

Provenance Metadata in a Triplestore



Goal

Evaluate the triplestore database organisation for implementing the IVOA Provenance data model.

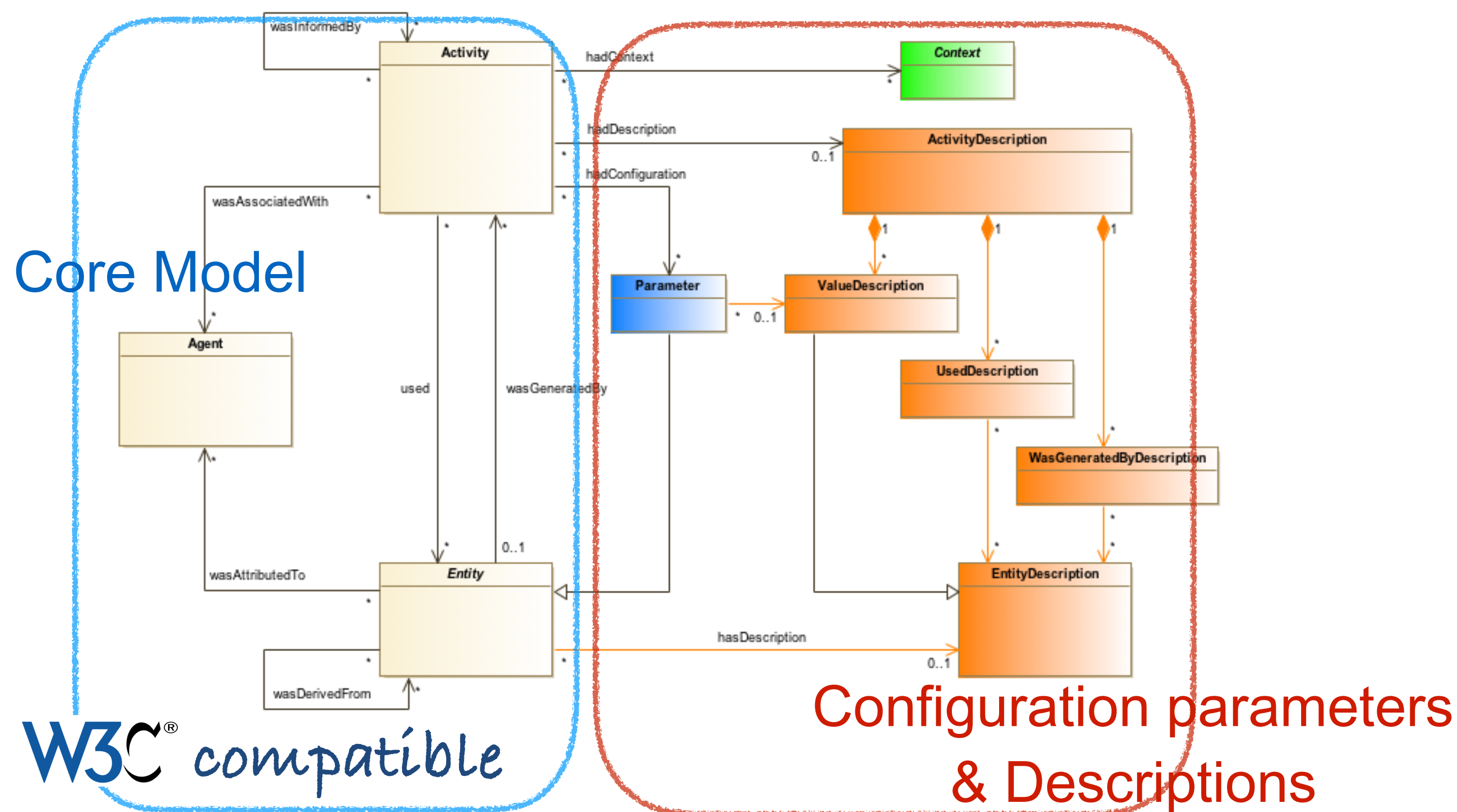
This model extends the PROV-DM defined by W3C.

In the IVOA framework, Entities typically represent data products, Activities the tasks consuming and producing Entities. Credits or responsibility is given to Agents for each Entity or Activity. Parameters and Descriptions for the methods applied in an Activity together with the roles of Entities in the scenarios are described in specific relations and classes.

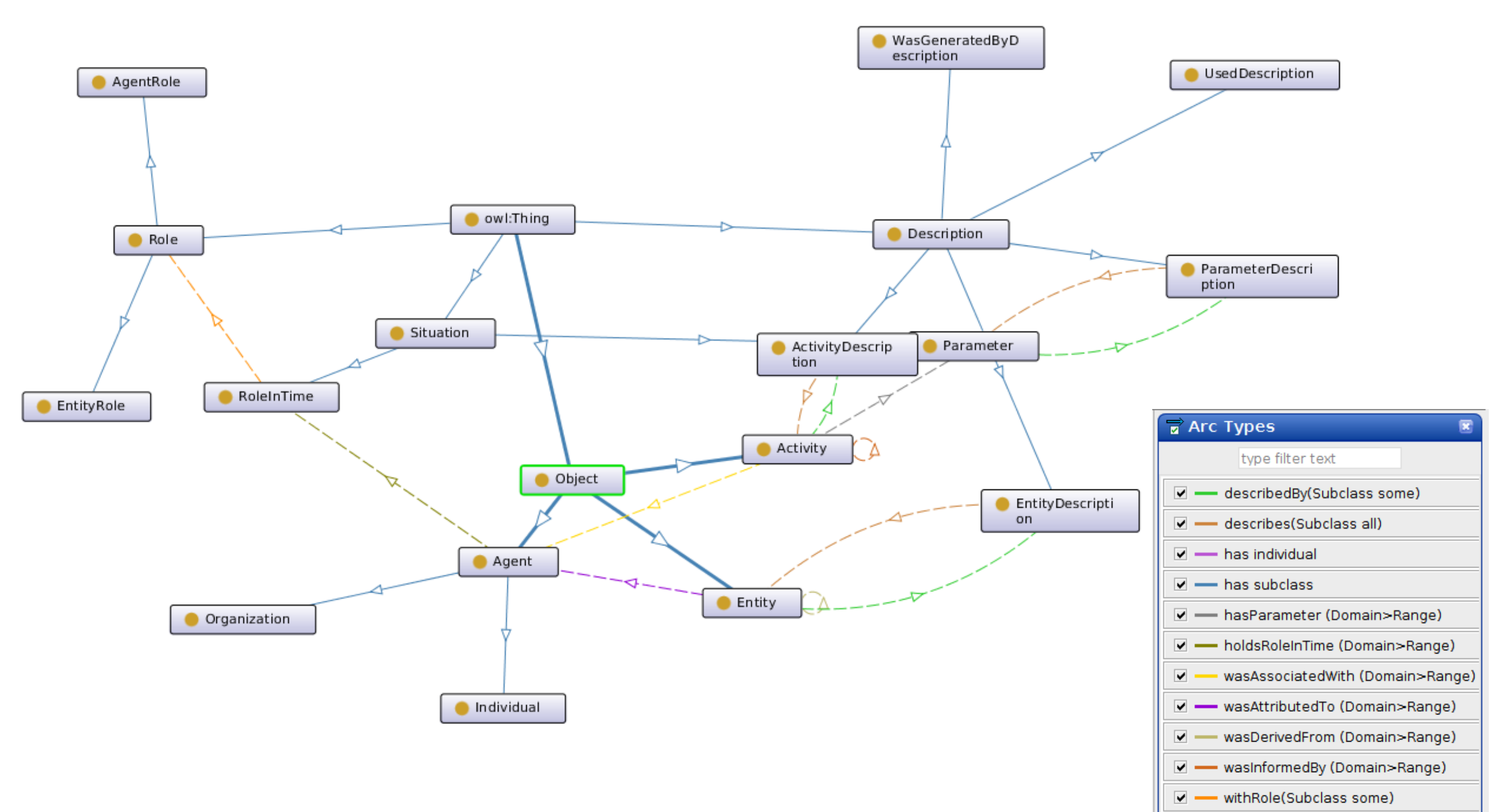
CDS Prototype Image Database

A test database tracing the processing of image data sets, from plates through files with digitization, cut-outs, RGB combination, HiPS conversion has been used for testbed. It implements IVOA Provenance DM in PostGres and supports a TAP/ADQL query interface.

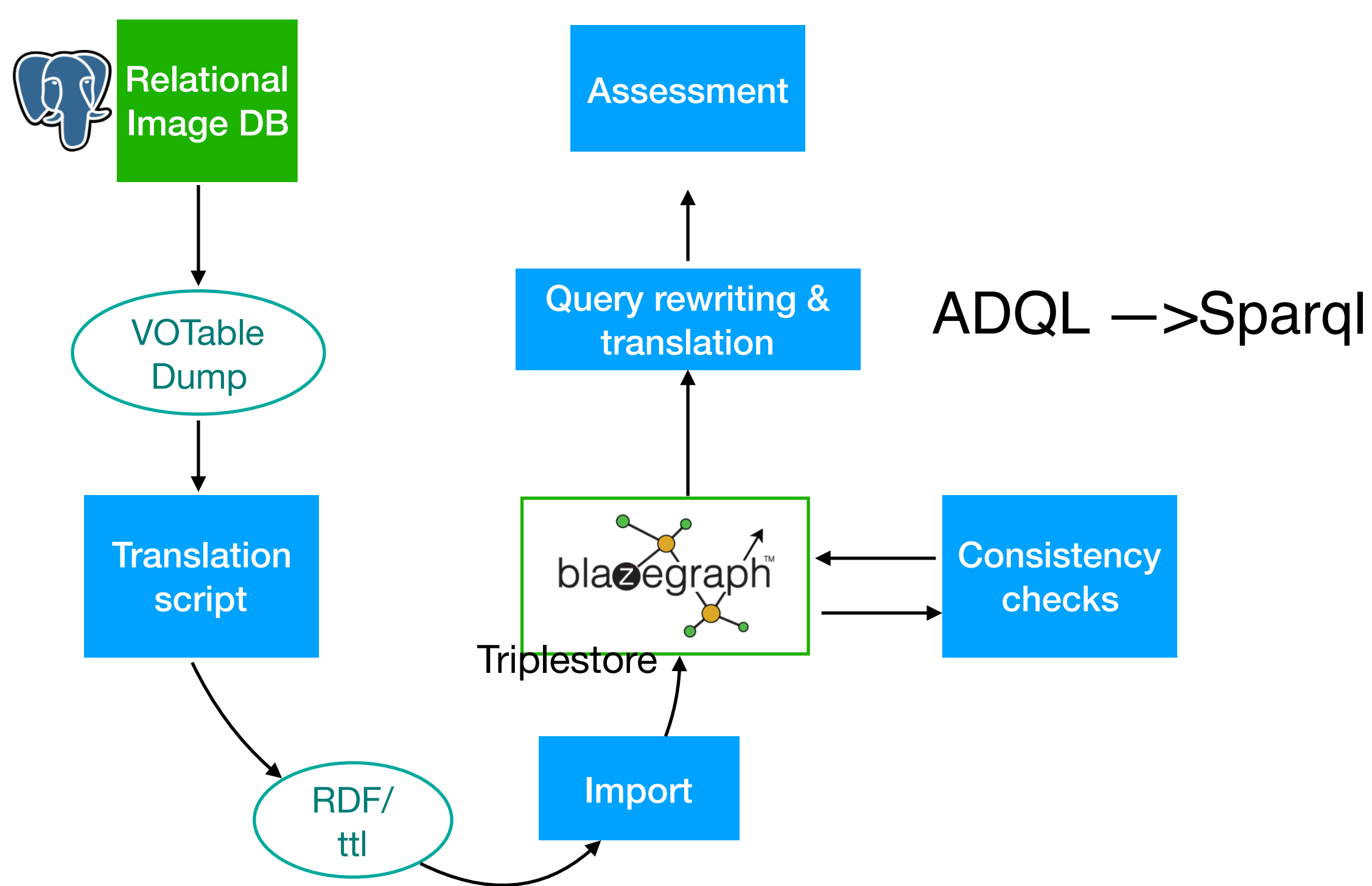
IVOA Provenance DM



PROV-O Ontology Extension



Implementation and Testing



SPARQL query testing

« 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 :holdsRoleInTime/:withRole ?role .
  }
  FILTER regex( str(?relation), "wasAttributedTo", "i") .
}
UNION
{ ?activity :used/refersTo :*E* .
  ?activity ?relation ?y .
  ?y :refersTo ?name .
  ?y :holdRoleInTime/:withRole ?role .
  FILTER regex( str(?relation), "wasAssociatedwith", "i") .
}
```

Sparql filters relations on their names and so can avoid many joins
Set of comparison queries:

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

Lessons Learned

The Triplestore RDF/ttl offers :

- **equivalent support** for queries in SPARQL compared to ADQL
- **flexibility** to code relations and add new properties
- **extensibility** if the model grows with new properties of classes/relations
- **expressibility** of searching criteria
- **scalability** very stable and efficient with many relations and instances.
- Blazegraph together with a spatial index code scale properly up to 8,5 million objects extracted from the Simbad database.

The IVOA Provenance data model can :

- circulate in **multiple serialization formats** : IVOA (VO Table) and semantic web (RDF/ttl)
- propose an interoperable framework to trace provenance info
- adjust to various compliance levels, from simple to rich descriptions
- answer a **large variety** of queries for provenance use-cases
- reuse the W3C provenance concepts with some degree of freedom and adaptability



Mireille Louys, L. Holzmann, F.-X. Pineau, F. Bonnarel

mireille.louys@astro.unistra.fr



Université

de Strasbourg