



# IVOA PROVENANCE DM CURRENT STATUS

**Kristin Riebe, Mathieu Servillat, François Bonnarel,  
Mireille Louys, Michèle Sanguillon, Laurent Michel  
and the IVOA Data Model Working Group**

# A few figures

- **When** :Initiated end of 2015
- **Who**: A project stemming from IVOA members involved in the distribution of new data products
- **Where**: A focus group within the Data Model working group (~ 8-9)
- Links to data access protocols (DAL WG) and GWS WG for web services and authorization + Apps group

# IVOA Goals for Provenance

---



- **Describing how data sets were produced:**
  - Observing process and observing conditions (summary only)
  - Data reduction, selection and extraction methods applied to raw measures to build up science-ready data products (source lists, event lists, spectra, light curves, images, ...)
  - Workflows executed to build theoretical data (spectra, images, ...)
- **Helping VO users to:**
  - Derive selection criteria to filter out suitable data for his/her scientific needs
  - Estimate better which data release fits the best for their needs
  - Run his/her own reduction method on intermediate data products in order to refine data analysis
  - Apply customized / dedicated workflows

# Scope for Provenance

---



- **Not only history**
  - More than a log file
  - Attached to each science dataset distributed by a project archive if needed
  - *Traceability* : What happened to these data?
  - *Reproducibility* : How can I run the task again with the same or a different set of parameters

# Design for Provenance

---



- **Bring together commonalities**
  - General view for different archives /projects
  - Make it accessible via IVOA data access protocols
  - Enhance existing dataset distribution services by linking the provenance description.
    - Not a new « full picture » model
- **Benefit from outside experience and framework**
  - W3C Provenance model
  - Existing serialisation formats

# Provenance in the W3C<sup>®</sup>

---

- **4 recommendations (30/04/2013)**

**PROV-DM**: the PROV data model → *extensible with new attributes*

**PROV-O**: the PROV ontology

**PROV-Constraint**: Constraints of the PROV Data Model

**PROV-N**: a notation for provenance aimed at human consumption

- **and a number of non-prescriptive notes**

**PROV-XML**: an XML schema for the PROV data model

**PROV-AQ**: Provenance access and query

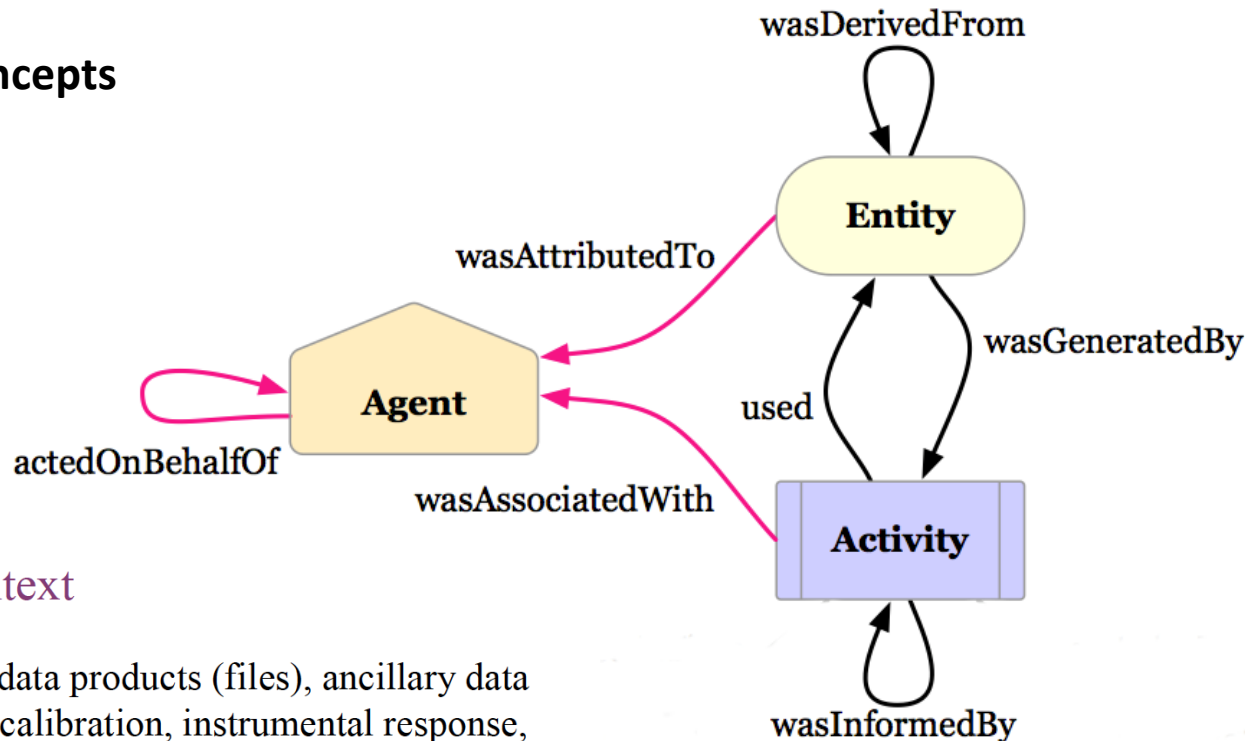
- **Tools to translate from one format to another & experience**

Southampton Provenance suite

<https://provenance.ecs.soton.ac.uk/>

**Visualization**

## Core Concepts

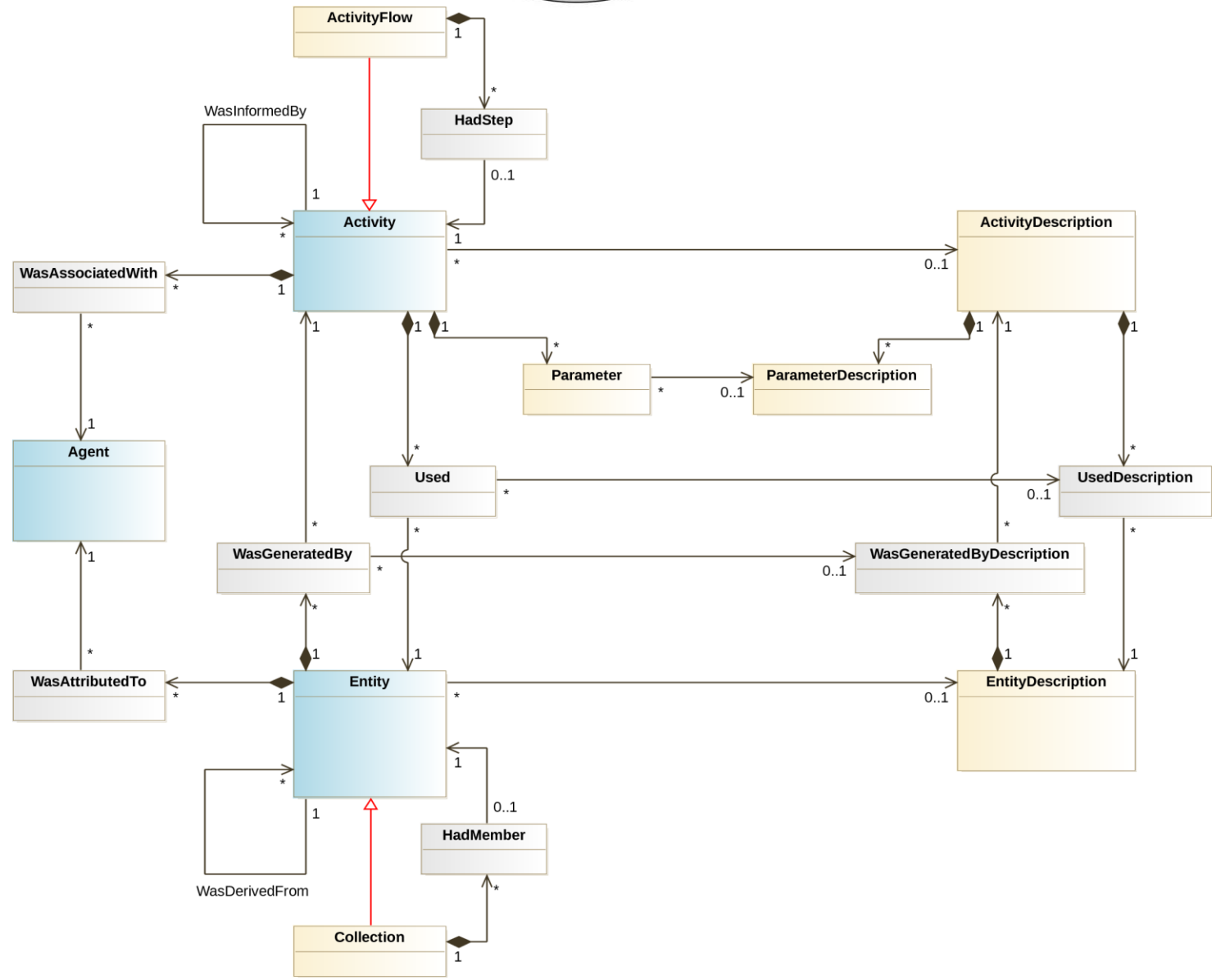


## In our context

- Entity** ■ data products (files), ancillary data (calibration, instrumental response, etc.), processing parameter files
- Activity** ■ data acquisition, mosaicing, regriding, fusion, calibration, ..., transformation
- Agent** ■ Telescope astronomer, pipeline operator, principal investigator, etc.



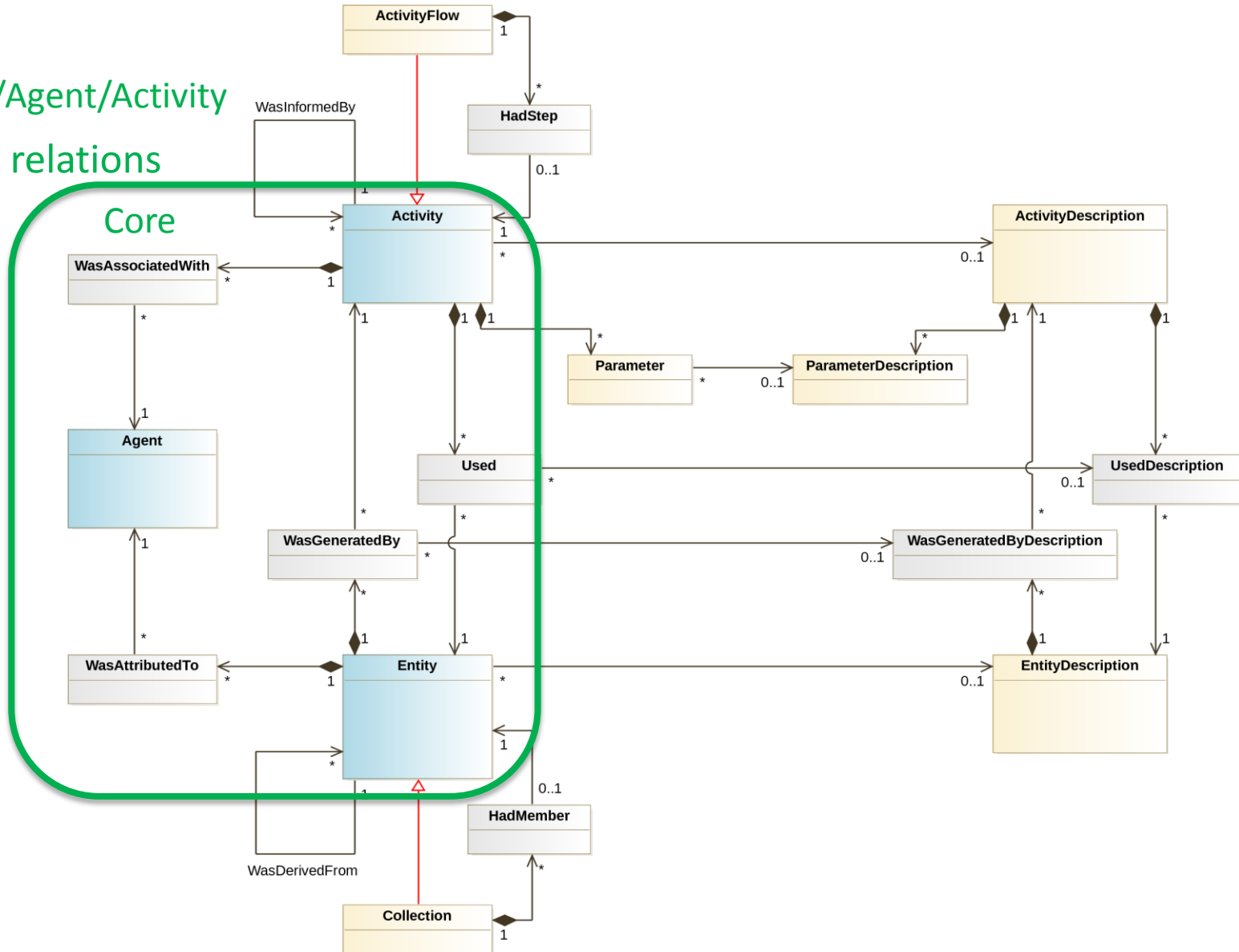
# Provenance DM



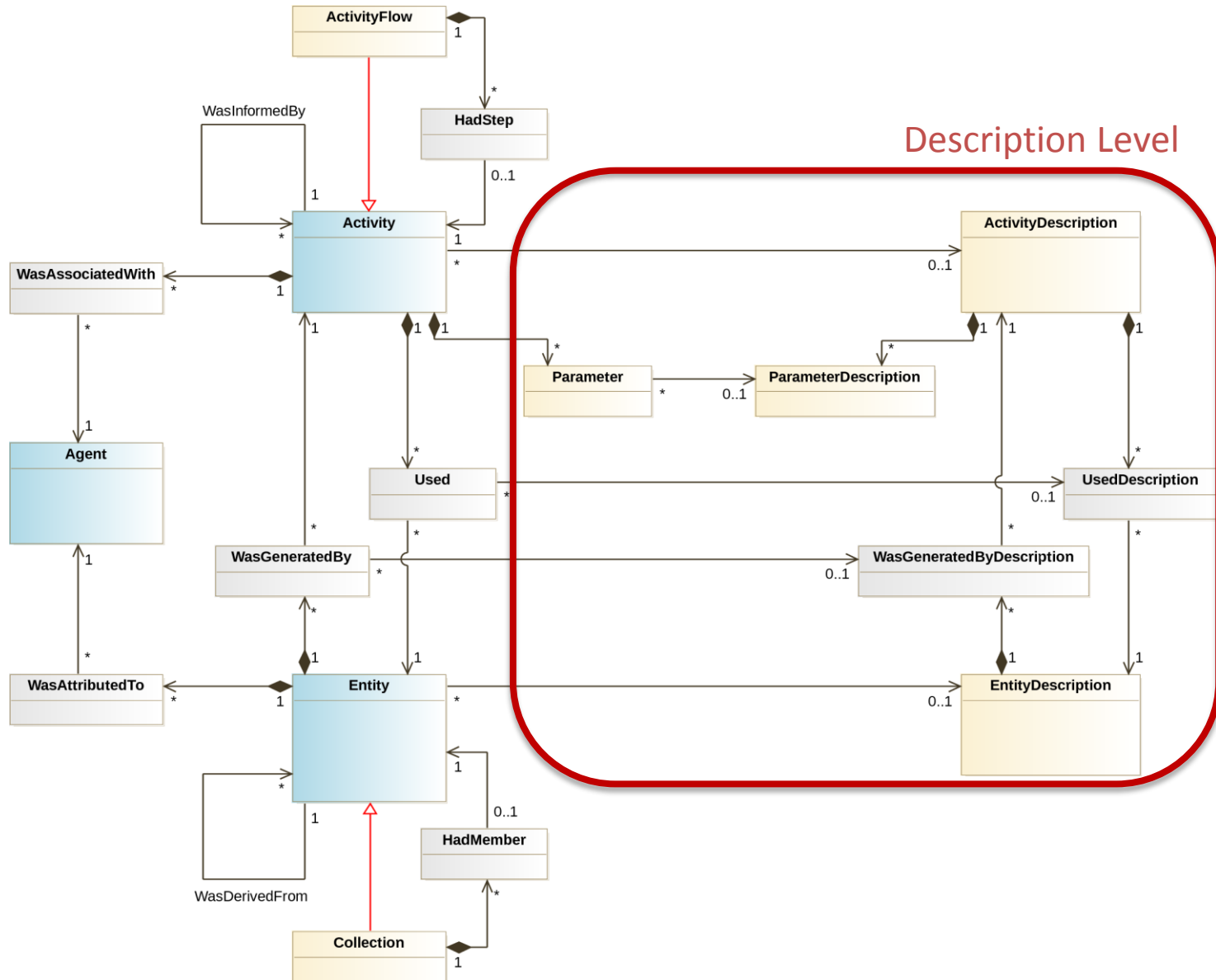


# IVOA Provenance DM

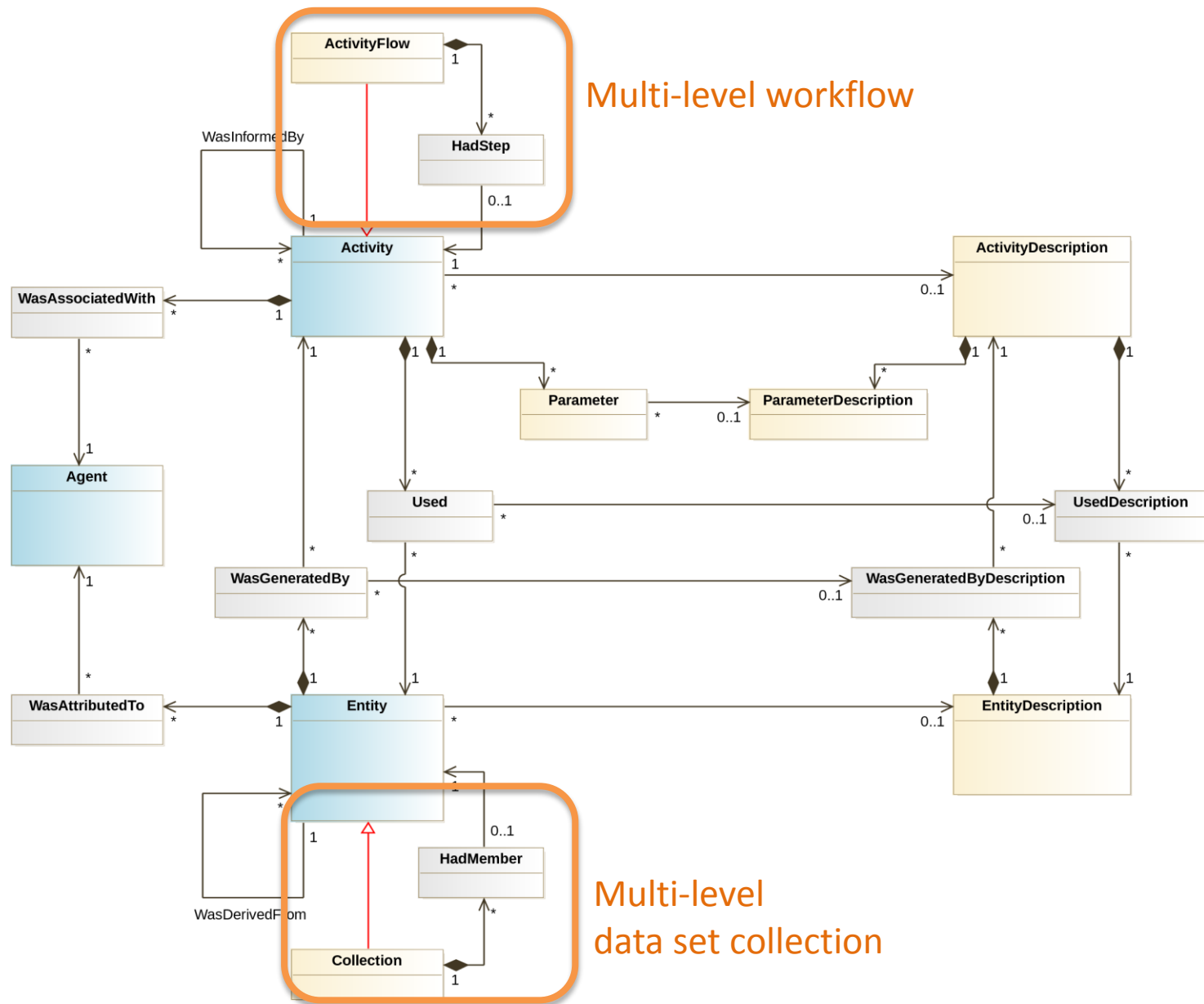
Entity/Agent/Activity  
+ relations



# IVOA Provenance DM



# IVOA Provenance DM



# Current status of IVOA Provenance DM

---

- A Core Provenance Model proposed by the IVOA data model working group at [ivoa.net/Documents](http://ivoa.net/Documents)  
See [WD-ProvenanceDM-1.0-20170921](http://ivoa.net/Documents/WD-ProvenanceDM-1.0-20170921)  
Currently discussed / slight updates to deliver  
Comments from the projects are very welcome
- Implementation examples
- TAP\_SCHEMA representation in progress

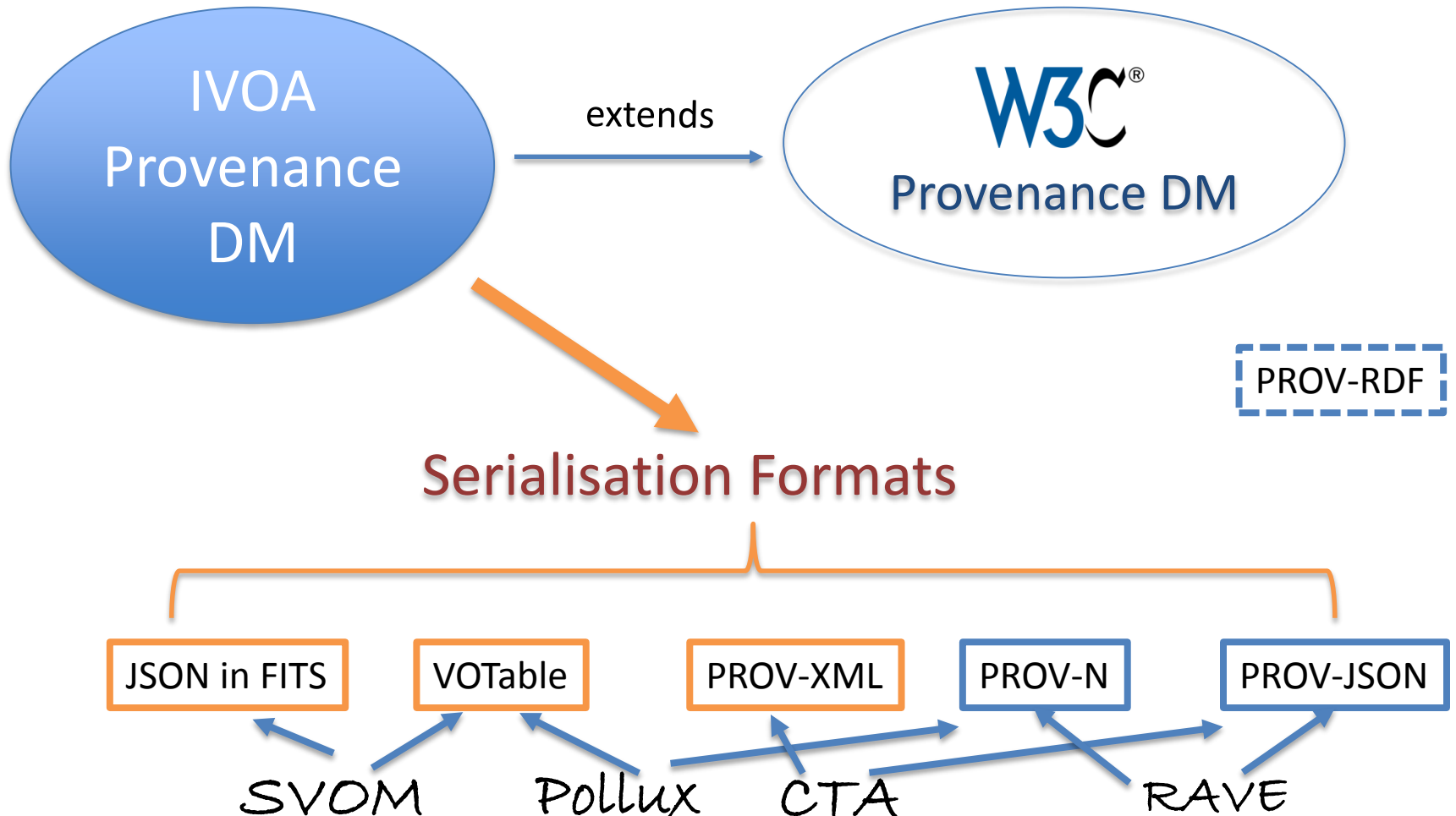
# IVOA Implementations

---

## Based on four use-cases

- [CTA](#) ( see M. Servillat's talk)
- [RAVE – the Radial Velocity Experiment](#)
- [POLLUX](#) (synthetic stellar spectra service)
- [SVOM](#) gamma ray burst /transients
- Prototype TAP-based API for images in an archive (@CDS)

# DM relations and serialisation



# PROV-N format

activity(MAMANUM-1, -, -)

entity(ivo://SERC/Plate#SERC.J.444)

entity(ivo://gepi/MAMA/Num#SERC.J-MAMA.444)

used(MAMANUM-1, ivo://SERC/Plate#SERC.J.444, -)

wasGeneratedBy(ivo://gepi/MAMA/Num#SERC.J-MAMA.444, MAMANUM-1, -)

activity(cds\_cutoutj444, -, -)

entity(ivo://cds/P/MAMA/SERC#SERC.J-MAMA.444)

used(cds\_cutoutj444, ivo://gepi/MAMA/Num#SERC.J-MAMA.444, -)

wasGeneratedBy(ivo://cds/P/MAMA/SERC#SERC.J-MAMA.444, cds\_cutoutj444, -)

activity(AlaRGB6, -, -)

entity(ivo://cds/P/MAMA/ESO#ESO.R-MAMA.444)

used(AlaRGB6, ivo://cds/P/MAMA/ESO#ESO.R-MAMA.444, -)

# Various usage of the model

- Lessons learnt from implementation

Project/ implemented features	Activity Flow Multi-level Activity	Activity /Entity description	Serialisation Formats
RAVE	Yes	Yes	PROV-JSON , PROV-N
CTA	Yes for future	Yes	VOTable, PROV-JSON
SVOM	?	yes	PROV-JSON , JSON in Fits
POLLUX	No	Yes	All

- Various profiles to be defined into the IVOA Provenance specification



# Looking for feedback

- The datamodel is general enough to fit for various usecases
- Collect feedback to define consistent profiles and provide implementation templates
- Is ASTERICS the network to gather this experience ?
- Will you consider to use it and contribute?

# Provenance WP management

---



Provenance days : meeting site

<http://wiki.ivoa.net/twiki/bin/view/IVOA/ObservationProvenanceDataModel>



Reports and discussions at DM sessions during the IVOA Interoperability meetings



Connections to interdisciplinary projects  
See WG for « Provenance Patterns »

More material :

[http://wiki.ivoa.net/internal/IVOA/InterOpOct2017DM/servillat\\_IVOA2017CL\\_provenance.pdf](http://wiki.ivoa.net/internal/IVOA/InterOpOct2017DM/servillat_IVOA2017CL_provenance.pdf)

dm@ivoa.net

**Thanks**