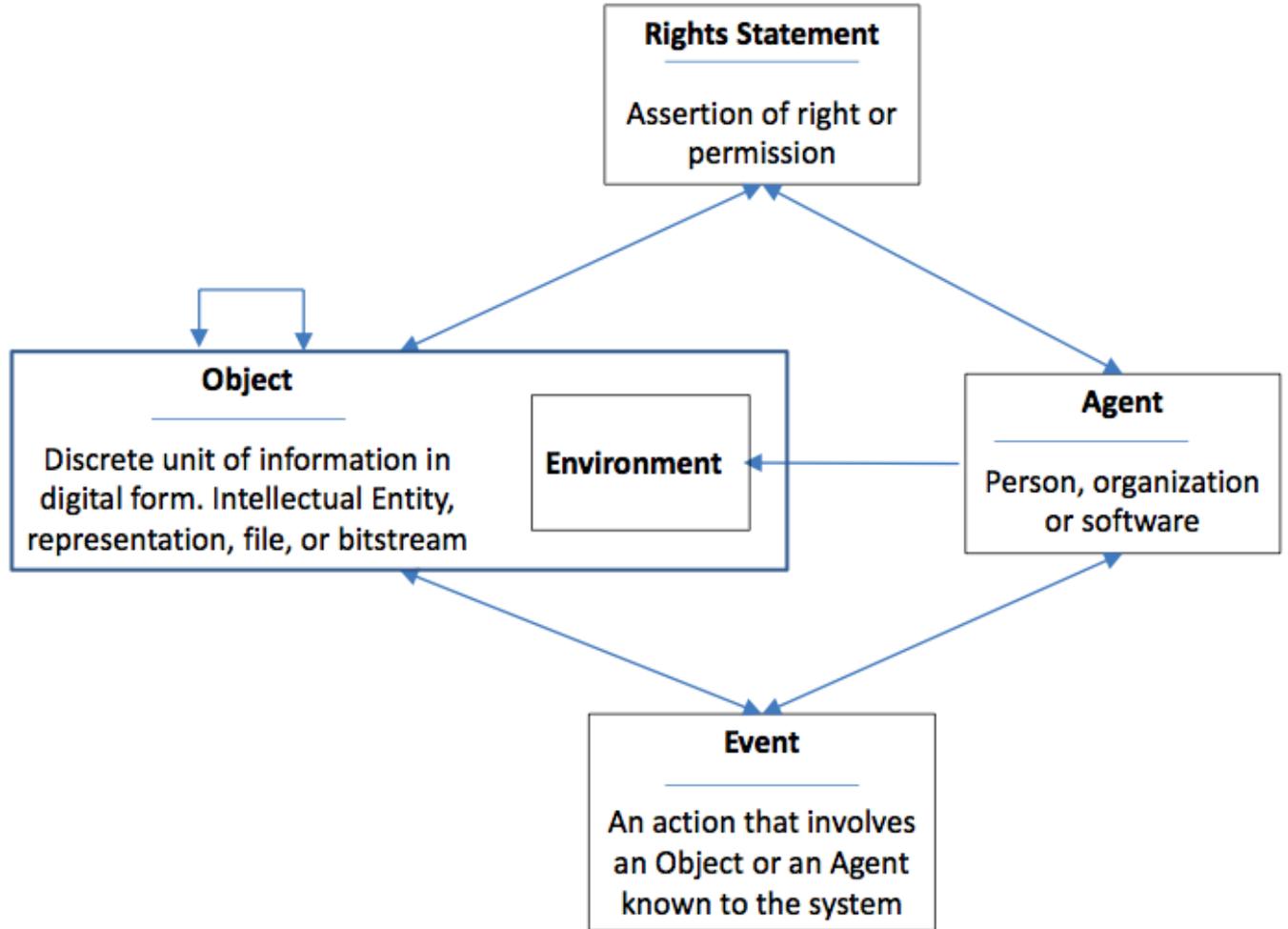
Use cases: PREMIS metadata as Linked Data

- 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

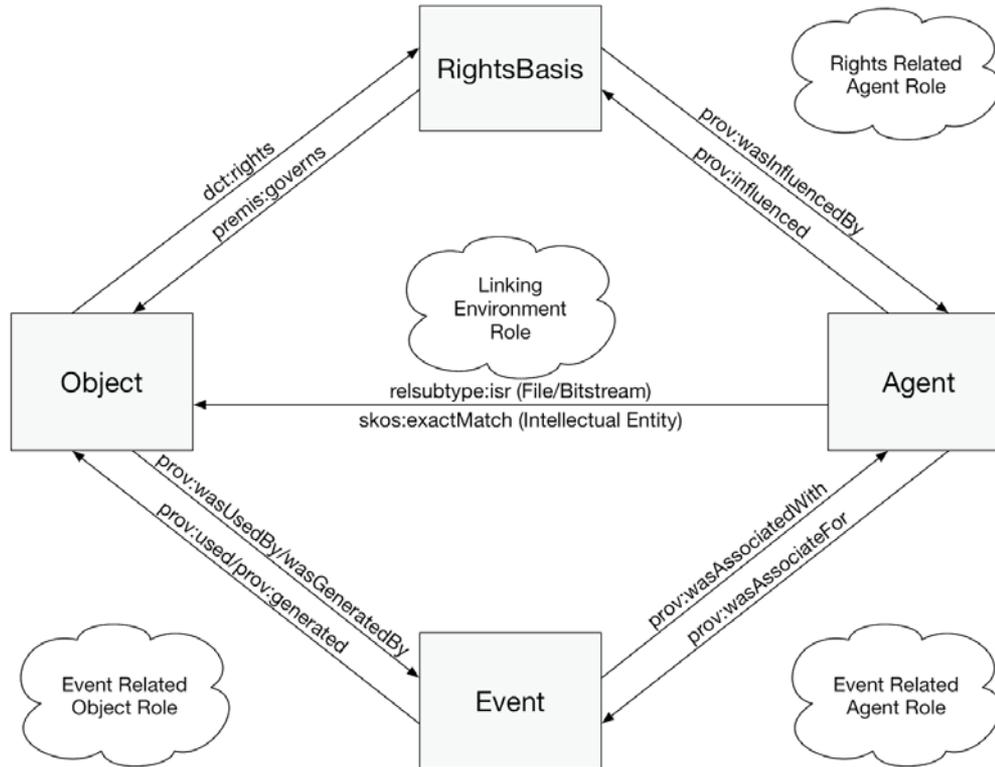
Principles

- Make the ontology as simple as possible
- Reuse existing elements from other ontologies where semantics agree
- Make relationships (equivalencies, hierarchical) with elements from other ontologies with appropriate RDF constructs
 - `skos:closeMatch`, `skos:exactMatch`
 - RDF/OWL subproperty/subclass relationships
- Use Linked Data aware controlled vocabularies for enumerated lists, e.g. <http://id.loc.gov/preservationdescriptions>
- Follow RDF rather than XML constructions
- Stay faithful to PREMIS Data Dictionary and model as much as possible

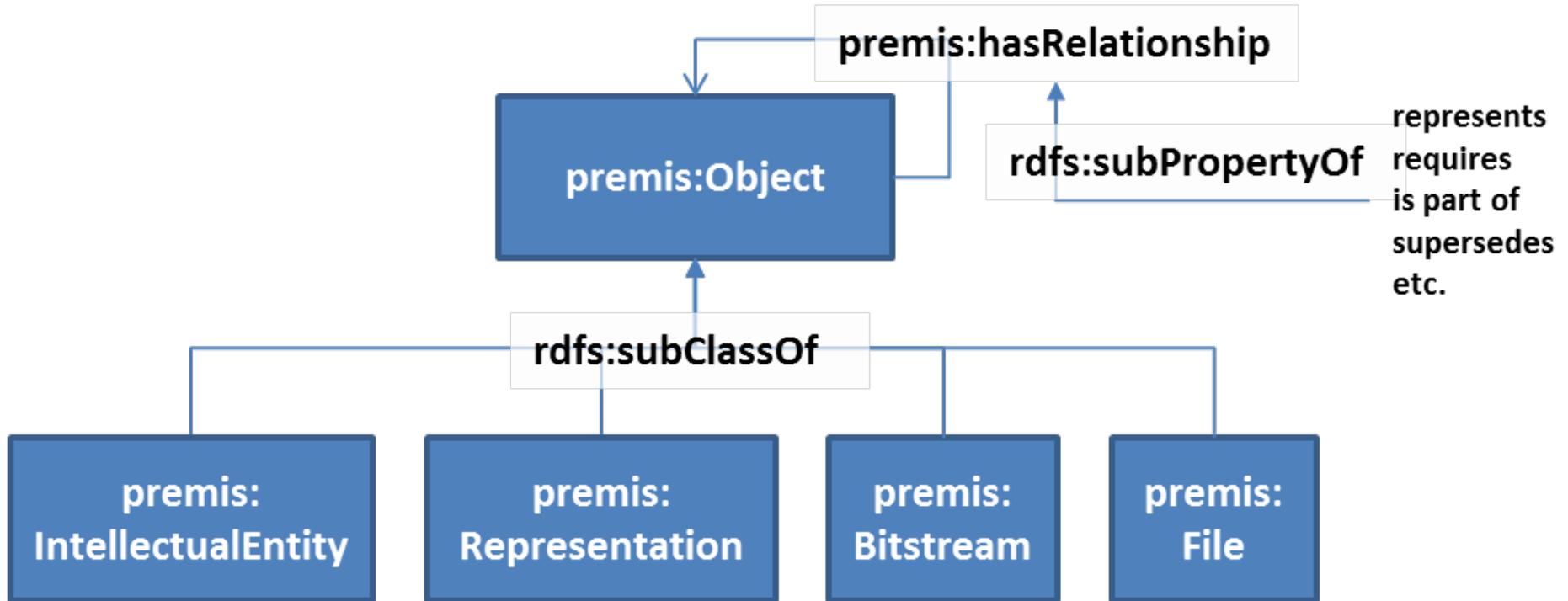
PREMIS Data Model version 3.0



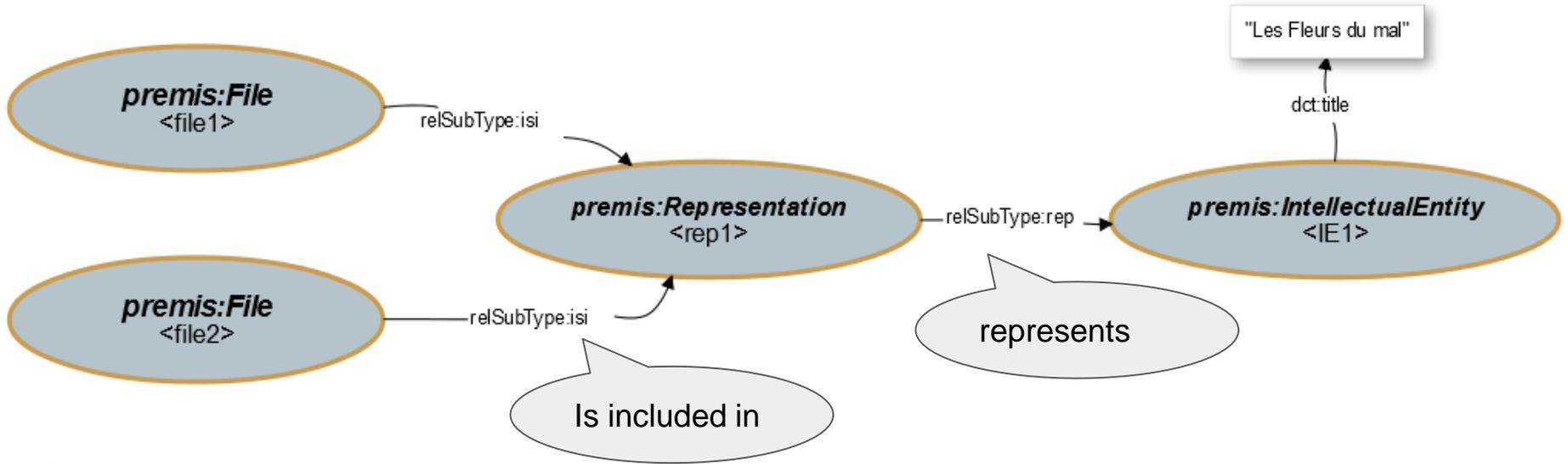
PREMIS ontology data model: the big picture



The Object Entity

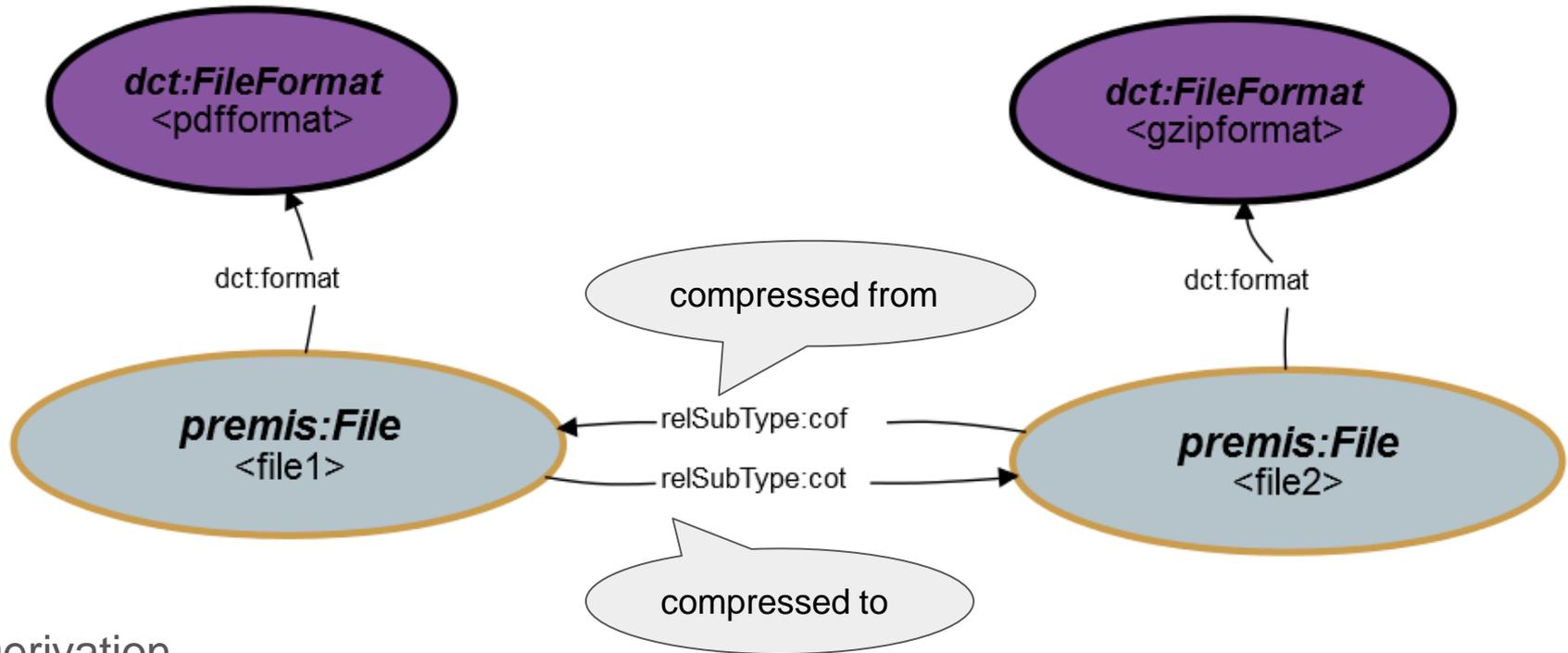


Relationships between Objects: examples



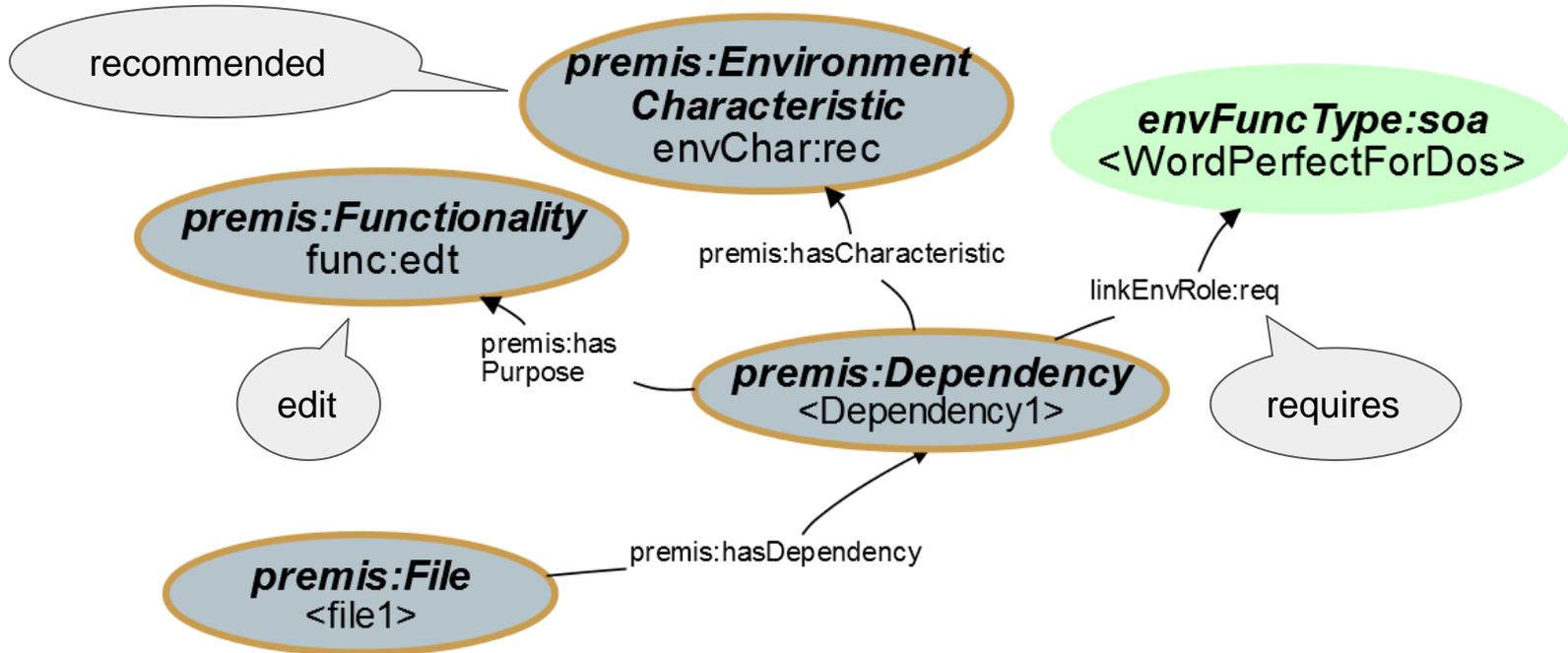
Structural

Relationships between Objects: examples

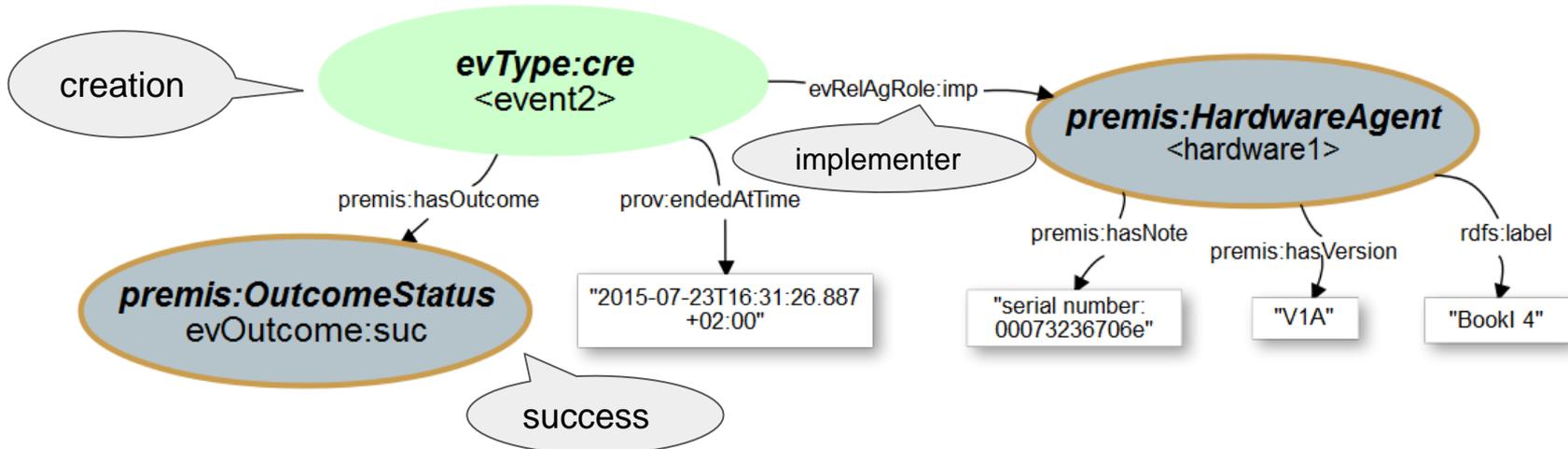
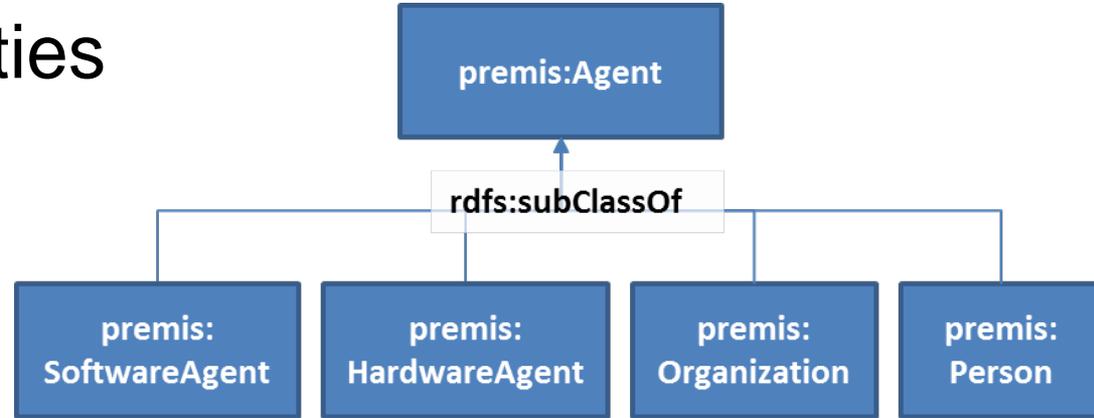
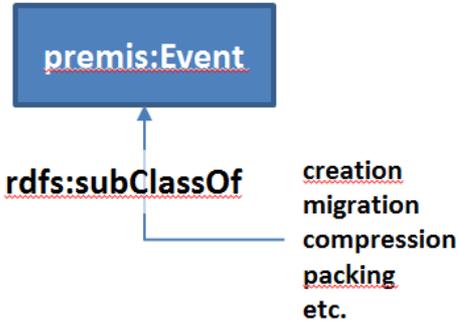


Derivation

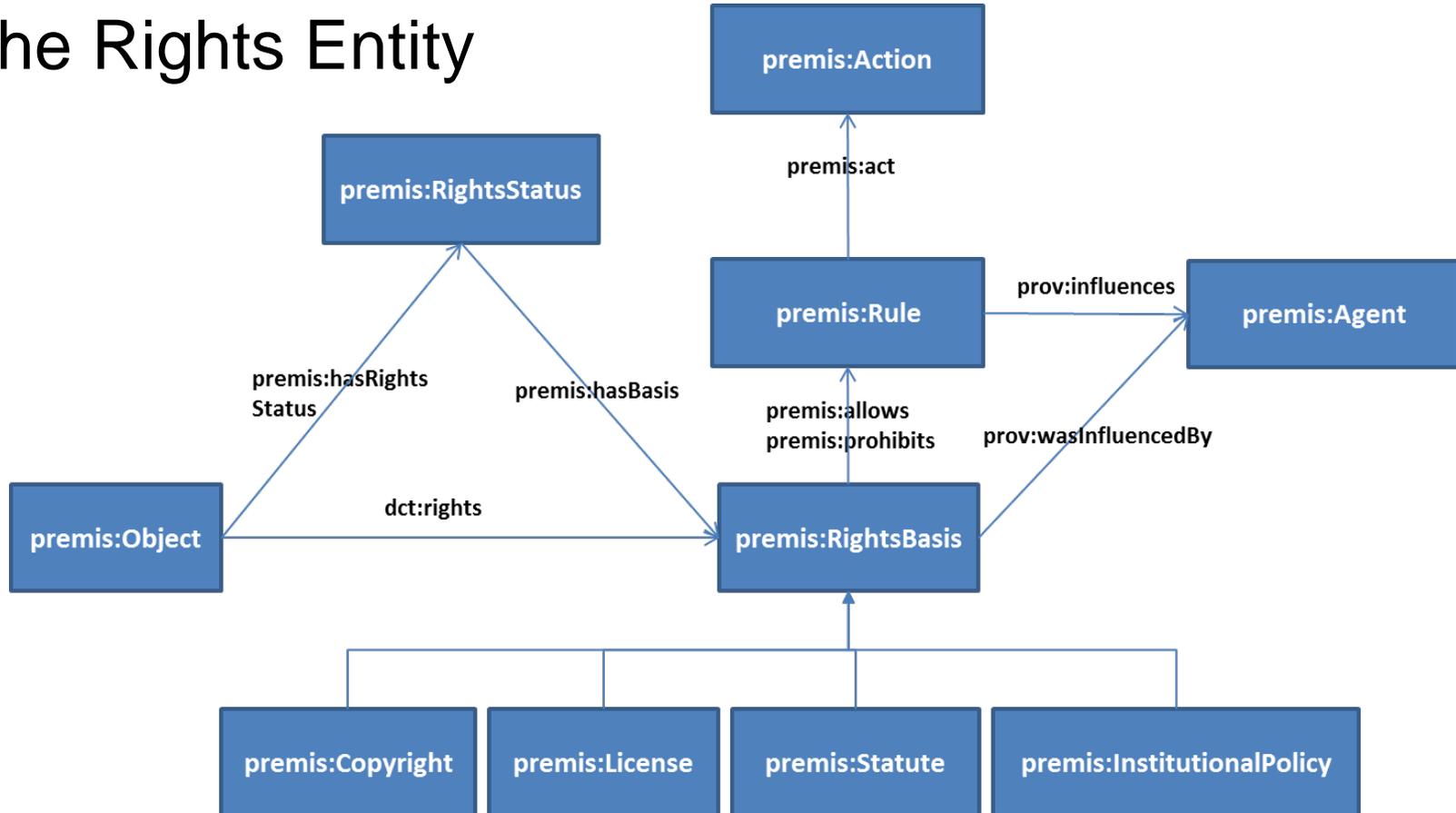
Relationships between Objects: examples



Event and Agent Entities



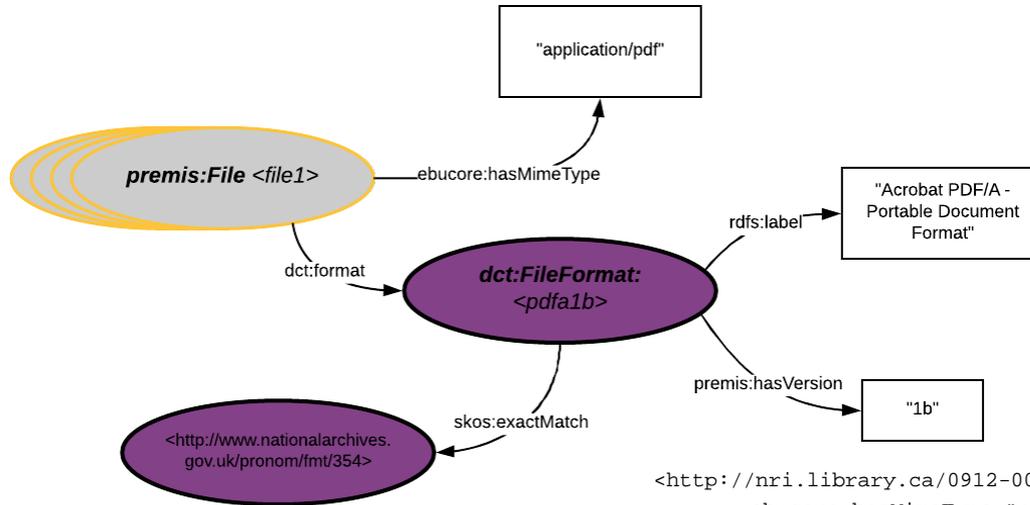
The Rights Entity



Reusing other ontologies

- Why reuse terms from other ontologies?
 - The essence of linked data is using common terms that have common meanings. No need to create a unique term for a concept that already exists, then maintain information on how the unique term maps to other ontologies.
 - The more a term is used across multiple ontologies, the more supported and sustainable it is.
 - The PREMIS ontology should be seen as a framework which can be extended by other ontologies if desired.
- What other ontologies did we look to?
 - Primarily dublin core and prov
 - Also rdfs, skos, foaf, edm, odrl, ebucore

Example: format

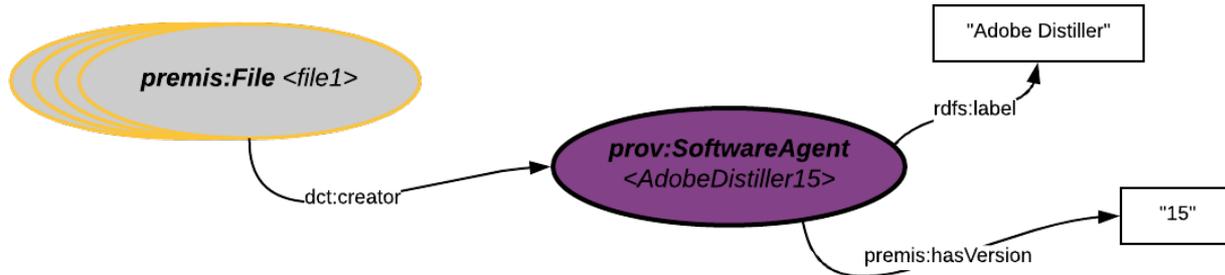


```
<http://nri.library.ca/0912-0001.pdf> a premis:File ;
  ebucore:hasMimeType "application/pdf" ;
  dct:format <http://nri.library.ca/pdfalbformat> .
```

```
<http://nri.library.ca/pdfalbformat> a dct:FileFormat ;
  rdfs:label "Acrobat PDF/A - Portable Document Format" ;
  premis:hasVersion "1b" ;
  skos:exactMatch
```

```
<http://www.nationalarchives.gov.uk/pronom/fmt/354> .
```

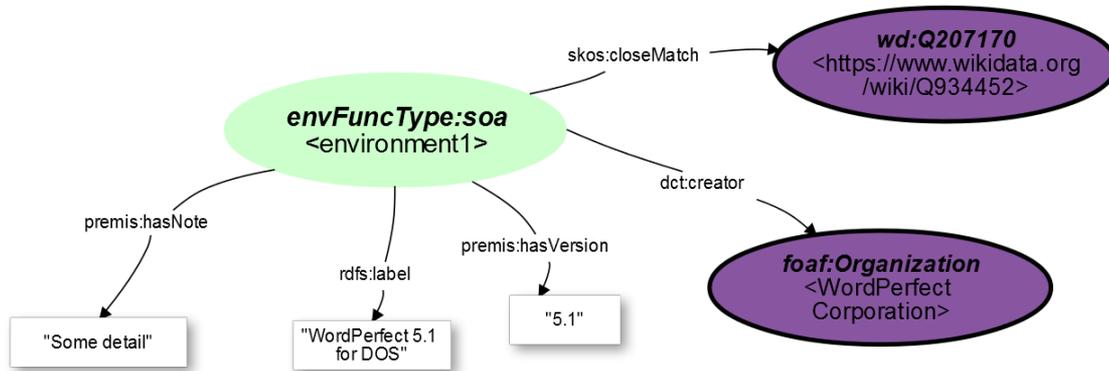
Example: creating application



```
<http://nri.library.ca/0912-0001.pdf> a premis:File ;  
  dct:creator <AdobeDistiller15> ;  
  prov:generatedAtTime "2016-05-10T14:23:30Z" .
```

```
<http://uri.adobedistiller15> a prov:SoftwareAgent ;  
  rdfs:label "Adobe Distiller" ;  
  premis:hasVersion "15.0" .
```

Example: a software application described as an environment



```
<environment1> a envFuncType:soa ;  
  rdfs:label "WordPerfect 5.1 for DOS" ;  
  premis:hasVersion "5.1" ;  
  premis:hasNote "Some detail" ;  
  dct:creator <WordPerfect Corporation> ;  
  skos:closeMatch <https://www.wikidata.org/wiki/Q934452> .
```

```
<WordPerfect Corporation> a foaf:Organization .
```

Using LOC preservation vocabularies

The screenshot shows the Library of Congress website interface. At the top, there is a navigation bar with the Library of Congress logo, buttons for "ASK A LIBRARIAN", "DIGITAL COLLECTIONS", and "LIBRARY CATALOGS", and a search bar with the text "Search Loc.gov" and a "GO" button. Below the navigation bar, the breadcrumb "The Library of Congress > Linked Data Service" is visible.

The main content area is titled "Event Type". It features a search box with the placeholder text "Enter Keyword or Phrase". Below the search box, a dropdown menu is open, showing a list of terms: "Environment Function Type", "Environment Purpose", "Environment Registry Role", "Event Related Agent Role", "Event Related Object Role", and "Event Type". The "Event Type" term is highlighted in orange. To the right of the search box, there is a note: "**Please Note: LC Classification entries are not included in general search results. You must explicitly select LC Classification in order to search the scheme. This is temporary while the impact of adding LCC to the current system is better understood." Below the note are "Search" and "Reset" buttons.

Below the search area, there is a "Details" tab. The "Event Type" details section includes the following information:

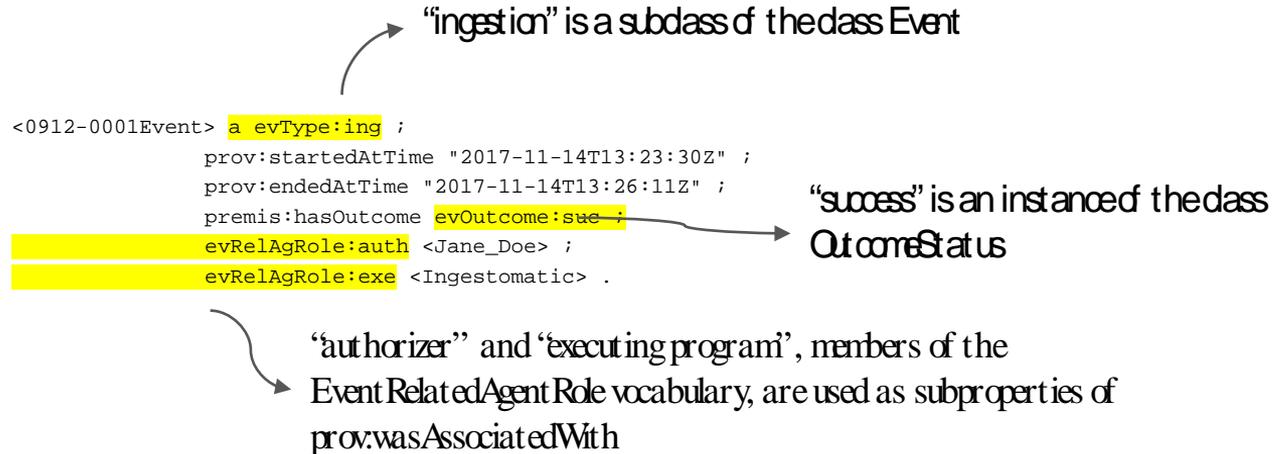
- Event types are actions performed on digital objects within a preservation repository.
- URI(s)
 - > <http://id.loc.gov/vocabulary/preservation/eventType>
- Instance Of
 - > [MADS:RDF-MADSScheme](#)
 - > [SKOS:ConceptScheme](#)
- Scheme Members
 - > [accession](#)
 - > [appraisal](#)
 - > [capture](#)
 - > [compression](#)
 - > [creation](#)
 - > [deaccession](#)
 - > [decompression](#)
 - > [decryption](#)
 - > [deletion](#)
 - > [digital signature generation](#)

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/preservation/evOutcome> .
@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:hasOutcome evOutcome:suc .
```

LOC vocabulary terms: subclasses, instances and subproperties



Changes to the LOC vocabularies

- Writing the ontology has prompted us to review the existing vocabularies and propose some changes
 - In some cases it is a matter of declaring members of an existing vocabulary sub-classes of a class, or making them sub-properties of a property. Some examples:
 - Terms in the **Cryptographic Hash Functions** vocabulary need to be named sub-classes of the class **Fixity**.
 - Terms in the **Agent Type** vocabulary need to be named sub-classes of the class **Agent**.
 - Terms in the **Relationship Subtype** and **Linking Environment Role** vocabularies need to be named sub-properties of **premis:hasRelationship**.

Changes to the LOC vocabularies

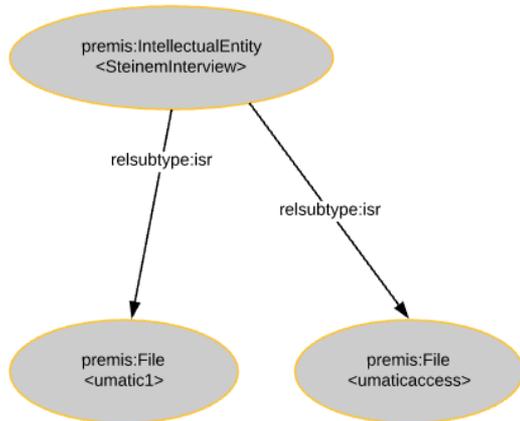
- For others, new concepts require a reconsideration of some of the vocabularies. Some examples:
 - We propose creating a new **Event Outcome** vocabulary, to capture outcomes such as “success” and “warning” as URIs instead of plain text.
 - We propose deprecating the **Environment Purpose, Hardware Type, Software Type and Inhibitor Target** vocabularies and merging their terms into a new, more generic **Functionality** vocabulary. We also propose adding some new terms such as “characterize” and “identify”, derived from new terms in the **Event Type** vocabulary.
 - We are reviewing the possible overlap between the **Actions Granted, Event Type** and newly-proposed **Functionality** vocabularies.
- All of this work is ongoing and we’d love to have your input.

Putting it all together: some examples

A simple example: Digitized video

Use case:

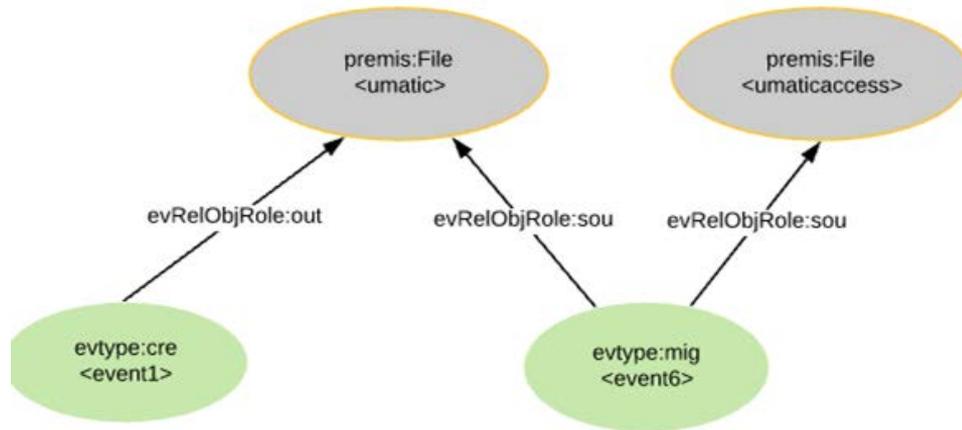
Emory University holds the Alice Walker papers, which includes a umatic videotape of Gloria Steinem interviewing Alice Walker in 1984. For preservation reasons, it has been digitized into a .mov file and an .mp4 access copy has been generated. Both files have been ingested into the repository. An access copy of the video is available in the reading room only.



```
<WalkerPapers> a premis:IntellectualEntity ;  
relSubtype:hsp <SteinemInterview> .
```

```
<SteinemInterview> a premis:IntellectualEntity ;  
relSubtype:isr <umatic1>, <umaticaccess> .
```

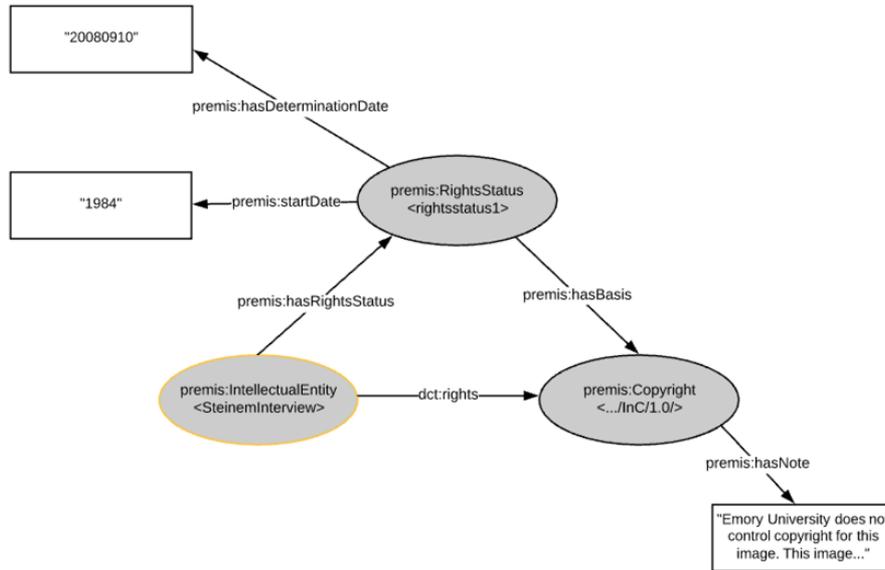
A simple example: Digitized video



```
<event1> a evtype:cre ;  
  dct:date "2017-08-22T16:48:07.487Z" ;  
  premis:hasOutcome evOutcome:suc ;  
  premis:hasOutcomeNote ".mov successfully created." ;  
  evRelAgRole:imp <tbunn> ;  
  evRelAgRole:exe <finalcutpro> ;  
  evRelObjRole:out <umatic1> .
```

```
<event6> a evtype:mig ;  
  prov:startedAtTime "2017-08-25T00:07:20.656946" ;  
  prov:endedAtTime "2017-08-25T08:50:56.487Z" ;  
  premis:hasOutcome evOutcome:suc ;  
  premis:hasOutcomeNote ".mp4 successfully created" ;  
  evRelAgRole:exe <finalcutpro> ;  
  evRelObjRole:sou <umatic1> ;  
  evRelObjRole:out <umaticAccess> .
```

A simple example: Digitized video



```

<SteinemInterview> a premis:IntellectualEntity ;
relSubtype:isr <umaticl>, <umaticAccess> ;
dct:rights <http://rightsstatements.org/vocab/InC/1.0/> ;
premis:hasRightsStatus [
    a premis:RightsStatus ;
    premis:startDate "1984" ;
    premis:hasDeterminationDate
    "20080910" ] .
  
```

```

http://rightsstatements.org/vocab/InC/1.0/ a premis:Copyright ;
    premis:hasNote "Emory University does not control
copyright for this image. This image is made available for
individual viewing and reference for educational purposes only,
such as personal study, preparation for teaching, and research.
Your reproduction, distribution, public display or other re-use of
any content beyond a fair use as codified in section 107 of US
Copyright Law or other applicable privilege is at your own risk.
It is your sole responsibility to investigate the copyright status
of an image and obtain permission when needed. We are always
interested in learning more about our collections. If you have
information regarding this photograph, please contact
marbl@emory.edu." .
  
```

A complex example: Archival Disk Image

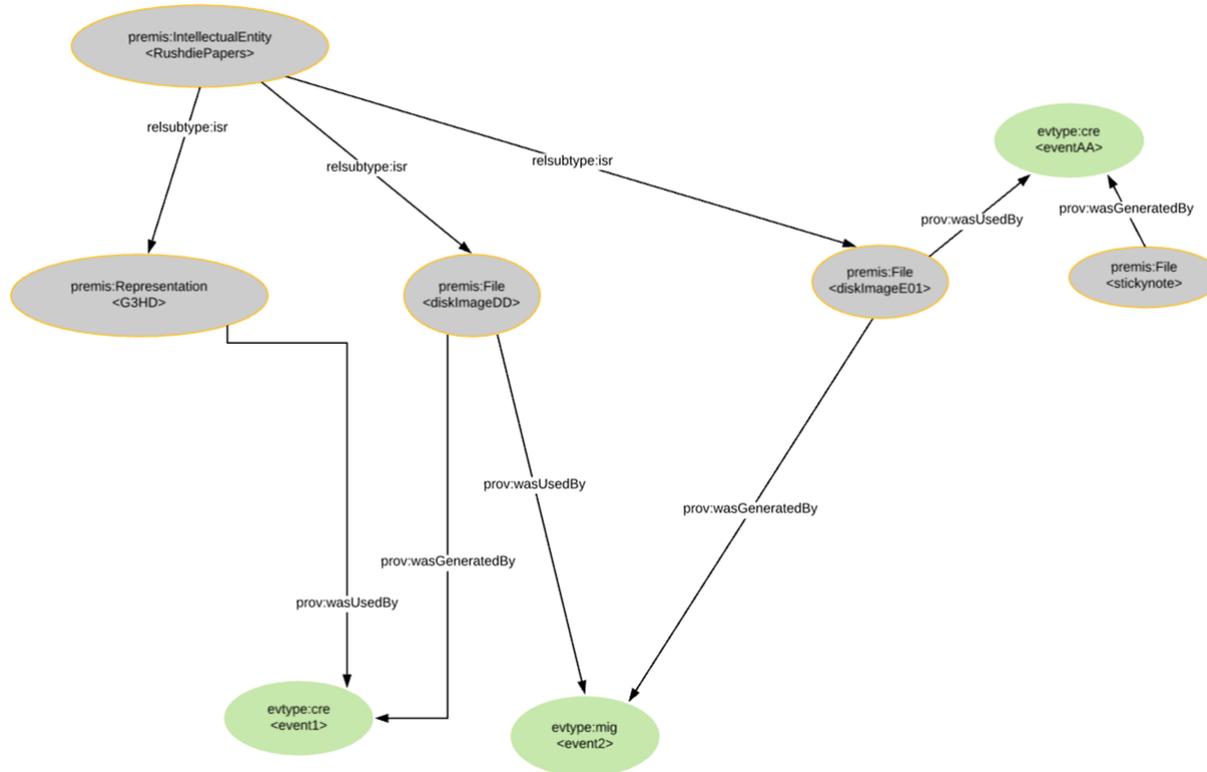
Use case:

Emory University acquired a laptop as a part of the Rushdie papers. Imaging of the laptop drive was outsourced to a vendor who used GNU ddrescue 1.19. Following receipt at Emory, a digital archivist reviewed the image, verified the MD5 checksum, and calculated a SHA1 checksum. Ewfacquire was used to migrate the disk image from .dd to .E01 in preparation for ingest into Emory's digital repository. Individual files (including a Mac sticky note) were extracted from the image using FTK Imager and ingested as separate files. All ingested files undergo validation, a virus check, and checksum generation as a part of the ingest process. Users may view files in a emulated environment, which uses SheepShaver 2.3 on a Mac Powerbook. This emulated environment has been imaged as well and ingested into the repository for preservation purposes. The disk images are restricted to the public but the sticky note (an example of an extracted file) is available for reading room use only per the deed of gift.

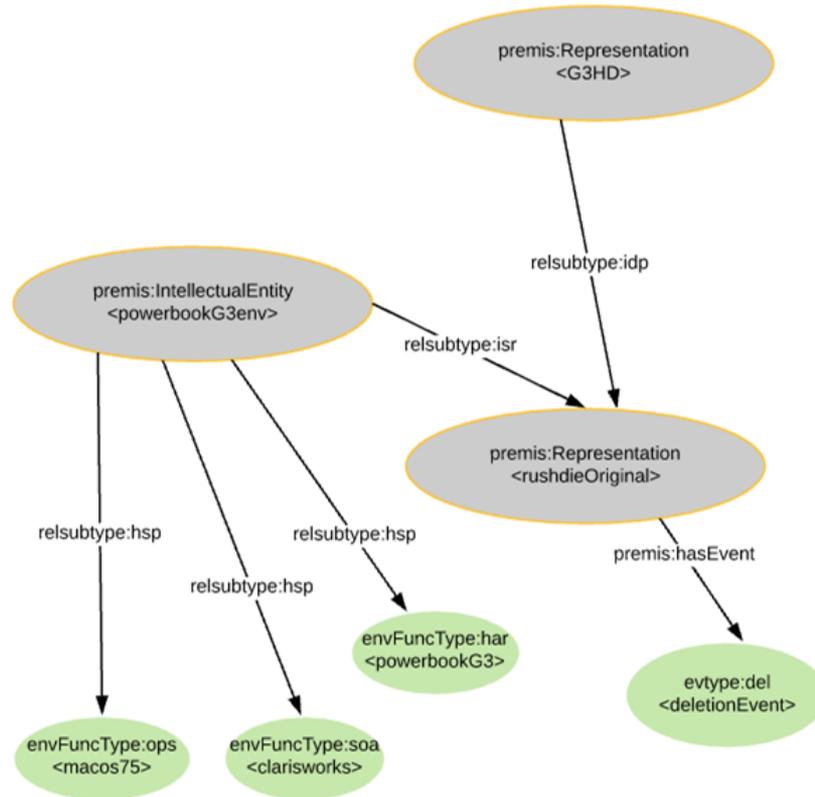
What we want to document/preserve:

- Disk images (.dd, .e01)
- Original environment of the physical hard drive
- Emulated environment
- Lifecycle of the intellectual object

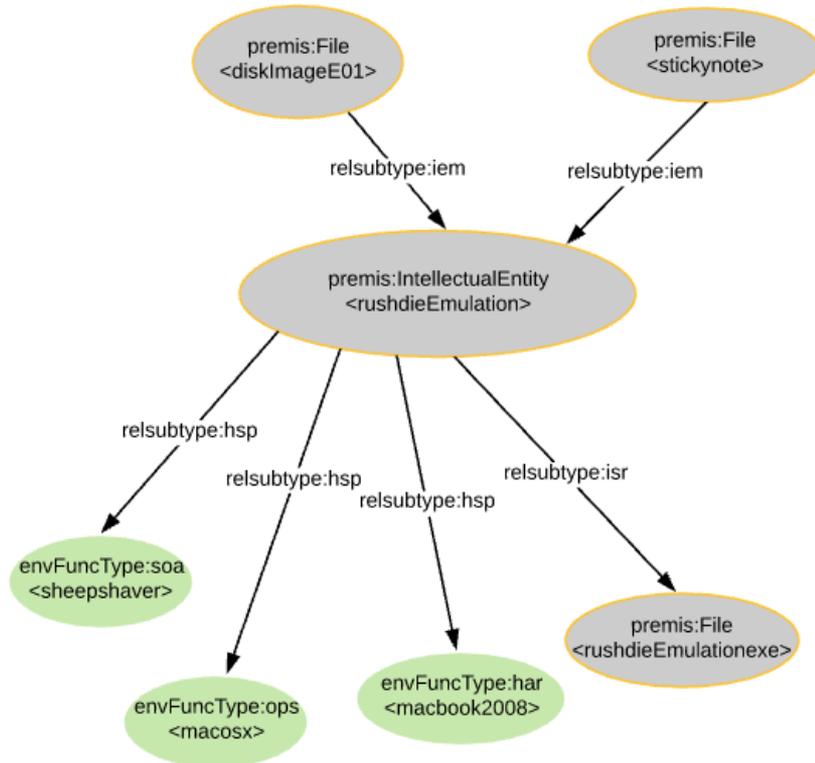
A complex example: Archival Disk Image



A complex example: Archival Disk Image



A complex example: Archival Disk Image



Review process

- The WG encourages people in the preservation, metadata and Linked Data communities to review and provide comments.
- The review period will end on **March 23, 2018**.
- Documents are available at: <http://www.loc.gov/standards/premis/ontology/owl-version3.html>.
- Comments may be sent to: premis-ontology-review-2018@googlegroups.com and/or you can join the conversation at <https://groups.google.com/forum/#!forum/premis-ontology-review-2018>
- Community members can read and respond to other comments as well as post their own.

Questions?

Also may be sent later to the link on the previous slide.