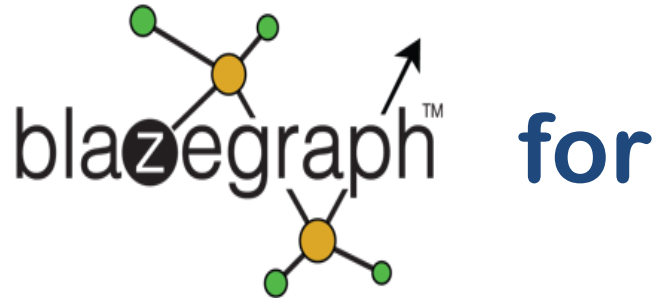




A triple store implementation with



the IVOA Provenance Data Model

Lucas Holzmann, ENSIIE Trainee April-August 2018

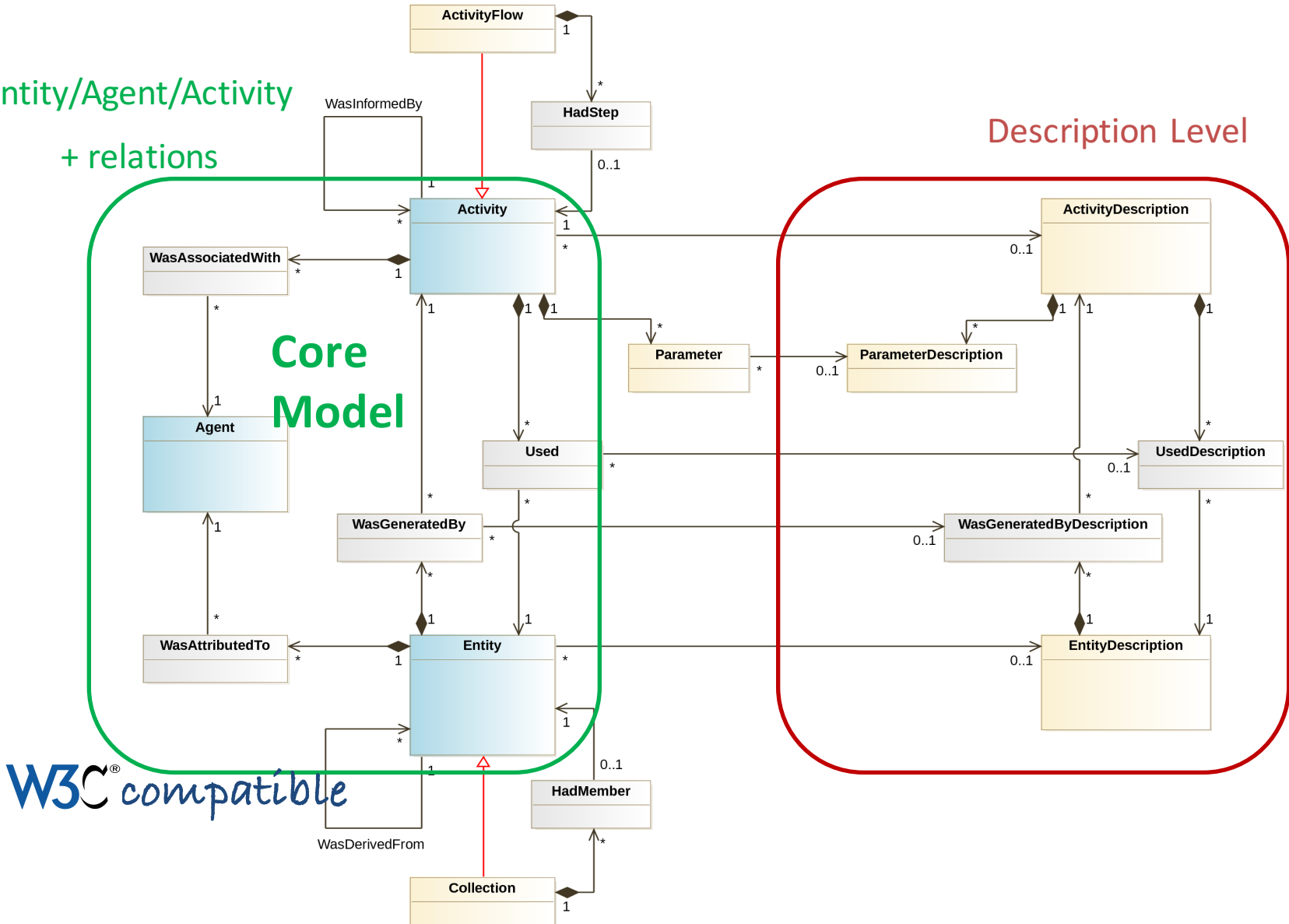
Mireille Louys, CDS & ICube, University Strasbourg

François Xavier Pineau, CDS

IVOA Provenance DM

Entity/Agent/Activity
+ relations

Description Level



VO ecosystem : VOTable & Relational DB

Agents

ident	name	type
1	ivo/icds	Centre de données astronomiques de Strasbourg
2	bonnarel	François Bonnarel
3	ivo/istsci	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://istsci/DS
MAMA digitizer pipeline	Plate digitization	Micro densitometer	http://gepi/mar
ut out service	soda	colour	http://cds.u-str
HPS Generation MEAN	HPSgen	HPSgen_MEAN	http://cds.u-str

Entity

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

isAParamOf	id	value	unit	ucd
HipsPARAM	hips_tile_format	jpeg fits		meta.code.class
HipsPARAM	hips_order	6		meta.code

WasGeneratedBy

head	tail	role
ivo/icds/P/DSS2/color/RGB_M07	AlaRGB7	voprov.output
ivo/icds/P/DSS2/POSSII#POSSII.N-DSS2.644	cds_coutout644	voprov.output
ivo/icds/P/DSS2/POSSII#POSSII.F-DSS2.644	cds_coutout644	voprov.output
ivo/icds/P/DSS2/POSSII#POSSII.N-DSS2.644	cds_coutout644	voprov.output

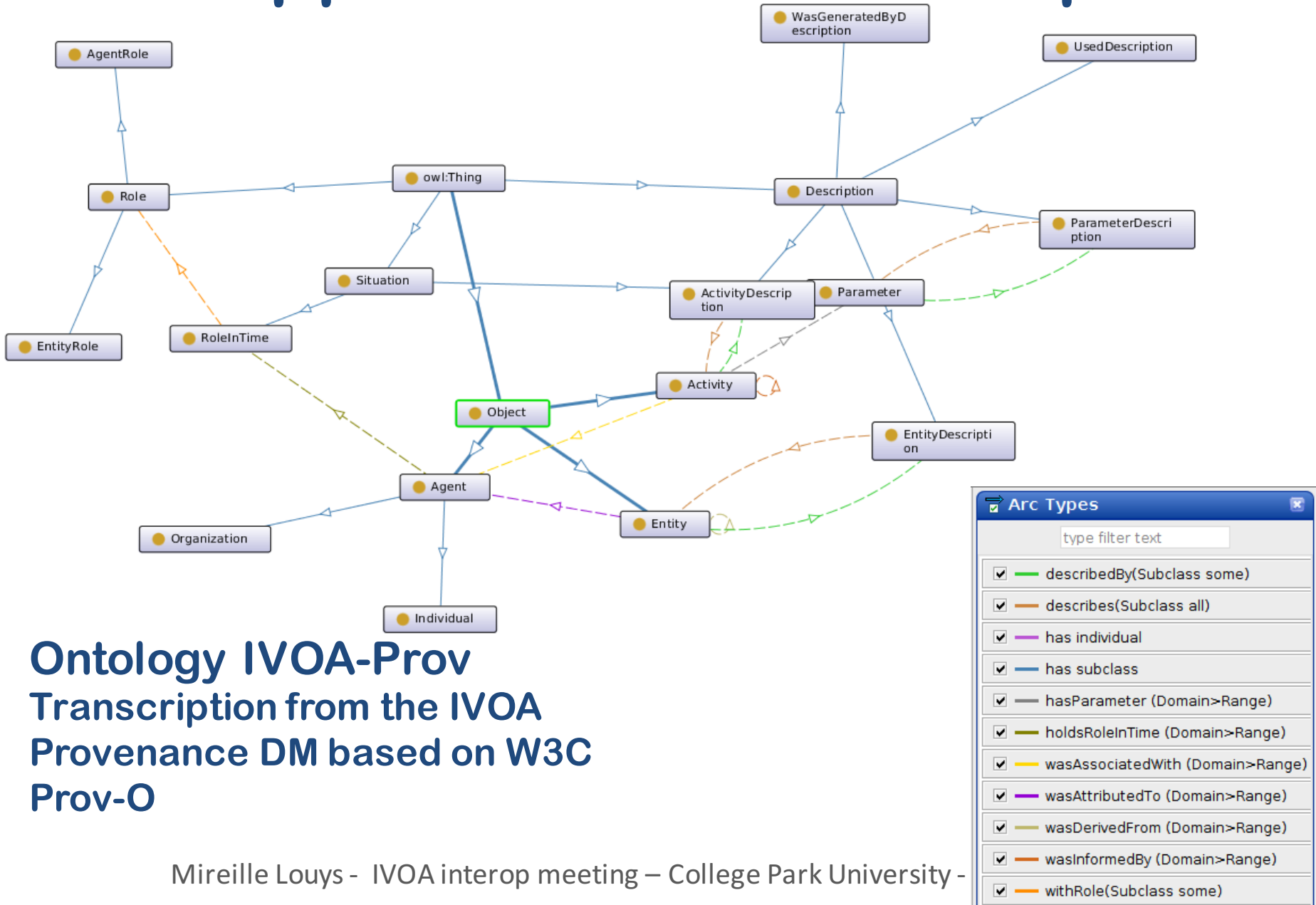
Used

head	tail	role
AlaRGB1	ivo/icds/P/DSS2/POSSII#POSSII.J-DSS2.143	voprov
AlaRGB1	ivo/icds/P/DSS2/POSSII#POSSII.F-DSS2.143	voprov
AlaRGB1	ivo/icds/P/DSS2/POSSII#POSSII.N-DSS2.143	voprov
AlaRGB2	ivo/icds/P/DSS2/POSSII#POSSII.J-DSS2.174	voprov
AlaRGB2	ivo/icds/P/DSS2/POSSII#POSSII.F-DSS2.174	voprov

WasAssociatedWith

head	tail	role
AlaRGB1	bonnarel	voprov.operator
AlaRGB1	ivo/icds	voprov.creator
AlaRGB2	bonnarel	voprov.operator
AlaRGB2	ivo/icds	voprov.creator
AlaRGB3	bonnarel	voprov.operator
AlaRGB3	ivo/icds	voprov.creator
AlaRGB4	bonnarel	voprov.operator

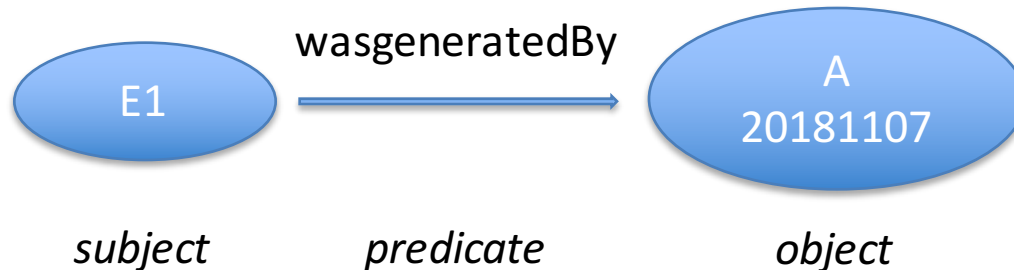
Tripplestore & BlazeGraph



Ontology IVOA-Prov
Transcription from the IVOA
Provenance DM based on W3C
Prov-O

Translation

- The table instances are translated to triples



- A change in paradigm :

The metadata organization and rules defined in the data model are translated into the ontology

- i.e. a validation of how subjects and objects can be used in predicates

<http://wiki.ivoa.net/internal/IVOA/ProvenanceRFC/provOntologyIVOA2018-v1-0.owl>

- Easy to script for the ingestion in the triplestore

Partial support for IVOA DM

Classes	Relations
Activity *	hadDescription
Activity	hadConfiguration
Entity *	used *
	wasGeneratedBy *
	wasDerivedFrom *
Agent *	wasAttributedTo
	wasAssociatedTo
ActivityDescription	hadDescription
Parameter	hadDescription
ParameterDescription	
UsedDescription	
WasGeneratedByDescription	



Triple store query in SparQL

« Give me all agents associated to an entity or to an activity which formerly used this entity named 'E' ».

```
SELECT ?name ?role ?relation WHERE {  
  { :*E* ?relation ?x .  
    ?x :refersTo ?name .  
    ?x :holdRoleInTime/:withRole ?role .  
  }  
  FILTER regex( str(?relation), "wasAttributed edwith", " i"). }  
}  
UNION  
{ ?activity :used/refersTo :*E* .  
  ?activity ?relation ?y .  
  ?y :refersTo ?name .  
  ?y :holdRoleInTime/:withRole ?role .  
  FILTER regex( str(?relation), "wasAssociatedwith", " i"). }  
}
```

Evaluation of the triplestore

- **Equivalent** Reproduces the existing classes and relations with their attributes from a relational db implementation like PostGres
- **Easy** Can encode the variety of relations and classes as (subject, predicate, object) triples (expand some nodes if needed)
- **Extensible** Extension and evolution of the knowledgebase is easy
 - The properties are coded into the instances
 - you can add a new property and handle a transition period for the data collection to grow and for the model to slightly evolve
- **Scalable**
 - A test has been performed implementing an excerpt of the Simbad Database : 8,5 millions of objects fit easily in the triplestore.

Evaluation for IVOA Prov

- Flexibility in the Triplestore to code the variety of relations and classes
 - Flat design : you can add a new property and handle a transition period for the data collection to grow and for the model to slightly evolve

- A large number of queries tested

<http://wiki.ivoa.net/internal/IVOA/ProvenanceRFC/ProvQuerytest-3store.pdf>

- The number of classes and possible relations does not matter
- Compatibility towards other efforts

Provenance Patterns DB in RDF → contribution/

