

To the W3C PROV working group:

This document summarizes my proposed changes and additions, which are based on my (potentially inaccurate) understanding of the PROV data model, dated July 24, 2012. Please see the accompanying document for a list of my detailed comments (which are also referenced below).

The most obvious change is the proposed addition of Role classes. In my experience, distinguishing between Entities and the Roles that they may play is a useful exercise that can more clearly define the concepts within a domain and improve the internal consistency of a model. In this case, adding Role classes substantially improved the clarity of the model for me, especially with respect to Entity Generation, Usage, and Invalidation. Furthermore, it helped me to identify relationships that were not modeled as rigorously within the PROV spec: Alternation, Specialization, Membership, and Adoption.

These proposals are illustrated in a number of diagrams that accompany this document. Those diagrams are summarized here:

- Logical Model – the main model, showing nearly all classes and relationships
- Entities – a summary of the defined entities
- Relationships – these diagrams show the specified relationships
  - Communication Initiation Termination
  - Association Adoption
  - Attribution Delegation
  - Derivation Interaction
  - Alternation Specialization Membership
  - Influence
- Relations – a summary of all relationships
- Roles – a summary of the role "helper classes" that were used to more explicitly define the specific semantics of each relationship

In the attached diagrams, attribute multiplicity is include after the attribute name and abstract classes are shown in italics. All other notations should be the same as were used in the published PROV data model. The color scheme used in the model is as follows:

- Yellow background – modeled relationships (as classes)
- Green outline – Entity classes
- Blue outline – Role classes
- Red outline – Activity classes

Please note that the answers to some of the questions in section 5.3 (see detailed comments) may require changes to the proposals below.

Regards,

Robert Freimuth, Ph.D.  
Mayo Clinic

- Major changes
  - Person, Software, and Organization were made specializations of Entity rather than Agent (see comments for section 5.3.1).

- Defined Agent as a Role that Entities can assume in a given context (see comments for section 5.3.1).
- Changed Primary Source from a type of Derivation to a role that an Entity can have in the Derivation of a new Entity (see comments for section 5.2.4).
- Minor changes
  - Added the attributes to Bundle and Collection that are inherited from Entity (see comments for sections 5.4.1 and 5.6.1).
  - Added the attributes to Specialization and Alternation that are inherited from Influence (see comments for sections 5.5.1 and 5.5.2).
  - Added start and end times to Association and Delegation to allow users to specify when these relationships took place (see comments for sections 5.3.3 and 5.3.4).
  - Defined Bundle as a type of Collection (see comments for sections 2.2.2 and 5.4.1).
  - Restricted Delegation to Agents that are Entities (not Activities) (see comments for section 5.3.4).
  - Defined a common Trigger class that can initiate or terminate Activities, which helps streamline the attributes required for those relationships and makes the models for the relationships more similar to each other. The relationships now capture the identifier of the triggering activity, while the trigger subclass (TriggeringActivity or TriggeringEntity) captures the type of trigger.
    - I am unclear whether this design poses an implementation issue using the current notation. If it does, a similar issue will need to be resolved for Association.
  - Renamed the "start" and "end" relationships to Initiation and Termination, respectively, to align more closely with the style used for other relationships.
  - Renamed the attribute "usedEntity" within the Derivation relationship to "sourceEntity" (this is a minor personal preference).
- Additions
  - Role classes were introduced to the model to help readers distinguish between activities and the modeled relationships (see comments for Table 2).
  - Defined EntityGeneration, EntityUsage, and EntityInvalidation as specializations of Activity to make the distinction explicit between those activities and the modeled relationships (see comments for sections 2.1.1, 5.1.3, 5.1.4, and 5.1.8).
    - This clarifies the difference between the act of creating an entity vs. expressing a relationship between an entity and the activity that created it.
    - Start and end time attributes were added to the Generation, Usage, and Invalidation relationships in place of the single "time" attribute to capture more explicitly the times related to those activities.
  - New class associations were added to support relationships that were explicitly or implicitly defined within the spec.
    - Specialization, Alternation, and Membership were defined according to the information in sections 5.5.1, 5.5.2, and 5.6.2.
    - Adoption was defined according to section 2.2.1.2. This will also allow users to specify that an Agent Adopted a Plan independently from its Association to an Activity.
    - ActivityComposition was added to support "activity wasPartOf activity" (see comments for section 2.1). This is a distinct concept from Communication, which is more temporal in nature.