Provenance Data Model a synthetic view

Mireille Louys , CDS & ICube , Strasbourg University

François Bonnarel, CDS
Mathieu Servillat, LUTH, Paris

Michèle Sanguillon, LUPM, Montpellier

Laurent Michel, Observatoire de Strasbourg

Mark Cresitello-Dittmar CfA, Boston

Goals for an IVOA Provenance data model

A: Tracking the production history

Find out which steps were taken to produce a dataset and list the methods/tools/software that was 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

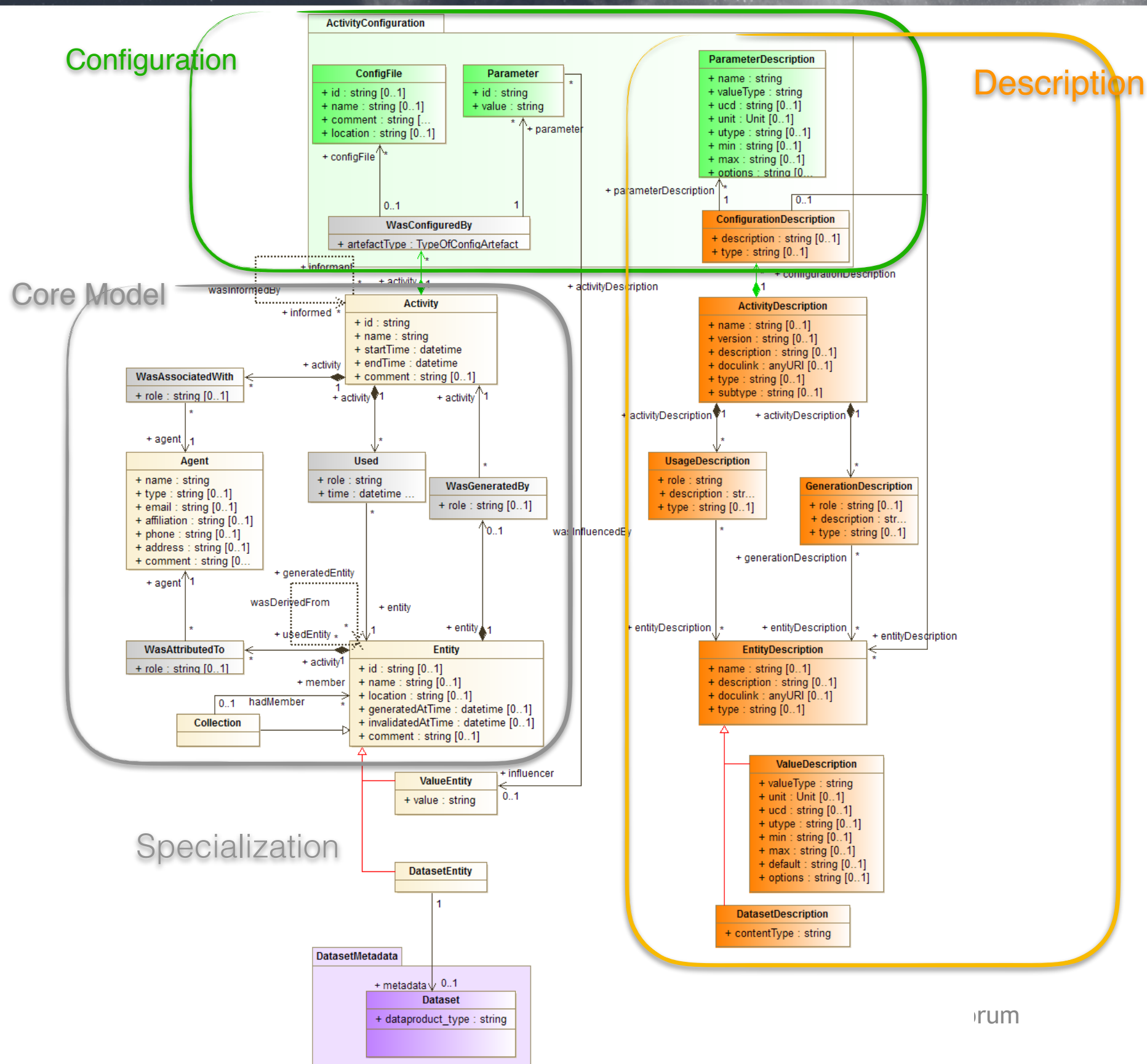This would allow one to also do a "forward search", i.e. locate derived datasets or outputs.

Time line

- Oct 2018 : Proposal for the IVOA Provenance DM standard :Proposed Recommendation
- Nov-Dec 2018: Implementation feedback for 3 major applications
  - CTA pipeline / OPUS jobs submission
  - CDS HIPS Image database
    - Prov-TAP prototype
    - Image DB prototype in Triplestore
  - Validation of the Activity configuration needs
    - searchable parameters
    - factorized description for Activities and Entities and their usage
    - definition for a Provenance Vocabulary
- Jan-Feb 2019
  - Working examples : ProvFocusAstericsExamples
  - Revision of Data model Class Diagram ProvFocusAsterics
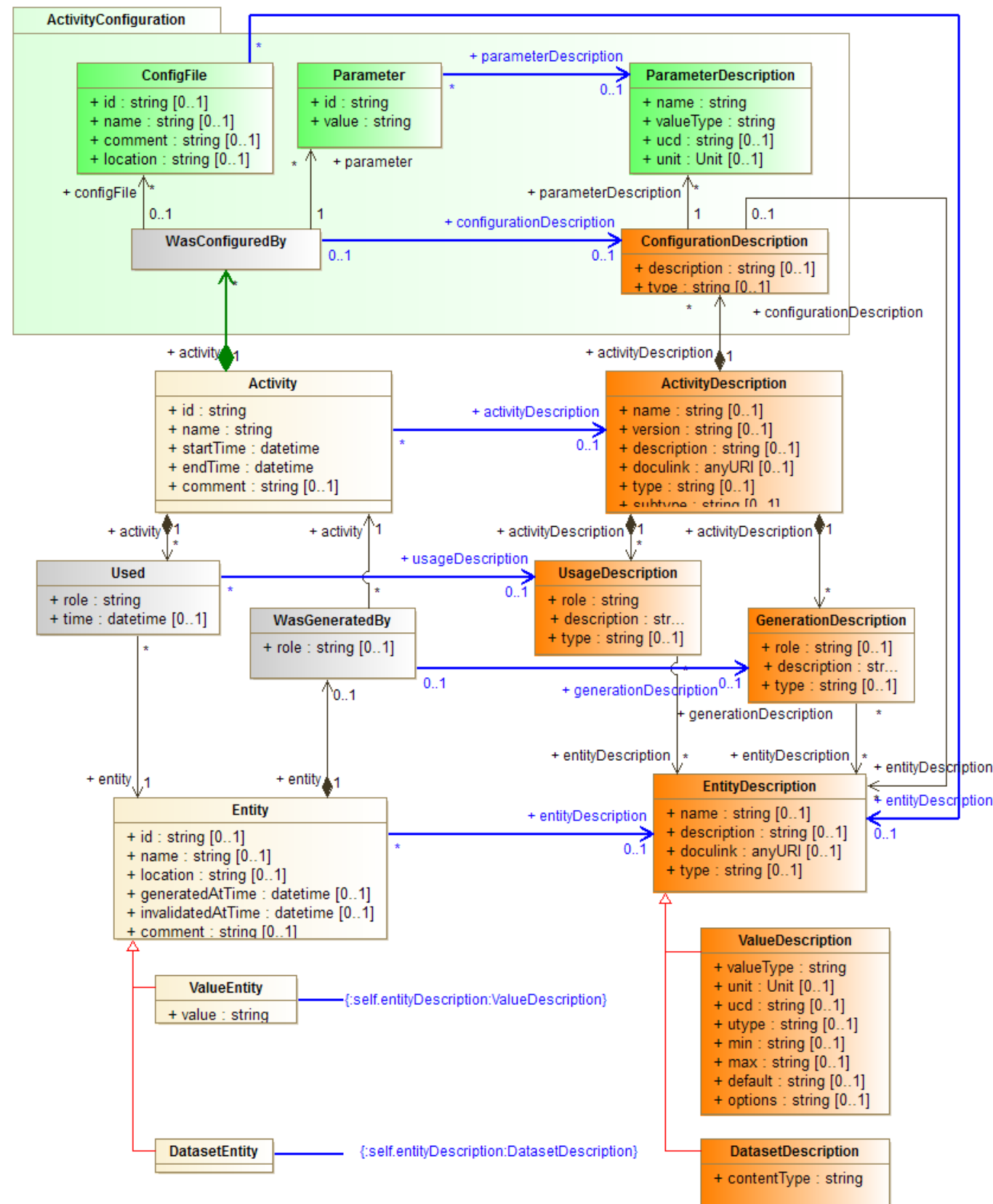 - Document revision for the 2nd iteration towards PR

Reorganizing the DM parts

Working examples

- CTA Pipe : OPUS job submission and execution (LUTH)

- HiPS Image database sample (CDS)

- Cube segmentation example (CDS)

- See the details on
https://wiki.ivoa.net/twiki/bin/view/IVOA/ProvFocusAstericsExamples

☐ Implementations

- Ready :
  - CTA Pipe : OPUS job submission and execution (LUTH). VOTable, JSON
  - Image database prototype in Triplestore (CDS)   RDF
  - HiPS Image database (CDS)
    - PROV-TAP (Table access protocol)   VOTable
  - RAVE implementation (AIP, Postdam)
    - Simple access (Prov-SAP) prototype Prov-N, PROV-JSON
  - Provenance for Pollux DB & *voprov* library (LUPM) VOTable, Prov-N, JSON
- Under study :
  - CTA pipe implementation for raw data DL0 - DL1 / LUPM Montpellier
  - SVOM pipeline execution tracking   JSON

Feature coverage

DM Feature/ Project	Institute	Core	ValueEntity/ DatasetEntity	Parameter and configuration	Descripttion
CTA Opus	LUTH paris	x	x	x	full
CDS Hips database	CDS	x	x	x	full
Image DB Triplestore	CDS	x	x	x	full
Pollux	LUPM Montpellier	x			Activity only
RAVE AIP	IAP Postdam	x			Activity only
MuseWise prov Interface	IAP Postdam	x			Part

Deliverables for DADI WP

- **Feb 2019** Model Documentation from Modelio
- **March 2019** IVOA Proposed Recommendation document PR v2
- **March 2019** VODML xml representation of the model
- **March 2019** Example of serialization instances in VOTable, RDF, JSON
- **April 2019** Prov-TAP WD and implementation at CDS
- **April 2019** Provenance DM Implementation Note: How we implemented the data model on various databases

Project follow-up

- Documented on the IVOA wiki
- https://wiki.ivoa.net/twiki/bin/view/IVOA/ObservationProvenanceDataModel
  - Provenance days
  - IVOA meetings twice a year

- Interaction with the IVOA DAL group
  - Provenance simple access protocol
  - Provenance TAP protocol

Connection to other projects

- CTA pipeline integration into *ctapipe* software suite

- Contribution to Research Data Alliance
    - WG project Provenance Patterns DB
        - Image database prototype in Triplestore (CDS)

Thanks

Questions ?
Comments ?