

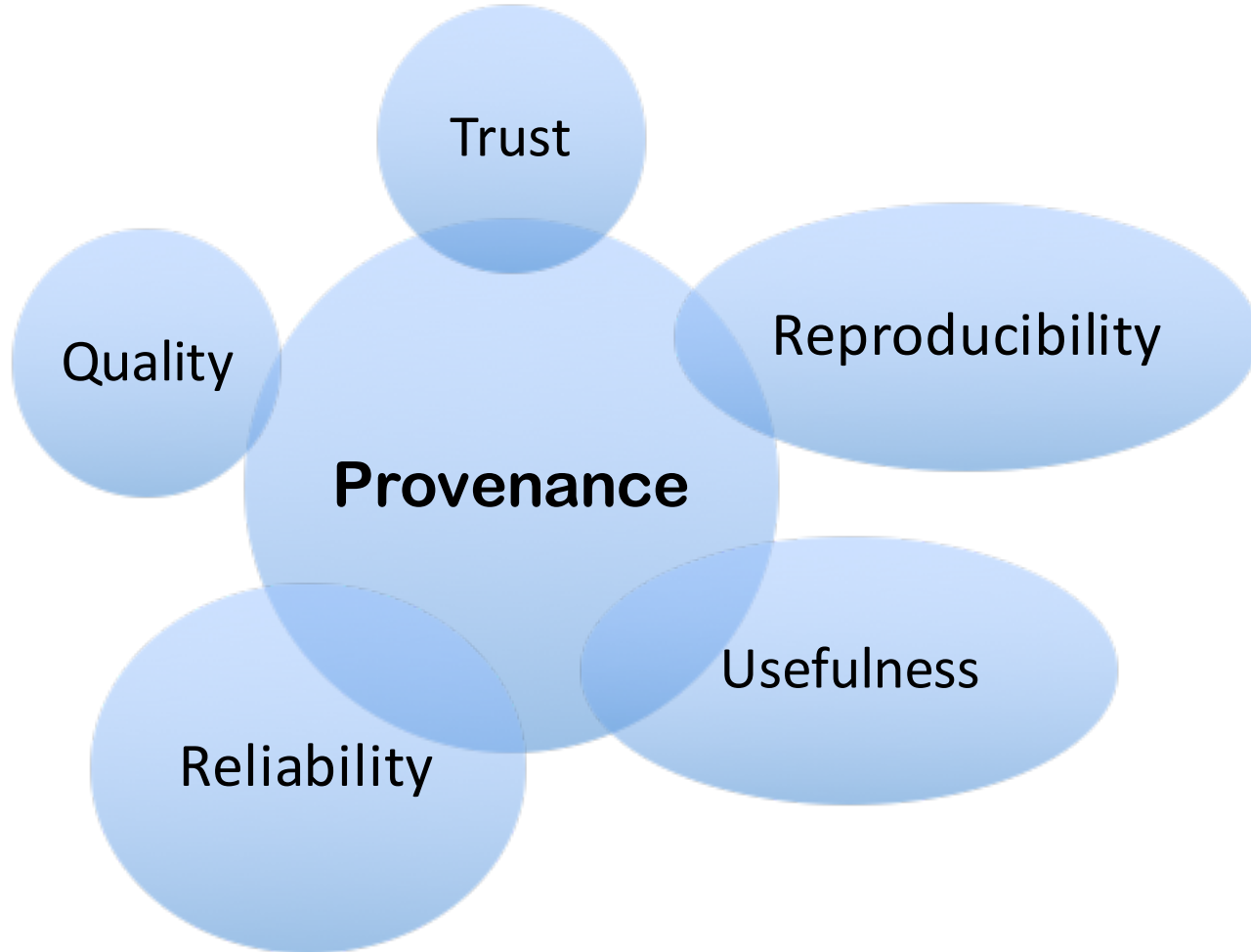


PROVENANCE @ IVOA

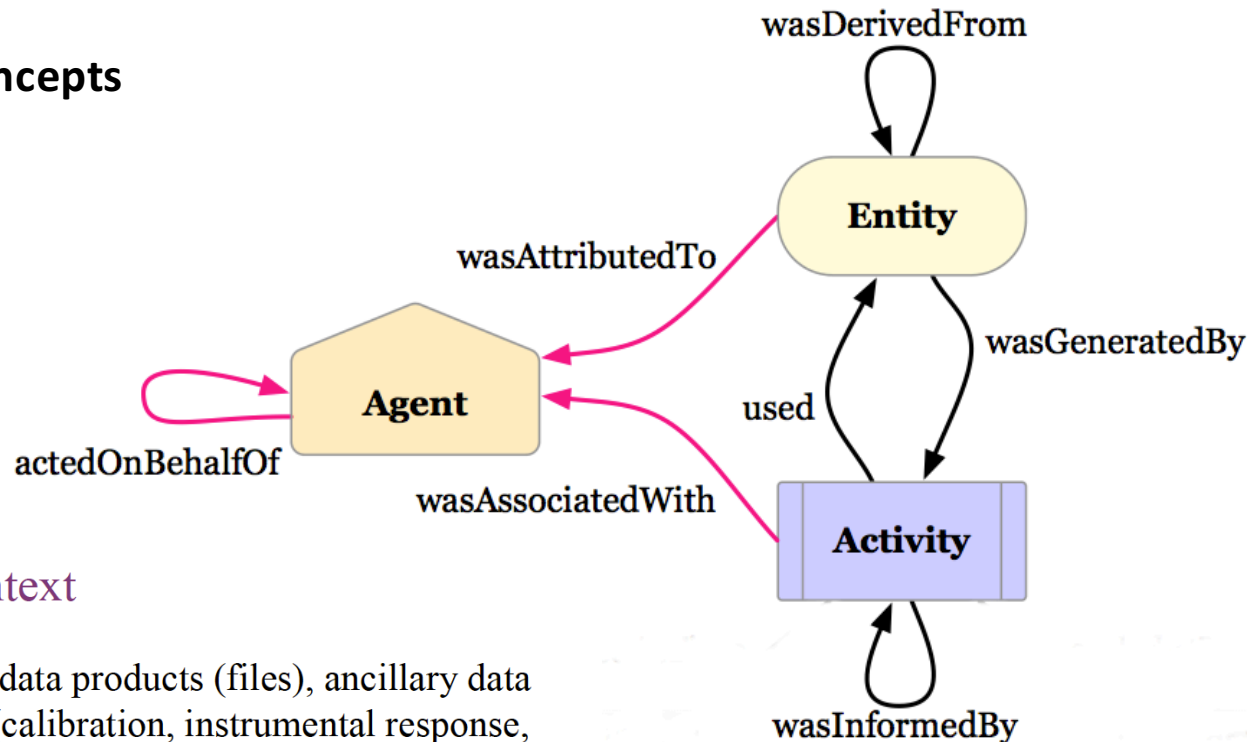
Kristin Riebe, [Anastasia Galkin](#), Ole Streicher, AIP
Mathieu Servillat, [Catherine Boisson](#), LUTH
[François Bonnarel](#), [Mireille Louys](#), CDS
Michèle Sanguillon, LUPM,
[Markus Nullmeier](#), Ari, Heidelberg, GAVO
Laurent Michel, SVOM, Observatoire de Strasbourg
and the IVOA Data Model Working Group

Goals for Provenance

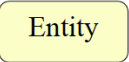
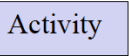

- **A: Tracking the production history**
 - Find out which steps were taken to produce a dataset and list the methods/tools/software that were involved.
- **B: Attribution and contact information**
 - Find the people involved in the production of a dataset, that need to be cited or can be asked for more information.
- **C: Locate error sources**
 - Find the location of possible error sources in the generation of a dataset.
- **D: Quality assessment**
 - Judge the quality of an observation, production step or dataset.
- **E: Search in structured provenance metadata**
 - Allow for “forward search”, i.e. locate derived datasets or outputs from a computation step.
 - Allow to get ancestors datasets of a given dataset



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 in the W3C[®]

- **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 available

- Southampton Provenance suite

<https://provenance.ecs.soton.ac.uk/> **translate between formats, visualize**

<https://lucmoreau.wordpress.com> **Tutorials**

<https://pypi.org/project/prov/> **Prov Python library**

IVOA Provenance effort

- Dedicated to the astronomical domain and based on the IVOA existing formats and tools and on good practices
- Datamodel: IVOA working draft close to final state : [WD-ProvenanceDM-1.0-20180530.pdf](http://ivoa.net/Documents/WD-ProvenanceDM-1.0-20180530.pdf) on ivoa.net/Documents
- Protocols :
 - PROV-TAP
 - PROV-SAP, simple access protocol
- Serialisation formats: VOTable, PROV-XML, JSON, PROV-N, FITS

IVOA Provenance Data Model Version 1.0



IVOA Working Draft 2018-05-30

Working group
DM

This version

<http://www.ivoa.net/documents/ProvenanceDM/20180530>

Latest version

<http://www.ivoa.net/documents/ProvenanceDM>

Previous versions

[WD-ProvenanceDM-1.0-20170921.pdf](#)

[WD-ProvenanceDM-1.0-20161121.pdf](#)

[ProvDM-0.2-20160428.pdf](#)

[ProvDM-0.1-20141008.pdf](#)

Author(s)

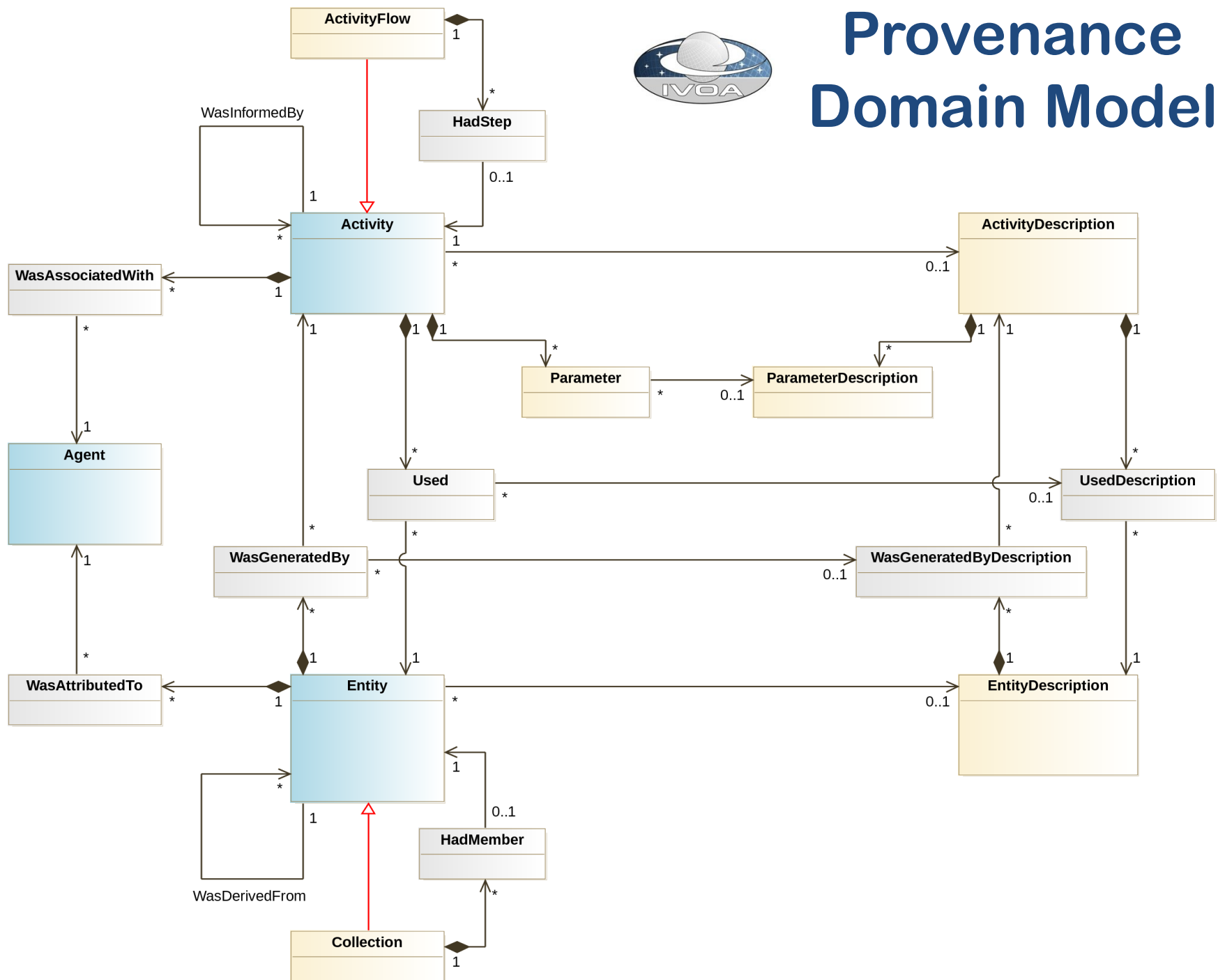
Kristin Riebe, Mathieu Servillat, François Bonnarel, Anastasia Galkin, Mireille Louys, Markus Nullmeier, Florian Rothmaier, Michèle Sanguillon, Ole Streicher, IVOA Data Model Working Group

Editor(s)

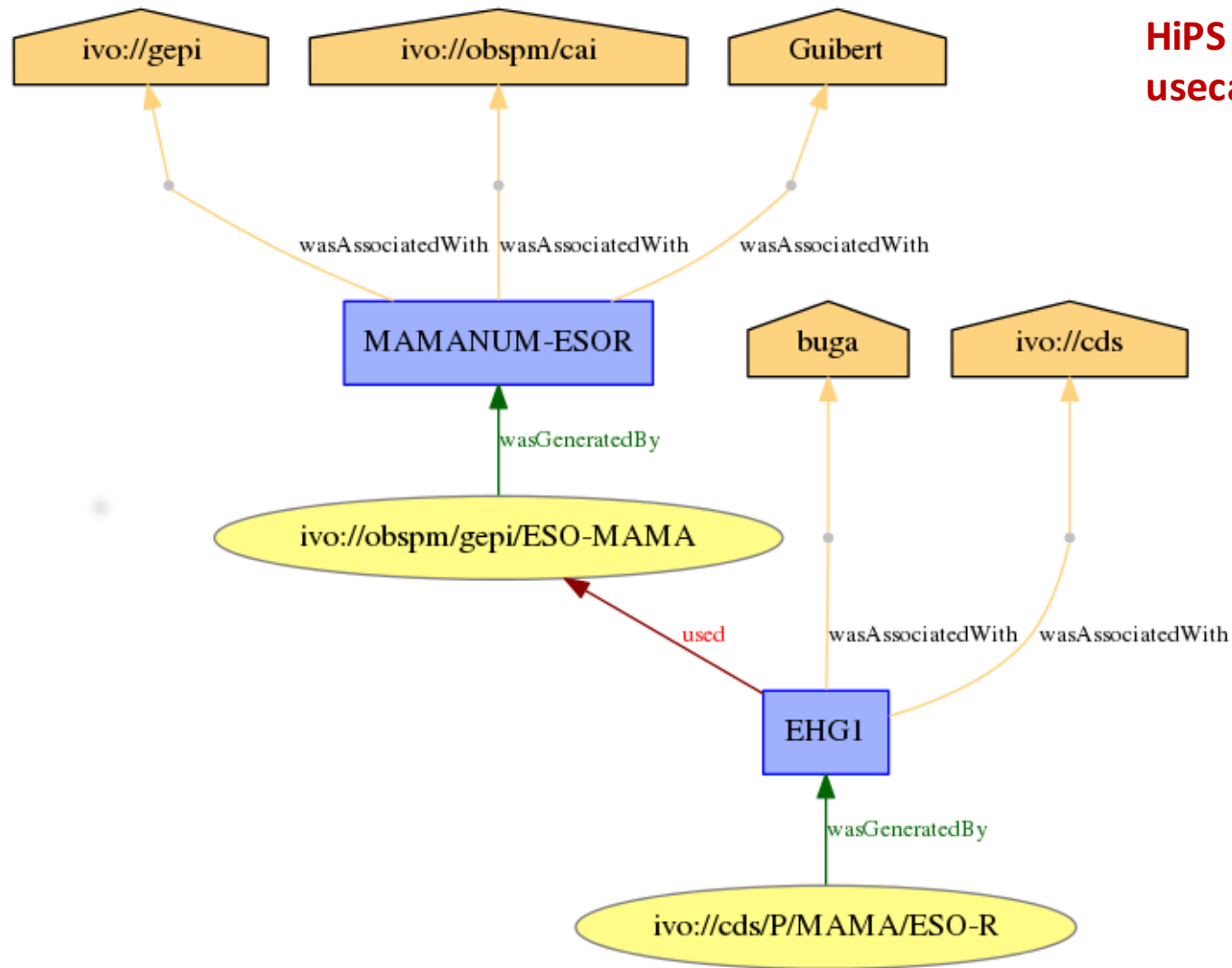
Kristin Riebe, Mathieu Servillat



Provenance Domain Model



W3C Graphical Representation



**HiPS generation
usecase**

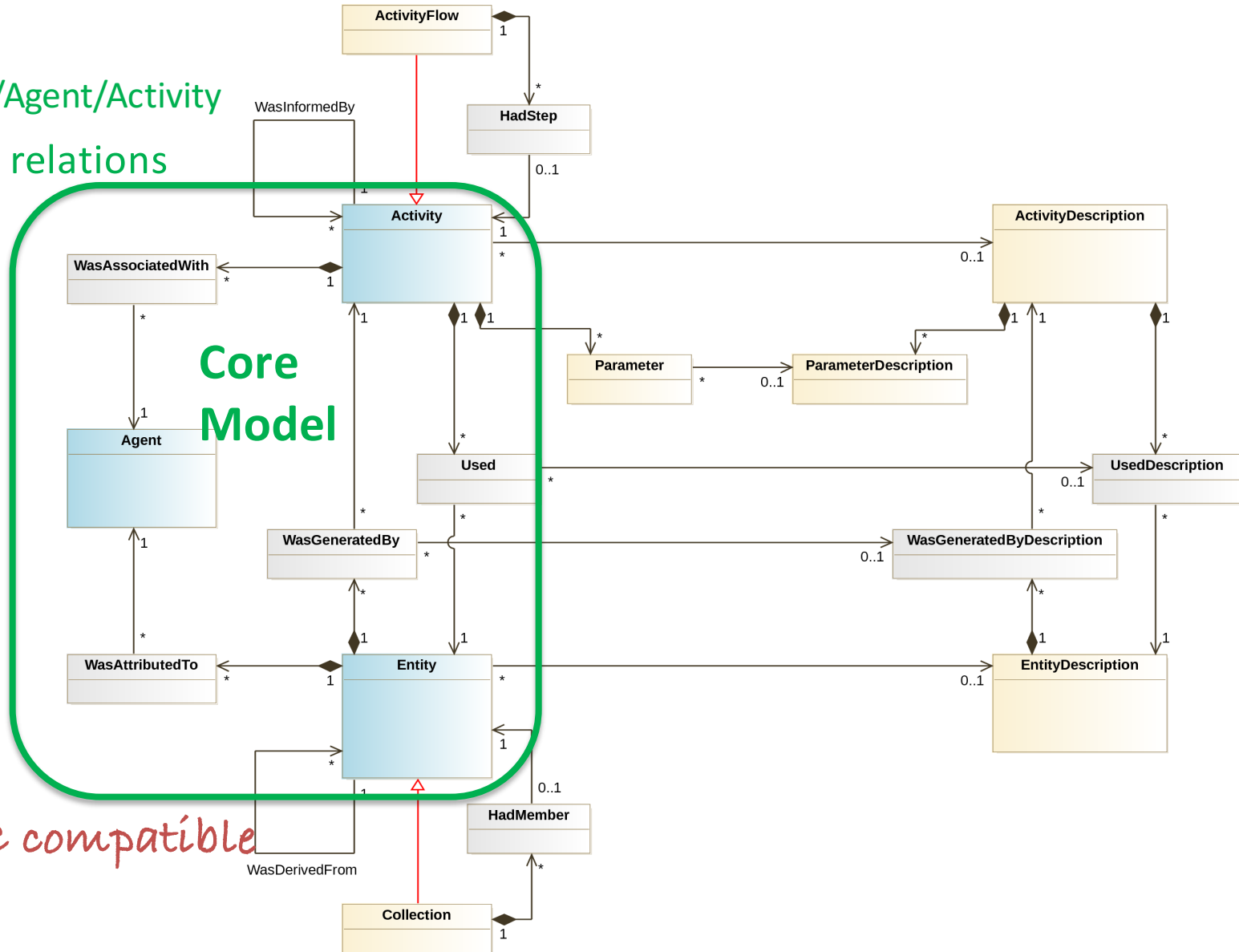
Metadata representation

**PROV-N format
as tuple in a relational DB**

```
activity(MAMANUM-1, date1, date2)
entity(ivo://SERC/Plate#SERC.J.444)
used(MAMANUM-1, ivo://SERC/Plate#SERC.J.444, -)
wasGeneratedBy(ivo://gepi/MAMA/Num#SERC.J-MAMA.444, MAMANUM-1, -)
entity(ivo://gepi/MAMA/Num#SERC.J-MAMA.444)
used(cds_cutoutj444, ivo://gepi/MAMA/Num#SERC.J-MAMA.444, -)
activity(cds_cutoutj444, -, -)
wasGeneratedBy(ivo://cds/P/MAMA/SERC#SERC.J-MAMA.444, cds_cutoutj444, -)
entity(ivo://cds/P/MAMA/SERC#SERC.J-MAMA.444)
used(AlaRGB6, ivo://cds/P/MAMA/SERC#SERC.J-MAMA.444, -)
activity(AlaRGB6, -, -)
entity(ivo://cds/P/MAMA/ESO#ESO.R-MAMA.444)
used(AlaRGB6, ivo://cds/P/MAMA/ESO#ESO.R-MAMA.444, -)
entity(ivo://cds/P/DSS2/SERC#SERC.I-DSS2.445)
used(AlaRGB6, ivo://cds/P/DSS2/SERC#SERC.I-DSS2.445, -)
wasGeneratedBy(ivo://cds/P/DSS2color#RGB_M83, AlaRGB6, -)
entity(ivo://cds/P/DSS2color#RGB_M83)
```

IVOA Provenance DM

Entity/Agent/Activity
+ relations



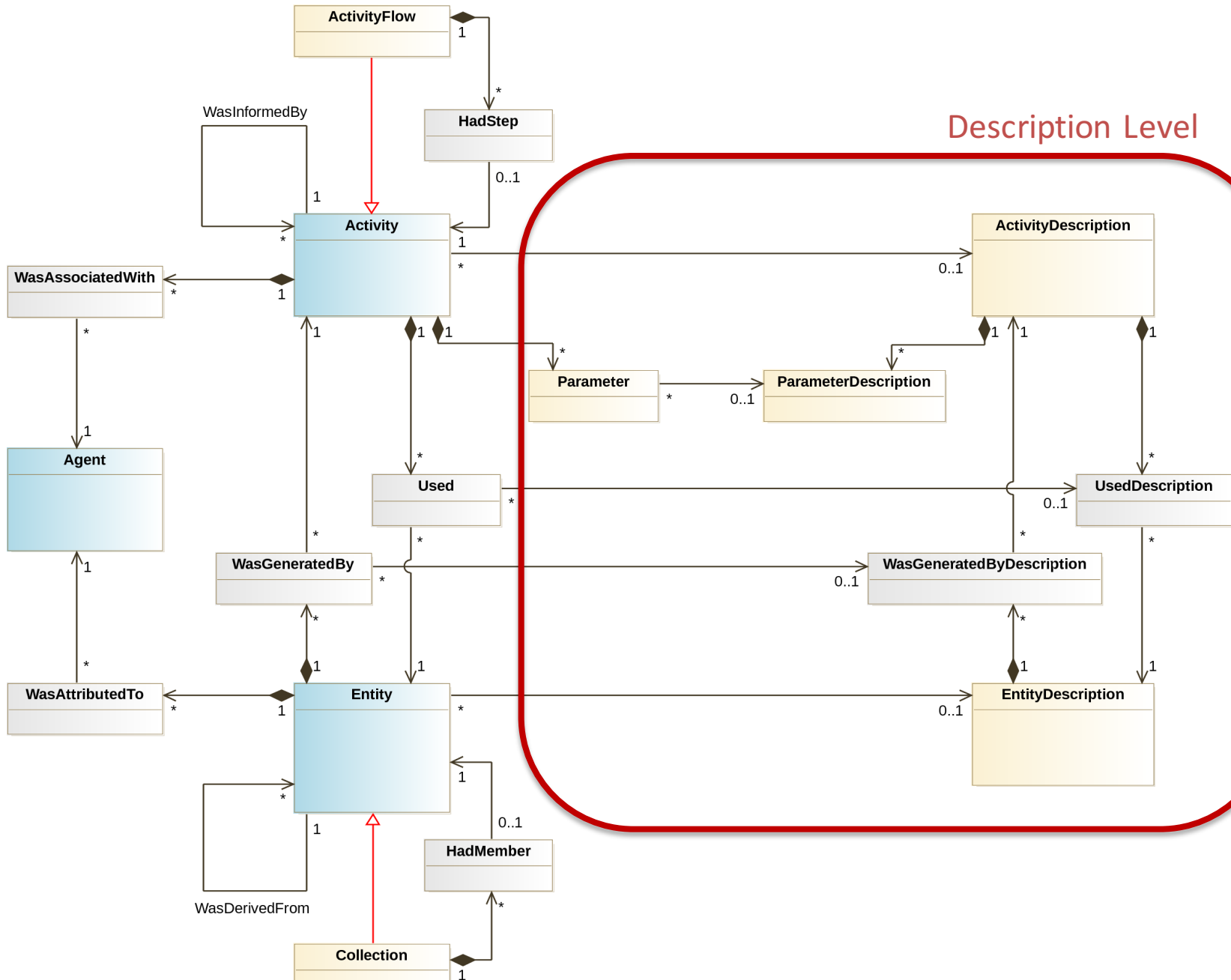
Core Model

W3C compatible

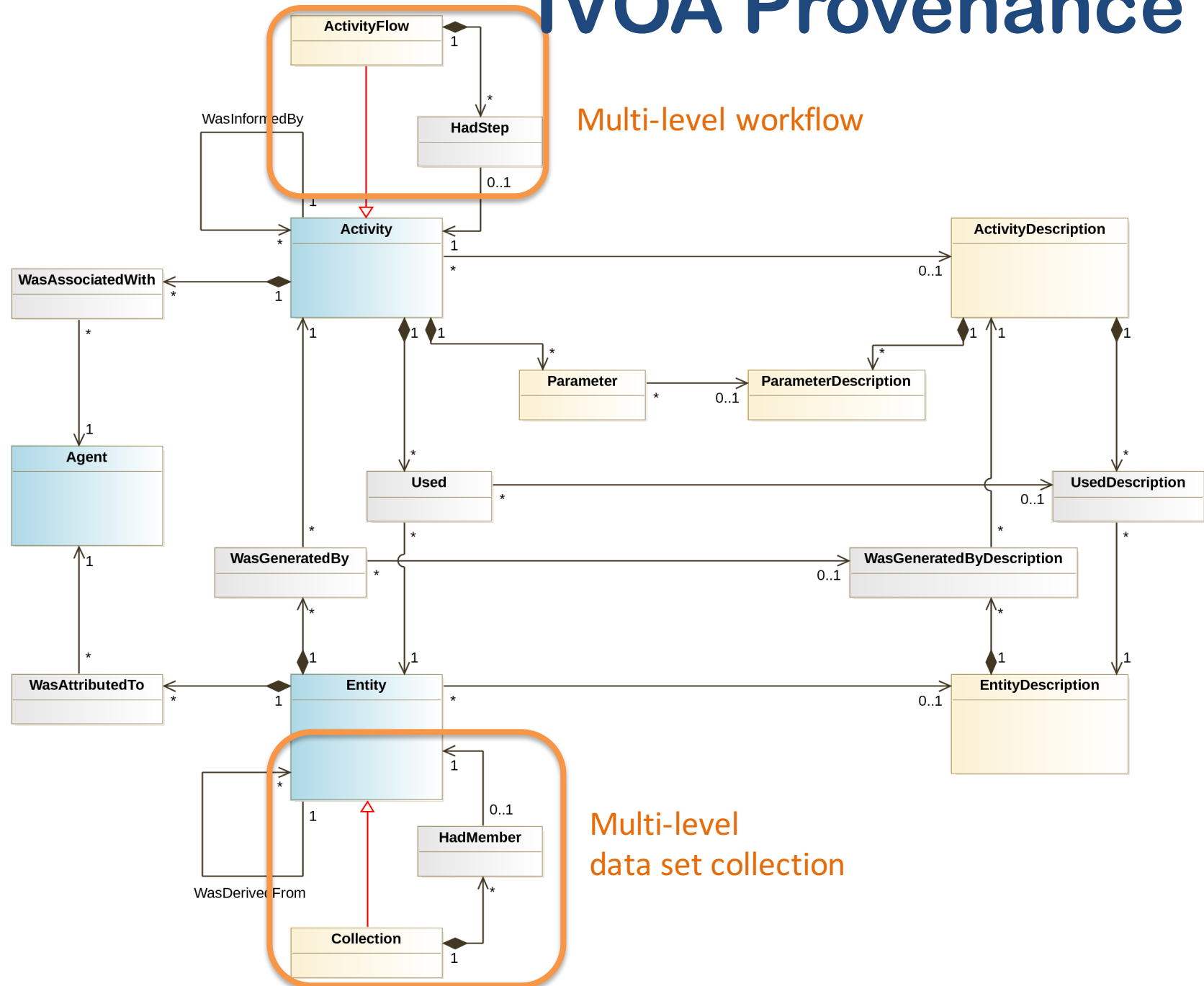
Core model → serialisation instances

- In **IVOA** framework
 - Easy to express in IVOA multiple tables (VOTable)
 - TAP compatible
 - Compatible with IVOA tools
- In **W3C** framework
 - All classes represented in IVOA DM also exist in W3C, just expand attributes
 - Easy to express in all W3C formats: PROV-XML, PROV-N, PROV-JSON, PROV-RDF

IVOA Provenance DM



IVOA Provenance DM

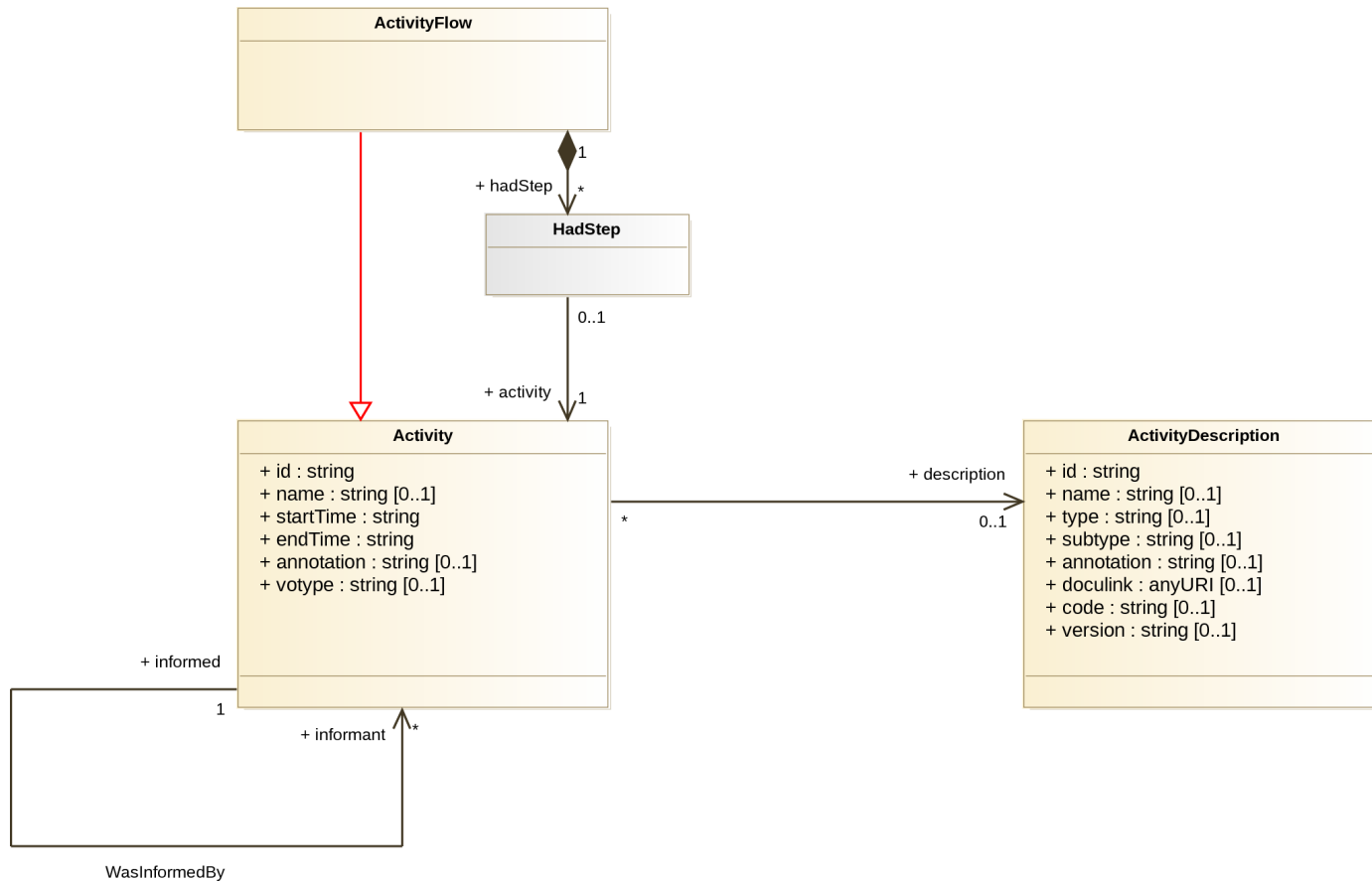


Multi-level workflow

Multi-level data set collection

IVOA Classes for Description

- Enrich the W3C classes by adding new classes
ActivityDescription, EntityDescription, ActivityFlow, etc.



In the VO ecosystem



Agents

ident	name	type
1	ivo./cds	Centre de données astronomiques de Strasbourg
2	bonnarel	François Bonnarel
3	ivo./stsci	Space Telescope
4	ivo./Palomar	Palomar Schmidt Telescope
5	ivo./ESO	ESO Schmidt Telescope
6	ivo./SERC	SERC Sliding Spring Schmidt Telescope
7	ivo./gepi	GEPi
8	ivo./obsnpvcal	centre analyse des images
9	buga	Mihaela Buga
10	Guibert	Jean Guibert

ActivityDescription

label	type	subtype	url
Aladin RGB image generation algorithm	RGBencoding		http://cds.u-str
ITSCI scan	DSS2 Plate digitization	Micro densitometer	http://stsci/DS
MAMA digitizer pipeline	Plate digitization	Micro densitometer	http://gepi/mar
ut out service	soda	cutout	http://cds.u-str
HIPS Generation MEAN	HIPsgen	HIPsgen_MEAN	http://cds.u-str

Entity

ident	label	type	annotation
61	ivo./POSSII/Plate#POSSII.N.143	POSSII Infra Red Survey DSS2 plate 143	This is the DSS2 digitization of
62	ivo./POSSII/Plate#POSSII.J.174	POSSII Blue Survey DSS2 plate 174	This is the DSS2 digitization of
63	ivo./POSSII/Plate#POSSII.F.174	POSSII Red Survey DSS2 plate 174	This is the DSS2 digitization of
64	ivo./POSSII/Plate#POSSII.N.175	POSSII Infra Red Survey DSS2 M101	This is the DSS2 digitization of
65	ivo./POSSII/Plate#POSSII.J.270	POSSII Blue Survey DSS2 270	This is the DSS2 digitization of
66	ivo./POSSII/Plate#POSSII.F.270	POSSII Red Survey DSS2 270	This is the DSS2 digitization of
67	ivo./POSSII/Plate#POSSII.N.270	POSSII Infra Red Survey DSS2 270	This is the DSS2 digitization of
68	ivo./POSSII/Plate#POSSII.J.413	POSSII Blue Survey DSS2 413	This is the DSS2 digitization of
69	ivo./POSSII/Plate#POSSII.F.413	POSSII Red Survey DSS2 413	This is the DSS2 digitization of
70	ivo./POSSII/Plate#POSSII.N.413	POSSII Infra Red Survey DSS2 413	This is the DSS2 digitization of
71	ivo./SERC/Plate#SERC.I.445	SERC Infra Red Survey 445	This is the Infra Red SERC Sch
72	ivo./SERC/Plate#SERC.J.444	SERC blue Survey 444	This is the blue SERC Schmid
73	ivo./ESO/Plate#ESO.R.444	ESO Red Survey 444	This is the DSS2 digitization of
74	ivo./ESO/Plate#ESO.R.445	ESO Red Survey 444	This is the DSS2 digitization of
75	ivo./ESO/Plate#ESO.R.446	ESO Red Survey 445	This is the DSS2 digitization of
76	ivo./ESO/Plate#ESO.R.447	ESO Red Survey 446	This is the DSS2 digitization of
77	ivo./cds/PMAMA/ESO-R	ESO-R MAMA HIPS at CDS	This is the HIPS version of ESO
78	ivo./obsnpvgepi/ESO-MAMA	ESO-R Survey MAMA GEPi	This is the ESO Schmidt survey

Activity

ident	label	start	stop	annotation	
1	AlaRGB1	Aladin RGB 1	2017-04-18T17:28:00	2017-04-19T17:29:00	Aladin RGB image generation for MCG 6946
2	AlaRGB2	Aladin RGB 2	2017-04-18T17:34:00	2017-04-19T17:35:00	Aladin RGB image generation for Messier 101
3	AlaRGB3	Aladin RGB 3	2017-04-18T17:41:00	2017-04-19T17:42:00	Aladin RGB image generation for Messier 33
4	AlaRGB4	Aladin RGB 4	2017-04-18T17:45:00	2017-04-19T17:46:00	Aladin RGB image generation for Messier 53
5	AlaRGB5	Aladin RGB 5	2017-04-18T17:47:00	2017-04-19T17:48:00	Aladin RGB image generation for Messier 81
6	AlaRGB6	Aladin RGB 6	2017-04-18T17:50:00	2017-04-19T17:51:00	Aladin RGB image generation for Messier 83
7	AlaRGB7	Aladin RGB 7	2017-04-18T17:53:00	2017-04-19T17:54:00	Aladin RGB image generation for Messier 87
8	stscNum-21	Num DSS2 POSSII 061 J	2006-06-29T16:32:50		DSS2 Digitization of plates at Stsci POSSII 061 J
9	stscNum-22	Num DSS2 POSSII 061 F	2006-04-04T16:39:18		DSS2 Digitization of plates at Stsci POSSII 061 F
10	stscNum-23	Num DSS2 POSSII 061 N	2006-04-04T16:52:00		DSS2 Digitization of plates at Stsci POSSII 061 N
11	stscNum-24	Num DSS2 POSSII 143 J	2006-04-04T16:10:12	2006-04-04T16:10:30	DSS2 Digitization of plates at Stsci POSSII 143 J
12	stscNum-25	Num DSS2 POSSII 143 F	2006-04-04T16:25:45	2006-04-04T16:25:55	DSS2 Digitization of plates at Stsci POSSII 143 F
13	stscNum-26	Num DSS2 POSSII 143 N	2006-04-04T16:31:01	2006-04-04T16:31:21	DSS2 Digitization of plates at Stsci POSSII 143 N
14	stscNum-27	Num DSS2 POSSII 270 J	2006-04-04T16:17:26	2006-04-04T16:17:36	DSS2 Digitization of plates at Stsci POSSII 270 J
15	stscNum-28	Num DSS2 POSSII 270 F	2006-04-04T16:17:27	2006-04-04T16:17:37	DSS2 Digitization of plates at Stsci POSSII 270 F
16	stscNum-29	Num DSS2 POSSII 270 N	2006-04-04T16:35:11	2006-04-04T16:35:21	DSS2 Digitization of plates at Stsci POSSII 270 N
17	stscNum-2a	Num DSS2 POSSII 174 J	2006-04-04T16:36:03	2006-04-04T16:36:13	DSS2 Digitization of plates at Stsci POSSII 174 J
18	stscNum-2b	Num DSS2 POSSII 174 F	2006-04-04T16:22:28	2006-04-04T16:22:38	DSS2 Digitization of plates at Stsci POSSII 174 F

Parameter

isParamOf	id	value	unit	ucd
1	HipsPARAM	hips_tile_format	jpeg fits	meta.code.class
2	HipsPARAM	hips_order	6	meta.code

WasGeneratedBy

head	tail	role	
1	ivo./cds/IDSS2/color/RGB_M07	AlaRGB7	voprov.output
2	ivo./cds/IDSS2/POSSII#POSSII.N-DSS2.644	cds_cutout644	voprov.output
3	ivo./cds/IDSS2/POSSII#POSSII.F-DSS2.644	cds_cutout644	voprov.output
4	ivo./cds/IDSS2/POSSII#POSSII.N-DSS2.644	cds_cutout644	voprov.output

Used

head	tail	role	
1	AlaRGB1	ivo./cds/IDSS2/POSSII#POSSII.J-DSS2.143	voprov.creator
2	AlaRGB1	ivo./cds/IDSS2/POSSII#POSSII.F-DSS2.143	voprov.creator
3	AlaRGB1	ivo./cds/IDSS2/POSSII#POSSII.N-DSS2.143	voprov.creator
4	AlaRGB2	ivo./cds/IDSS2/POSSII#POSSII.J-DSS2.174	voprov.creator
5	AlaRGB2	ivo./cds/IDSS2/POSSII#POSSII.F-DSS2.174	voprov.creator
6	AlaRGB3	ivo./cds/IDSS2/POSSII#POSSII.N-DSS2.174	voprov.creator
7	AlaRGB4	bonnarel	voprov.operator

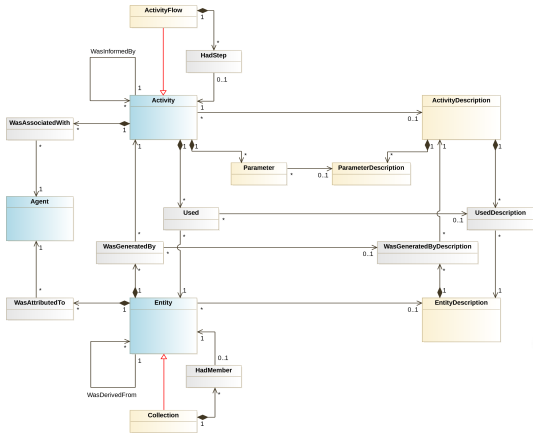
WasAssociatedWith

head	tail	role	
1	AlaRGB1	bonnarel	voprov.operator
2	AlaRGB1	ivo./cds	voprov.creator
3	AlaRGB2	bonnarel	voprov.operator
4	AlaRGB2	ivo./cds	voprov.creator
5	AlaRGB3	bonnarel	voprov.operator
6	AlaRGB3	ivo./cds	voprov.creator
7	AlaRGB4	bonnarel	voprov.operator

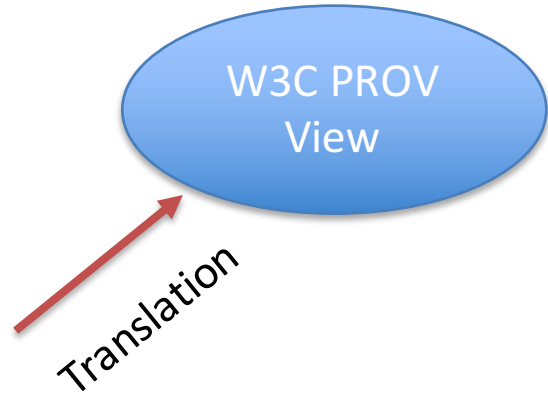
Specific classes Translation

- In IVOA ecosystem:
 - Use one table for each defined class and relation type
 - Fully extensible to relational DB
- In W3C:
 - No W3C structure corresponding to added IVOA classes for descriptions
 - Need some translation to the W3C existing constructs
 - **Refactoring of the IVOA PROV DM as a PROV W3C view.**

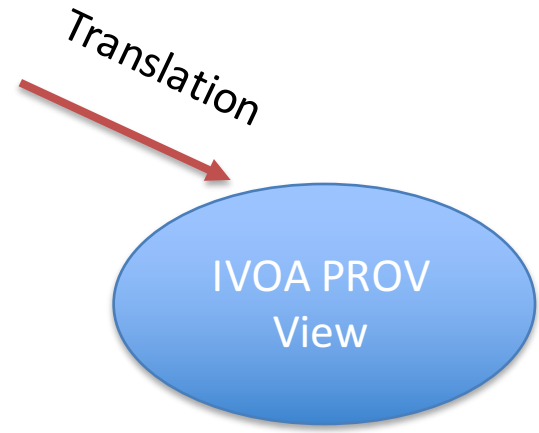
IVOA DM



All classes defined in the specification UML/VODML

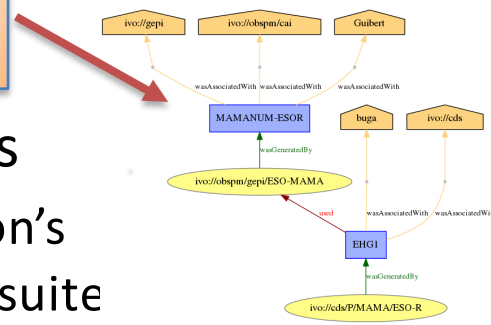


W3C PROV View



IVOA PROV View

PROV-JSON
PROV-N
PROV-XML



W3C Tools
Southampton's Provenance suite
ProvPython lib

Prov-SAP Simple Access Protocol

Prov-TAP TOPCAT TapHandle

VOTable
JSON
FITS

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)
- MUSE pipeline representation (O. Streicher, AIP)

Activity Description/ parameters



```
<RESOURCE ID="gammapy_maps" name="gammapy_maps" type="meta" utype="voprov:ActivityDescription">
  <DESCRIPTION>Use gammapy to generate a count map from a list of observations</DESCRIPTION>
  <!-- Service Descriptor -->
  <PARAM name="accessURL" datatype="char" arraysize="*" value="https://voparis-uws-test/rest/gammapy_maps" />
  <PARAM name="standardID" datatype="char" arraysize="*" value="ivo://ivoa.net/std/SODA#1.0" />
  <!-- Activity Description -->
  <PARAM name="type" datatype="char" arraysize="*" value="None" utype="voprov:ActivityDescription.type"/>
  <PARAM name="subtype" datatype="char" arraysize="*" value="None" utype="voprov:ActivityDescription.subtype"/>
  <PARAM name="annotation" datatype="char" arraysize="*" value="Use gammapy to generate a count map from a list of
  <PARAM name="version" datatype="char" arraysize="*" value="None" utype="voprov:ActivityDescription.version"/>
  <PARAM name="docuLink" datatype="char" arraysize="*" value="https://luthgitlab.obspm.fr/jlefaucheur/hess_release
  <PARAM name="contact_name" datatype="char" arraysize="*" value="Julien Lefaucheur" utype="voprov:Agent.name"/>
  <PARAM name="contact_email" datatype="char" arraysize="*" value="" utype="voprov:Agent.email"/>
  <!-- UWS job attributes -->
  <PARAM name="executionDuration" datatype="int" value="600" utype="uws:Job.executionDuration"/>
  <PARAM name="quote" datatype="int" value="120" utype="uws:Job.quote"/>
```

Activity Description

<!-- UWS parameters (Provenance Entities or Parameters) --> Parameters

```
<GROUP name="InputParams">
  <PARAM ID="obs_ids" arraysize="*" datatype="char" name="obs_ids" value="47802 47803 47804
  <DESCRIPTION>List of runs</DESCRIPTION>
  </PARAM>
  <PARAM ID="RA" datatype="double" name="RA" value="329.7169379" unit="deg"...>
  <PARAM ID="Dec" datatype="double"
  <PARAM ID="nxpix" arraysize="*" da
  <DESCRIPTION>Number of pixels
  <VALUES>
  <MIN value="0"/>
  <MAX value="1000"/>
  </VALUES>
  </PARAM>
  <PARAM ID="nypix" arraysize="*" da
  <PARAM ID="binsz" datatype="float"
</GROUP>
```

VOTable
DataLink Service Descriptor
UWS Job Description Language
Provenance ActivityDescription

```
<!-- Used Entities -->
<GROUP name="Used">
  <GROUP name="obs_ids" utype="voprov:UsedDescription" ref="obs_ids">
    <PARAM arraysize="*" datatype="char" name="role" utype="voprov:UsedDescription.role" value="DL3"/>
    <PARAM arraysize="*" datatype="char" name="location" utype="voprov:EntityDescription.location" value="" />
    <PARAM arraysize="*" datatype="char" name="content_type" utype="voprov:EntityDescription.content_type" v
  </GROUP>
</GROUP>

<!-- Generated Entities / UWS results -->
<GROUP name="Generated" utype="voprov:WasGeneratedBy">
  <GROUP name="count_map" utype="voprov:EntityDescription">
    <DESCRIPTION>Count map</DESCRIPTION>
    <PARAM arraysize="*" datatype="char" name="role" utype="voprov:UsedDescription.role" value="DL4 image"/>
    <PARAM arraysize="*" datatype="char" name="default" utype="voprov:Entity.id" value="count_map.fits"/>
    <PARAM arraysize="*" datatype="char" name="content_type" utype="voprov:EntityDescription.content_type" v
  </GROUP>
  <GROUP name="count_preview" utype="voprov:EntityDescription">
    <DESCRIPTION>Count map preview</DESCRIPTION>
```

Activity Description

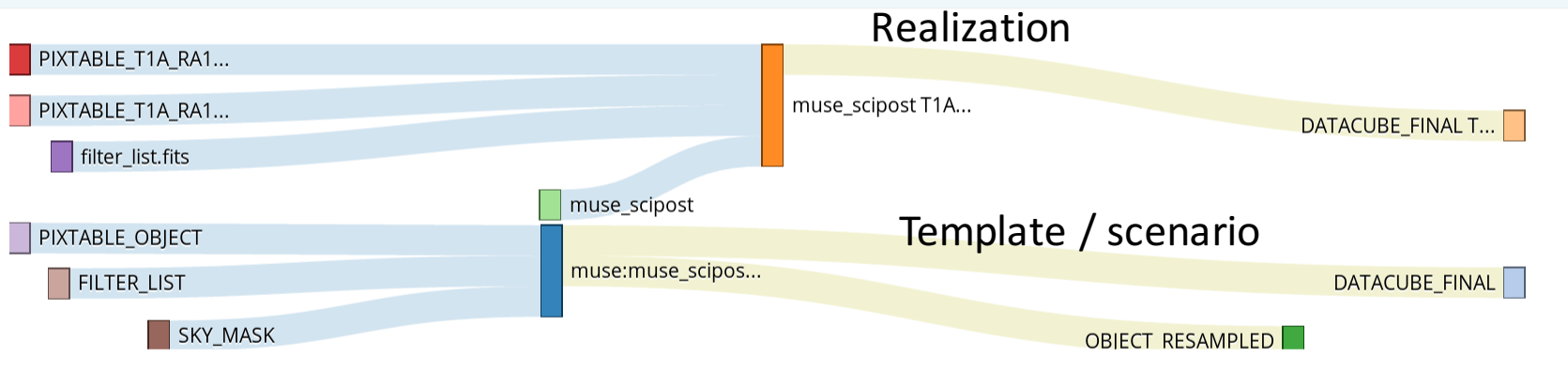
Muse scipost prov with template >

Visualizations

Sankey Wheel Hive Gantt

Select Visible Relations

Created on 17 Apr 2018 at 15:17 by olebole 2 views



<https://provenance.ecs.soton.ac.uk/store/documents/118181/>

Various usage of the model

Project/ implemented features	Protocol	Activity Flow Multi-level Activity	Activity /Entity description / Parameters	Serialisation Formats
RAVE	Prov-SAP	Yes	Yes	PROV-JSON , PROV-N
CTA	TAP/ UWS	Yes for future	Yes	VOTable, PROV-JSON, JSON
SVOM	Prov-SAP?	?	Yes	PROV-JSON, JSON in Fits
POLLUX	Adhoc then Prov-SAP	No	Yes	All
CDS ImageDB Prototype	PROV-TAP	No	Yes	VOTable, JSON, PROV- RDF
MUSE	AstroWise Pipeline	no	Yes as Bundles	PROV-N, PROV-JSON

- Various profiles to propose : Workflow view, data flow view, credits/rewards view depending on what the projects needs

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 »



Provenance Patterns

- <https://www.rd-alliance.org/groups/provenance-patterns-wg>
- Chaired by Nick Car, Geoscience Australia
 - A data base of elementary use cases
 - General enough to cover many application domains
 - Produced a set of Provenance patterns for these use cases
 - W3C Prov DM based
 - Implemented in Prov-RDF in TripleStore
- Interest shown for IVOA provenance use cases and how we have re-used the W3C concepts and serialisation formats
- One CDS intern to explore implementation of RDA Provenance patterns in RDF within a TripleStore for IVOA Provenance.

Contact questions & and suggestions
dm@ivoa.net

Thanks