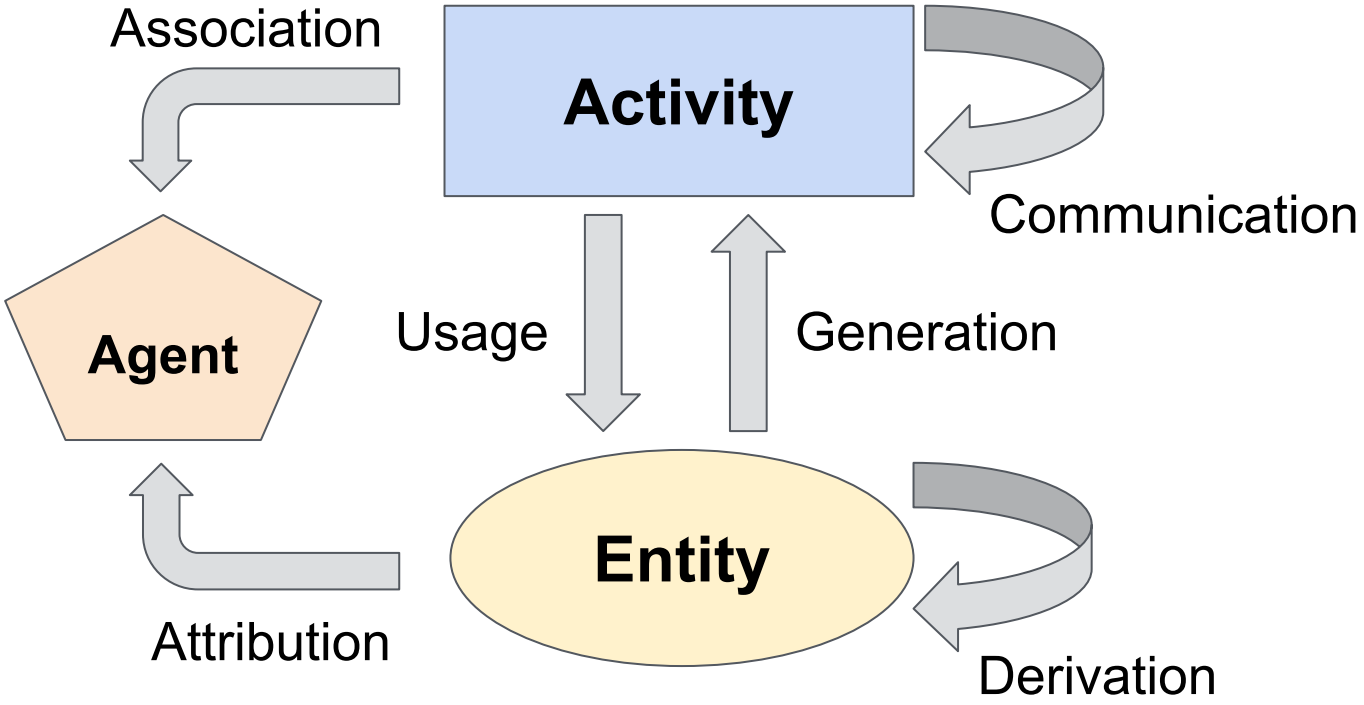
One-Step Provenance

IVOA Sydney 2024

Mathieu Servillat (LUTH - Observatoire de Paris / CNRS)
Catherine Boisson, François Bonnarel, Mireille Louys
+ ESCAPE participants
+ CTA members

Provenance glossary

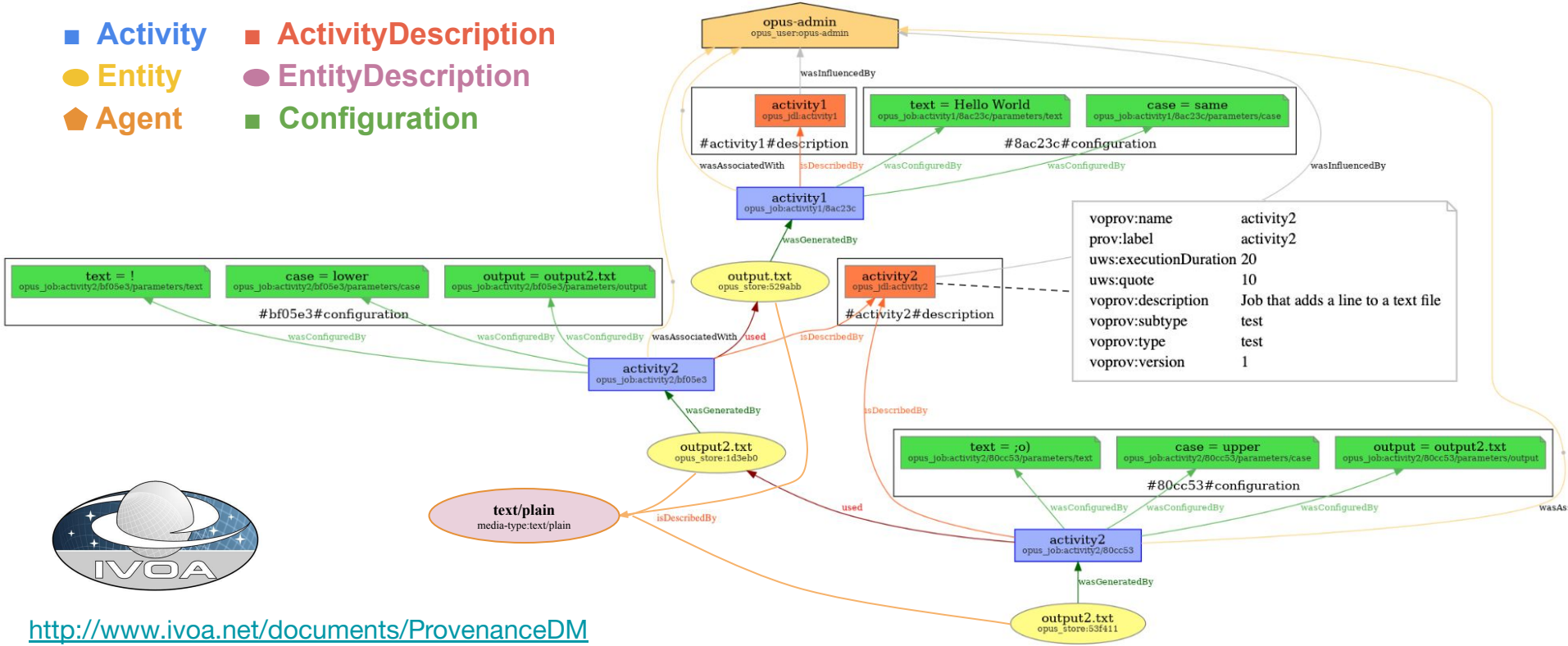


Word Wide Web Consortium

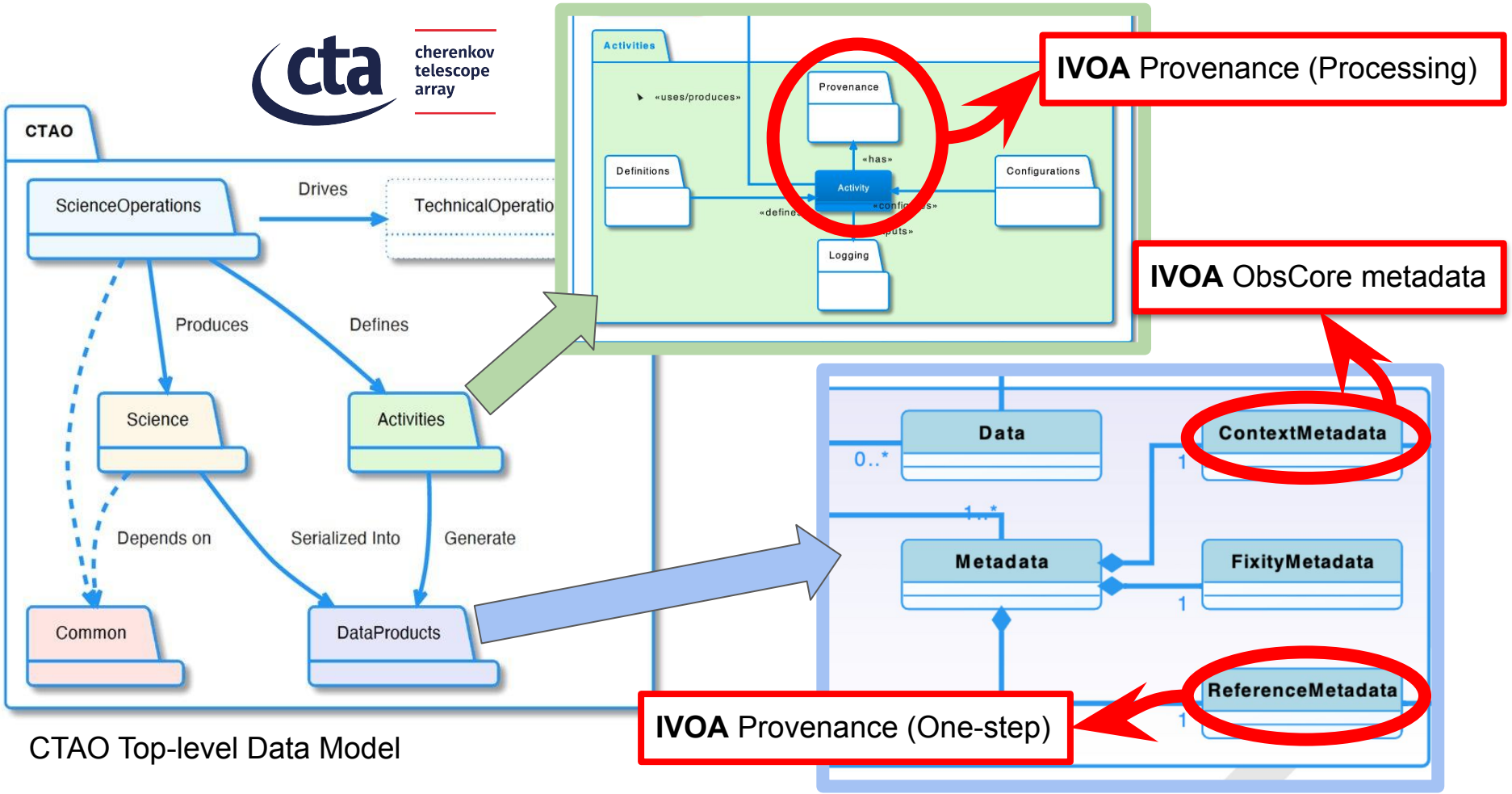
http://www.w3.org/TR/prov-overview

Full IVOA Provenance graph

- Activity
- ActivityDescription
- Entity
- EntityDescription
- Agent
- Configuration

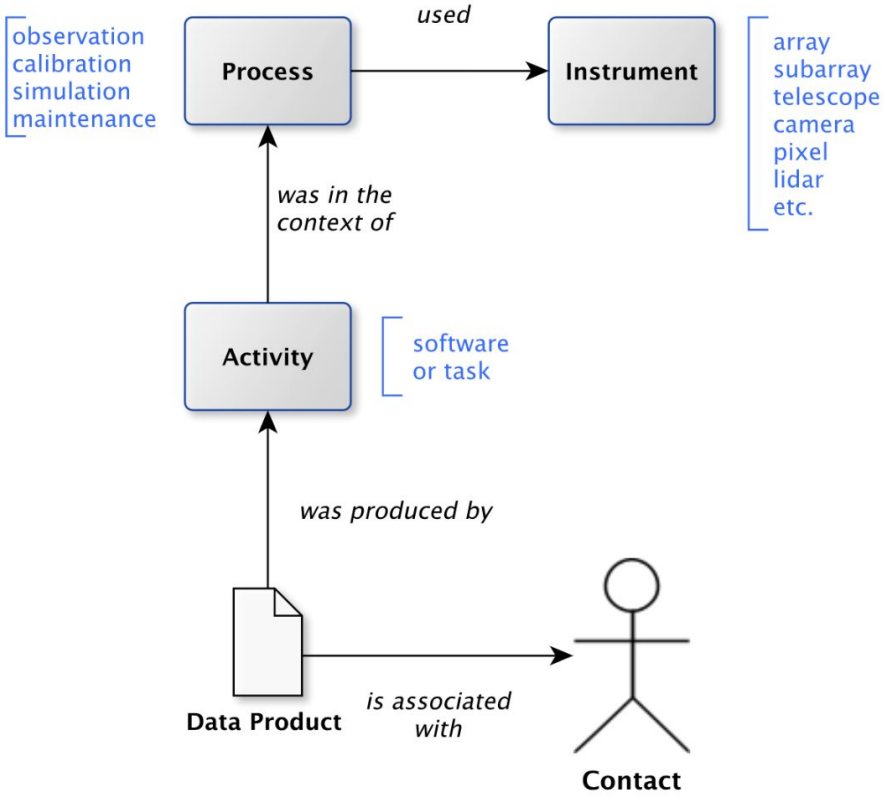


http://www.ivoa.net/documents/ProvenanceDM



CTAO Top-level Data Model

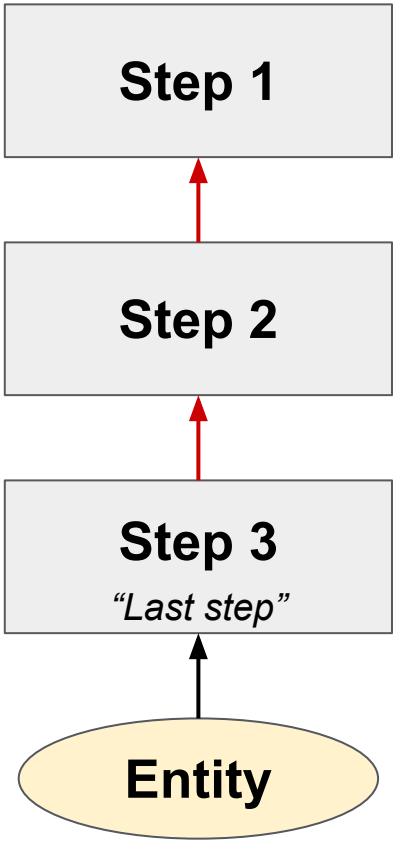
CTAO Reference Metadata (draft)



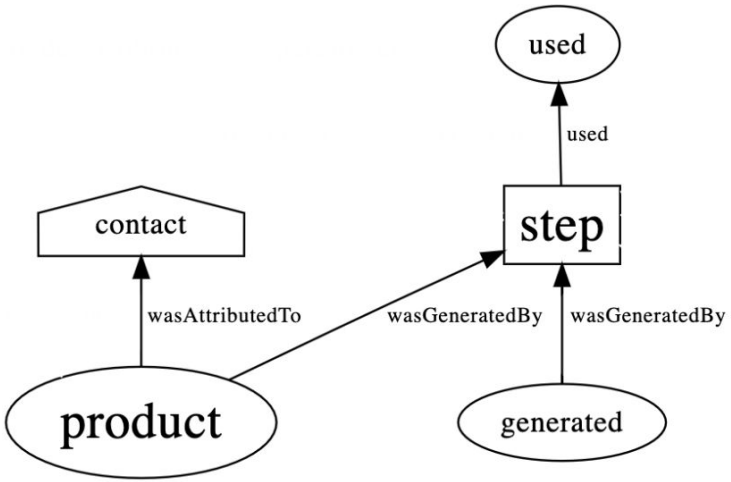

```
HIERARCH CTA METADATA VERSION = 1.0.0
HIERARCH CTA CONTACT ORGANIZATION = CTAO
HIERARCH CTA CONTACT NAME = CTAO Support
HIERARCH CTA CONTACT EMAIL = support@cta-observatory.org
HIERARCH CTA PRODUCT DESCRIPTION = Cat C DL3 event list
HIERARCH CTA PRODUCT CREATION_TIME = 2018-11-10 15:30:00
HIERARCH CTA PRODUCT ID = TBD
HIERARCH CTA PRODUCT DATA CATEGORY = C
HIERARCH CTA PRODUCT DATA LEVEL = DL3
HIERARCH CTA PRODUCT DATA TYPE = Event
HIERARCH CTA PRODUCT DATA ASSOCIATION = Subarray
HIERARCH CTA PRODUCT DATA MODEL NAME = open-gamma-astro
HIERARCH CTA PRODUCT DATA MODEL VERSION = v0.2
HIERARCH CTA PRODUCT DATA MODEL URL = https://github.com/open-gamma-
HIERARCH CTA PRODUCT FORMAT = fits
HIERARCH CTA PROCESS TYPE = observation
HIERARCH CTA PROCESS SUBTYPE = standard
HIERARCH CTA PROCESS ID = 12345
HIERARCH CTA ACTIVITY NAME = pipeline-stage-3
HIERARCH CTA ACTIVITY TYPE = software
HIERARCH CTA ACTIVITY ID = 5367fcf8-e75f-11e8-9692-3c15c2d6877e
HIERARCH CTA ACTIVITY START = 2018-11-10 15:24:11
HIERARCH CTA ACTIVITY END = 2018-11-10 15:29:00
HIERARCH CTA ACTIVITY SOFTWARE NAME = ctapipe
HIERARCH CTA ACTIVITY SOFTWARE VERSION = 2.2.0
HIERARCH CTA INSTRUMENT SITE = CTA-South
HIERARCH CTA INSTRUMENT CLASS = subarray
HIERARCH CTA INSTRUMENT TYPE = standard
HIERARCH CTA INSTRUMENT SUBTYPE =
HIERARCH CTA INSTRUMENT ID = MST20
```

Back to IVOA Provenance

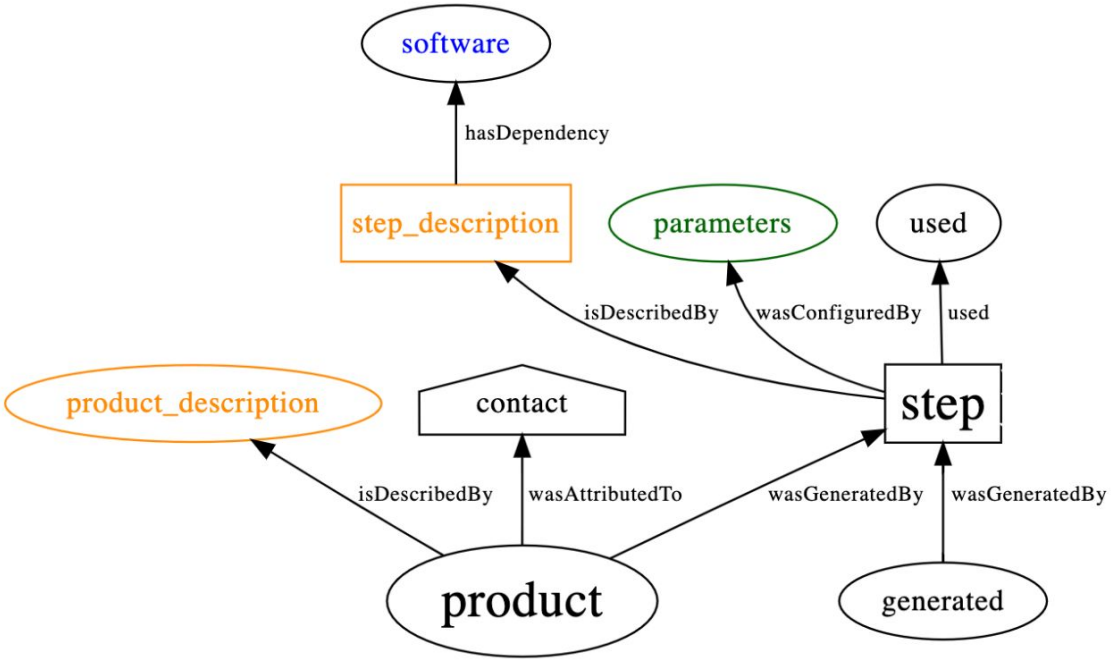
- Use cases
  - CTA data products header
  - Workshops with ESCAPE partners, within ASOV
  - ADASS BoF session in 2020
    https://hal.science/obspm-03343907v1
- Requirements
  - Applied to digital object generation
  - Cite **software**
  - Record the **context of execution**
  - Include provenance attributes **inside** an entity
  - Handle different levels of **details**
- Definition of a “One-Step Provenance”
  - **List of attributes** to describe one step of data generation
  - Links between steps using **identifiers**
  - The “Last step” may be embedded in the entity
  - **Subgraph** designed with IVOA Provenance concepts



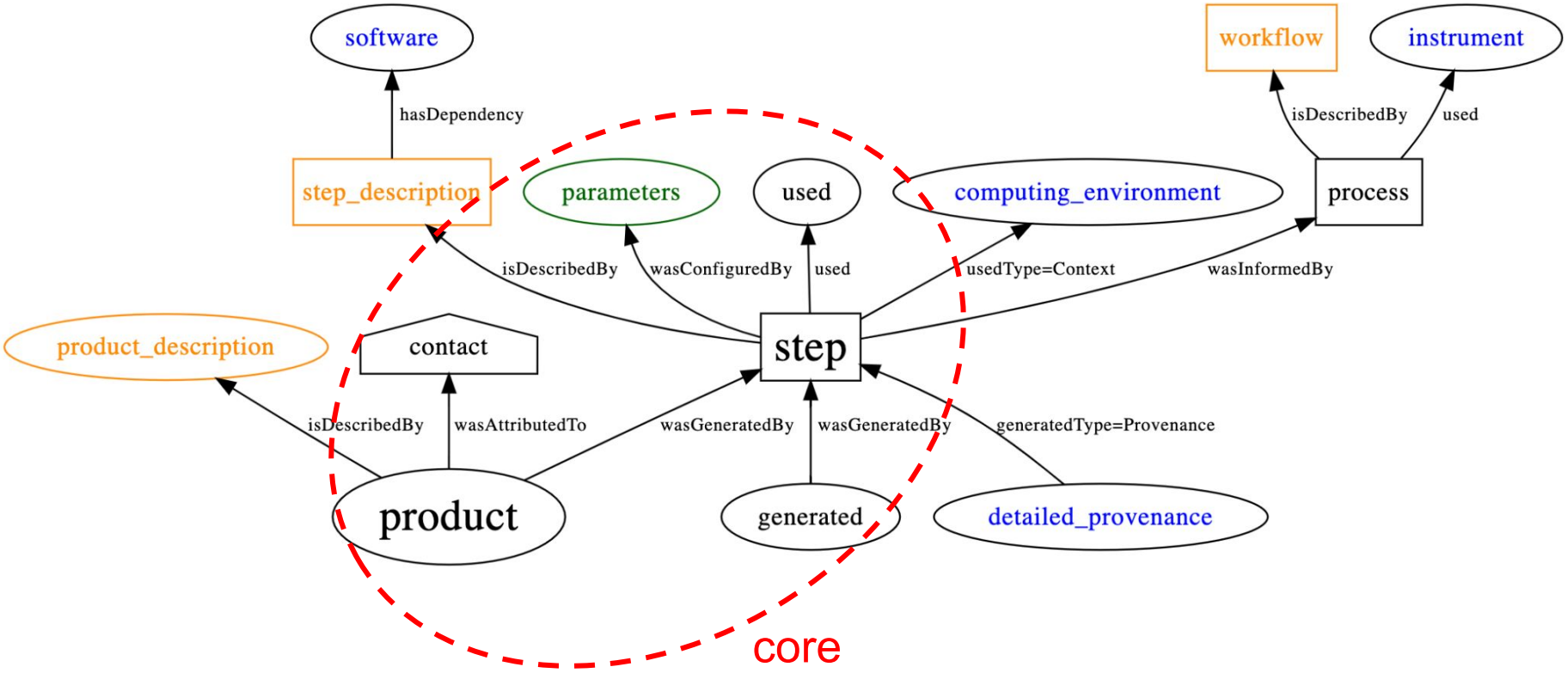
One-Step Provenance Data Model - Minimum



One-Step Provenance Data Model - Descriptions

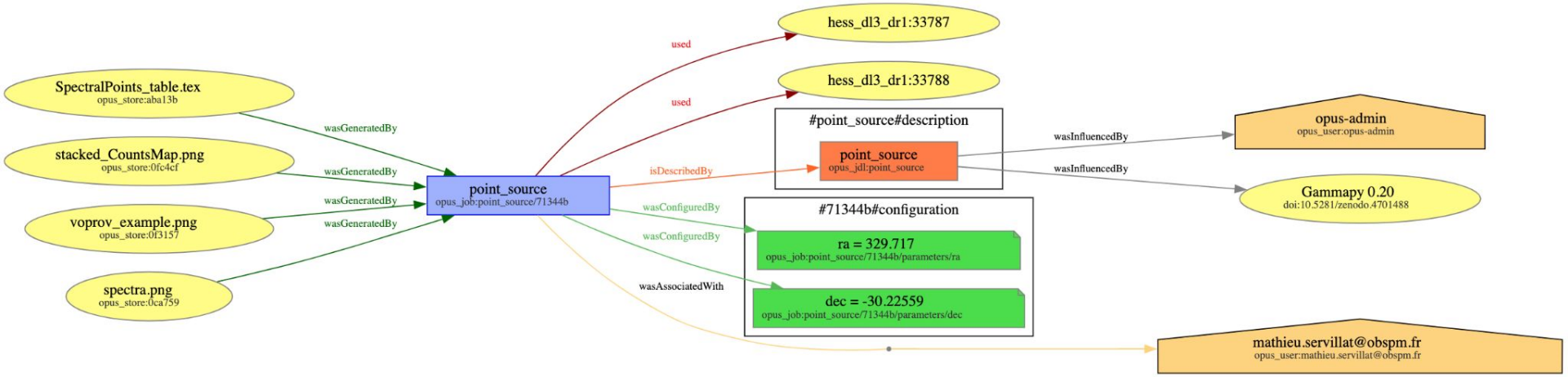


One-Step Provenance Data Model - Context



Example of implementation: OPUS

- Analysis of Cherenkov data with gammapy using OPUS
  - OPUS: job manager based on UWS (https://opus-job-manager.readthedocs.io)
  - job definition = expected input/output + configuration parameters + description
  - using the **`voprov`** Python package



Example of implementation: voprov

- **Extension of voprov Python package**
  - Thanks to Benjamin Parciany (internship)
  - **User friendly functions** :
    add_used entity()
    add_activity_description()
    add_agent()
    …
  - **Dedicated add_one_step() function**
    builds the subgraph
    from a dictionary

```
onestep_1 = {
    "product_id": "obs:image1b",
    "product_role": "bias-subtracted-image",
    "contact_id": "mservillat",
    "step_id": "ps:2459",
    "step_name": "ps:bias_subtraction",
    "step_parameters": {"threshold": 10, "skip
    "step_description": "remove the readout no
    "step_software": ["gammapy_v1.2"],
    "used_ids": ["obs:image1", "obs:bias"],
    "generated_ids": [],
    "process_id": "7531",
    "workflow_name": "ImageCalibration",
    "instrument_id": "CTAO",
}
```

```
pdoc.add_one_step(onestep_1)
```

```
<VOProvDataSetEntity: obs:image1b>
```

Example of implementation: ProvTAP

last step provenance in ProvTAP

- View (in postgres)

```
create view one_step_provenance as select
e.e_id AS product_id,  e.e_name AS product_name, e.e_location AS product_location, e.e_generated AS
product_generated, e.e_description as product_description,
   e.e_invalidated AS product_invalidated, e.e_comment AS product_comment,
 activity.a_name AS step_name, activity.a_starttime AS step_starttime,      activity.a_endtime AS step_endtime,
activity.a_comment AS step_comment, a.a_description as step_description,
 wasattributedto.wat_role AS contact_role, agent.ag_name AS contact_name, agent.ag_type AS contact_type,
   agent.ag_affiliation AS contact_affiliation, agent.ag_email AS contact_email, agent.ag_address AS contact_address,
   agent.ag_phone AS contact_phone, agent.ag_comment AS contact_comment,
string_agg(used.u_entity::text, ','::text) AS used_entities_list
FROM entity e
 JOIN wasgeneratedby ON e.e_id::text = wasgeneratedby.wgb_entity::text
 JOIN activity ON wasgeneratedby.wgb_activity::text = activity.a_id::text
join used on u_activity = a_id
join entity as ee on ee.e_id = u_entity
join wasattributedto on wat_entity = e.e_id
join agent on ag_id = wat_agent ;
```

@ François Bonnarel - DAL1
https://wiki.ivoa.net/internal/IVOA/InterOpMay2024DAL/ProvTAPEvolution.pdf