IVOA PROVENANCE METADATA

CURRENT STATUS

Mireille Louys

Kristin Riebe, MathieuServillat, Michèle Sanguillon, François Bonnarel, Markus Nullmeier

Working draft update

- Last working draft issued :

- [WD-ProvenanceDM-v1.0-20170512.pdf](#) at

[http://wiki.ivoa.net/twiki/bin/view/IVOA/ObservationProvenanceDataModel](#)

- Modelio 3.6 model available
  - To be completed for documenting each class/ attributes/relation

- VODML-Provenance.xml representation
  - Improved
  - Still a few problems in understanding the design constraints UML→ VOML

Draft Structure proposed

- A logical path :
  - Use cases
  - Domain model : the abstract shape in UML 2.0
  - Conception model : the UML translation in VOML
  - Description of classes and relations
  - How to use the model : discussion section
    - Experience in adopting the model for various use-cases
  - DB tables declarations examples
  - Queries examples and their response in various formats : JSON, VOTable, PROV-N
  - VOTools for Provenance : VOPROV-Python package
    - Serialisation in all formats W3C and VOTable

Domain Model - Provenance - 2017 May11
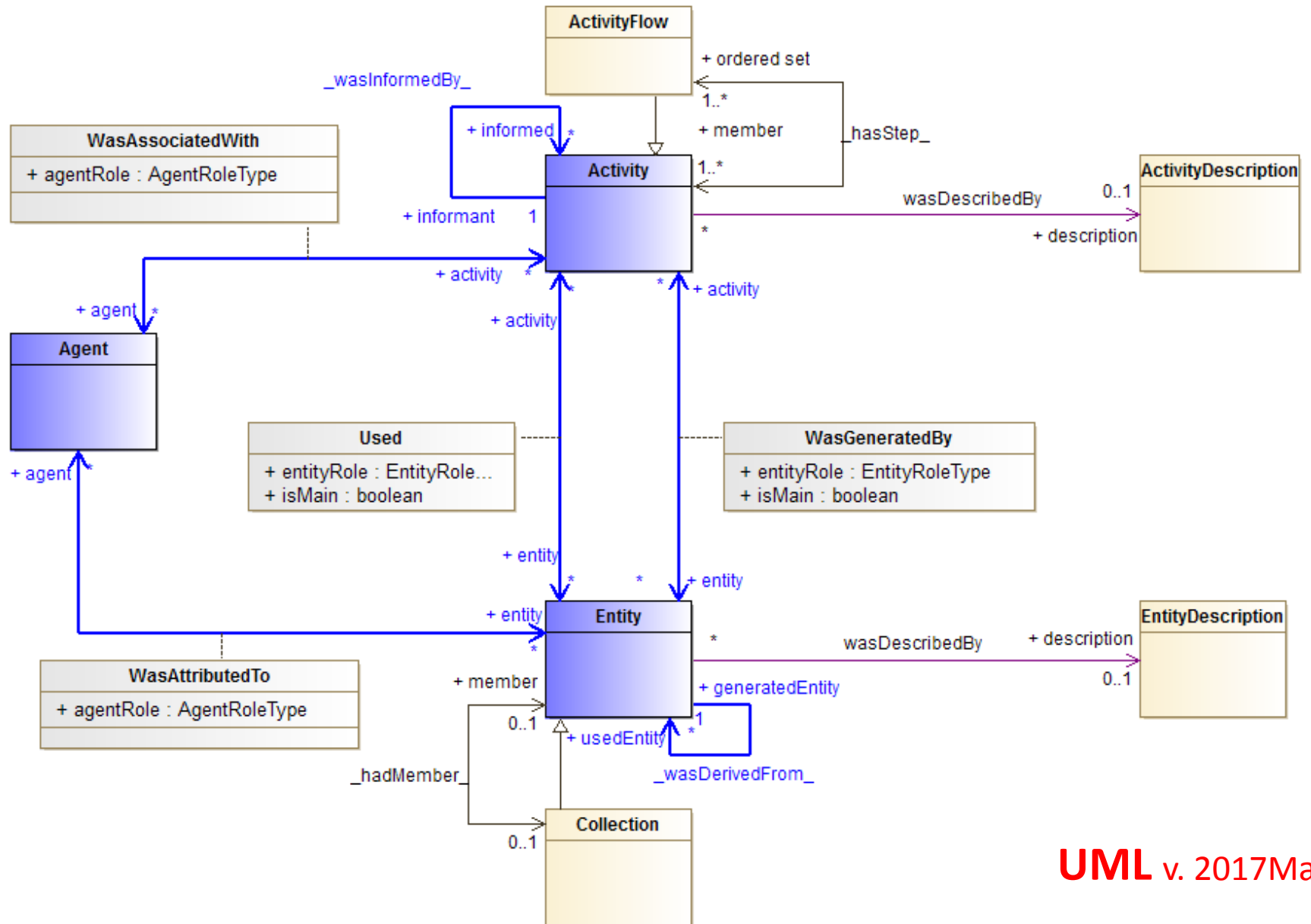
Conception Model - Provenance

Searching compliancy for VODML ingestion
DB implementation Focus
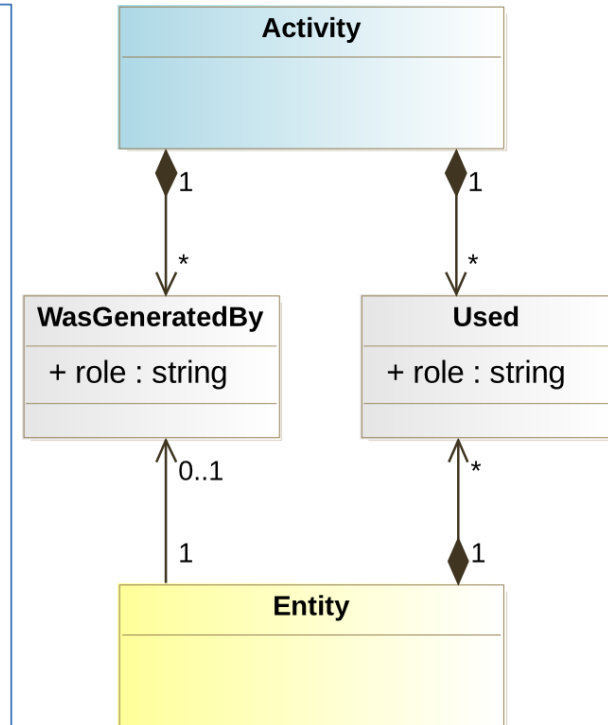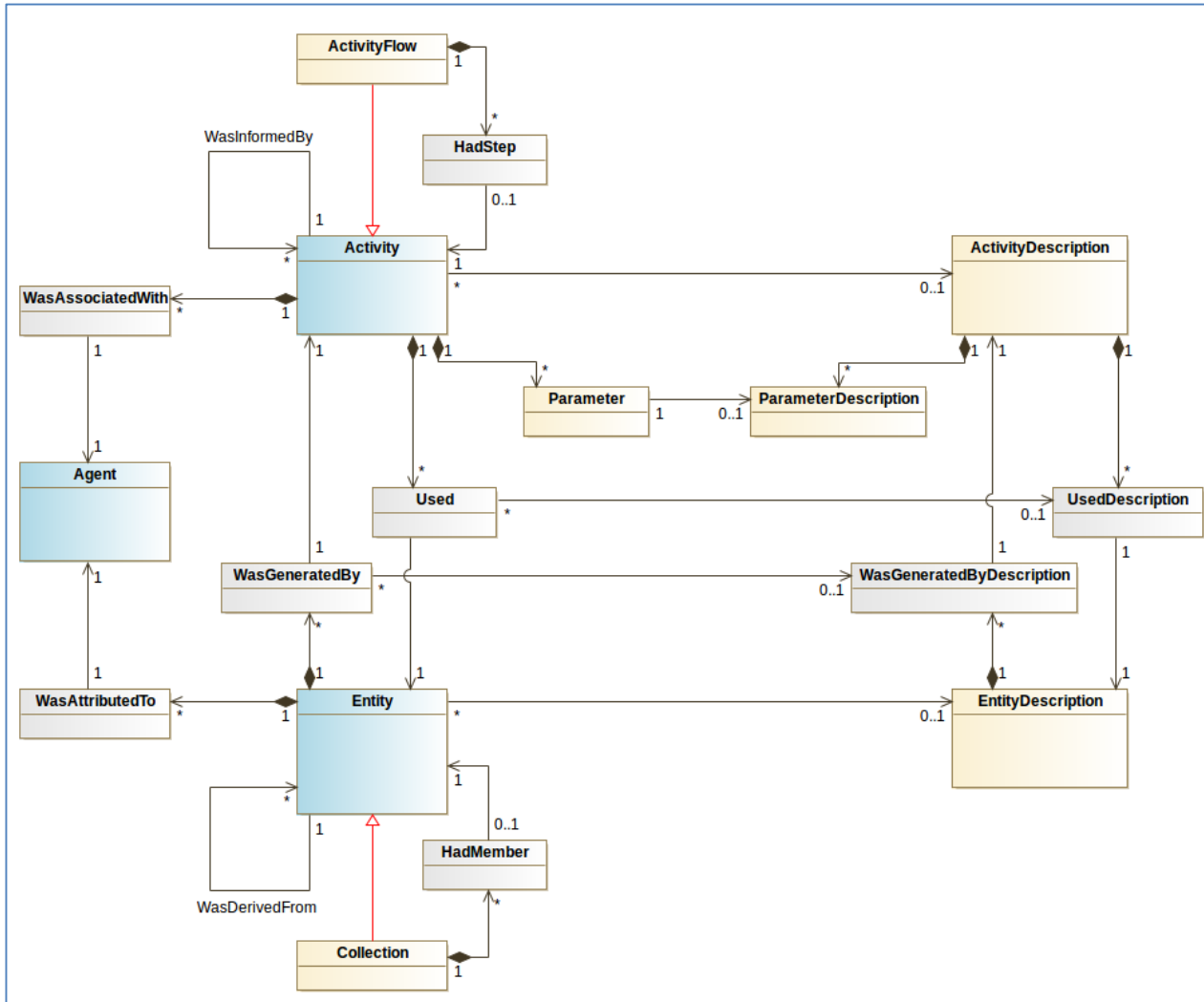
VODML
UML v. 2017May15

VODML compliancy

- Many *< ----> * relationships in the Domain model

- How to properly translate this to our conception model?

  – Using double aggregation/compositions:

    not allowed by VO-DML, produces validation errors

  – Use one composition + 1-1 reference?

    => Still being discussed

Current design questions

How to bind to subsets classes of existing
IVOA models and

represent ancillary data

Agent → ActivityDescription relation (MS)

- Describe the algorithm, goals, and limits of the code used by an Activity
- Bind the author of the code to the ActivityDescription class instead of Activity
- Distinguish :
  - Realisation/execution of a task → operator Agent
  - Description of this task → author Agent for the code
- Description classes as W3C entities
- They are published , versionned, bound to authors
- WasAttributedTo relation

Entity, Dataset,EntityDescription

- Entity Vs dataset
- Keep it general as w3C entities
  - Allow hooks for any category of data
    - Simulation, observation, log file, list of parameter , etc
  - Is the goal of Entity description to make explicit the binding to existing IVOA DMs
  - EntityDescription
    - Type :
    - *Observation* → ds:Obsdataset.dataproduct_type
    - *Simulation* → sim:Dataset.output , etc… cf mapping table in appendix
    - *Ancillary* → custom mapping //to be defined //

How to access Provenance metadata

Various scenarios to consider: *on going work*

- 2 step discovery (dataset focused)
  - Discover a data set (event list, image, etc.) as in ObsTAP
  - Select Provenance elements from a ProvTAP service using the Entity.id
  - Access provenance description via a datalink service
  - Embedded Provenance description into the Dataset serialisation (ex: Json serialisation in FITS extension , SVOM)
- Provenance TAP service
- Query Provenance DM from a SimDM implementation
  - Cosmological simulation data mapped to ProvDM (KR)
  - Mapping is easy due to existing provenance concepts in SimDM
    - see SimDM-ProvDM mapping table in WD, table 13.

Prototyped Use cases on the way

- Cosmological simulations (K. Riebe)
  - Links with SimDM
- CTA pipeline (M. Servillat)
  - recordind of Provenance embedded in the pipeline code
- Pollux data base A data product (M. Sanguillon)
  - Synthetic and theoretical spectra
- Hips generation (F. Bonnarel)
  - Tracing back the generation of HiPS surveys
- SVOM provenance design (L. Michel) starting

Outside VO community

- RDA meeting Barcelona April 2017
- A plan to have a Provenance WG in RDA
  - Charter beeing prepared, IG→ WG launch waiting for approval
  - Re-use of existing work Prov W3C DM and suite
  - Connection to IPAW conferences and ANDS Australian librarian circle
  - Plan to gather Provenance use-cases across disciplines ex: Blue Brain Project (Swiss, Samuel Kerrien)
  - Data + prov solutions
  - A catalog to be maintained by the WG group

Conclusion

- Adoption of the existing W3C provenance Model framework

- Implementation tests for 4 use-cases are on the way

- A variety of use-cases with different access and serialisation strategies

- Coupled interaction between astronomical projects and modeling → prototypes in an Agile-like process