

# Last-step provenance



*IVOA virtual 2022-04*

**Mathieu Servillat** (LUTH - Observatoire de Paris / CNRS)

Catherine Boisson, François Bonnarel, Mireille Louys, Michèle Sanguillon

+ ESCAPE participants

+ CTA members

# From F-A-I to FAIR

→ **ADASS XXXI** <https://hal.archives-ouvertes.fr/obspm-03516688v1>  
“FAIR high level data for Cherenkov astronomy”

**Findable**  
**Accessible**  
**Interoperable**  
**Reusable?**

- **Findable-Accessible-Interoperable**

- Use the **Virtual Observatory standards**, protocols and services
- Define community **standards** where required
- To be discussed early in projects, but **technical solutions exist**

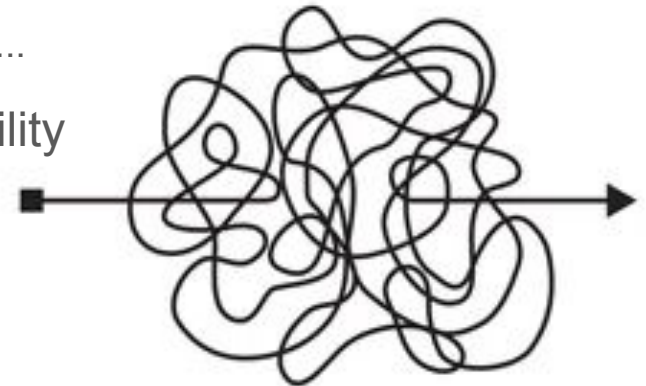
- **Reusability?**

- Based on the **quality / reliability / trustworthiness** of the products
- What calibration was applied? What tools were used and how?  
What assumptions were made during the data preparation?
- **Sustainability**: with time, key information may disappear...



- **Provenance** information as an answer to reusability

- Need for the **origin, trace**, and detailed manipulations
- Need to **structure** this information
- Need to **keep** it and **link** it to the data
- IVOA Provenance data model!



# IVOA Provenance

- **IVOA Provenance data model**

- Recommendation in April 2020, based on W3C PROV

- **ProvSAP: simple access protocole**

- Request provenance graph for an identifier (entity/activity)
- Takes advantage of the W3C serializations (JSON, XML, SVG, PNG...)
- `prov/voprov` Python package
- Implementations : Pollux, OPUS (CTA, MASER, CompOSE), ...

<https://voparis-uws-test.obspm.fr/client/proxy/provsap?ID=a6018b&DESCRIPTIONS=1&CONFIGURATION=1&ATTRIBUTES=0>

- **ProvTAP: table access protocole** (→ *DAL talk by François*)

- TAP Schema
- simplified views
- Implementations : HiPS tiles provenance, ...

International Virtual  
Observatory Alliance



IVOA Documents

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

IVOA Provenance Data Model  
Version 1.0

IVOA Recommendation 11 April 2020

Interest/Working Group:

<http://www.ivoa.net/wiki/bin/view/IVOA/IvoaDataModel>

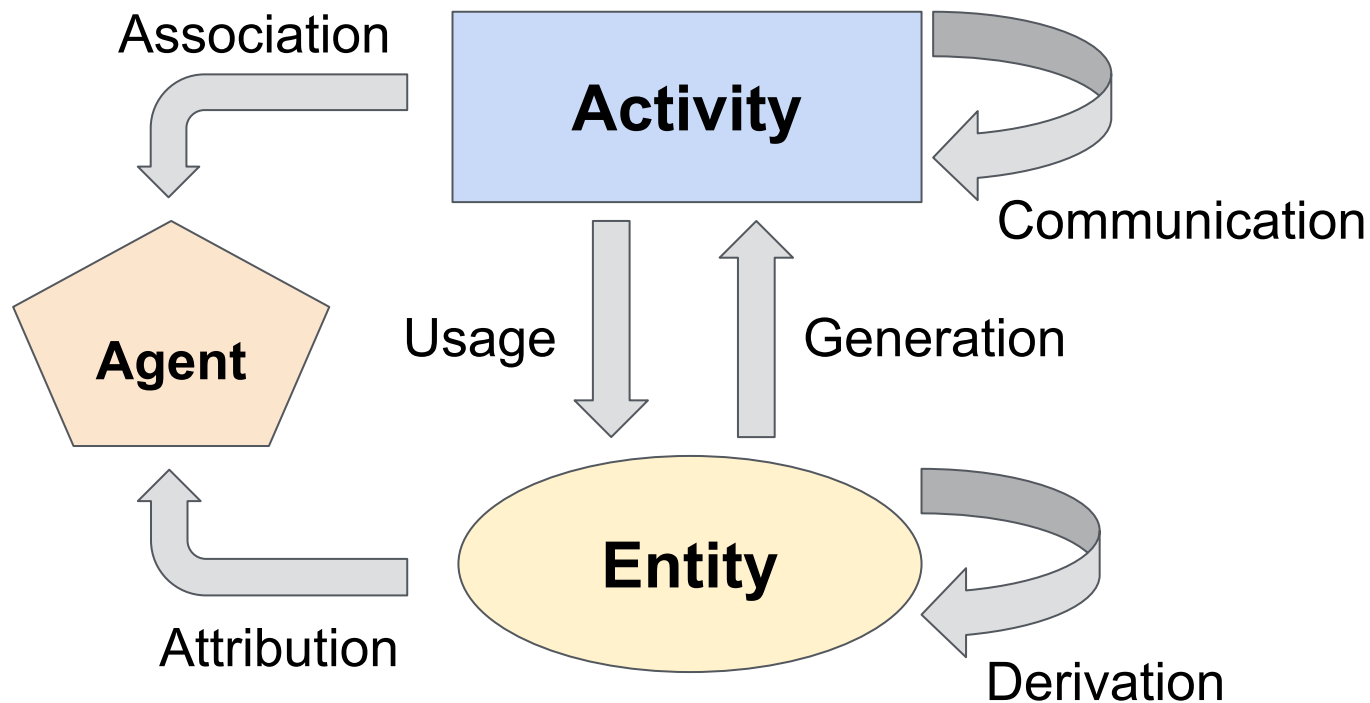
Author(s):

Mathieu Servillat, Kristin Riebe, Catherine Boisson, François Bonnarel, Anastasia Galkin, Mireille Louys, Markus Nullmeier, Nicolas Renault-Tinacci, Michèle Sanguillon, Ole Streicher

Editor(s):

Mathieu Servillat

# Provenance glossary

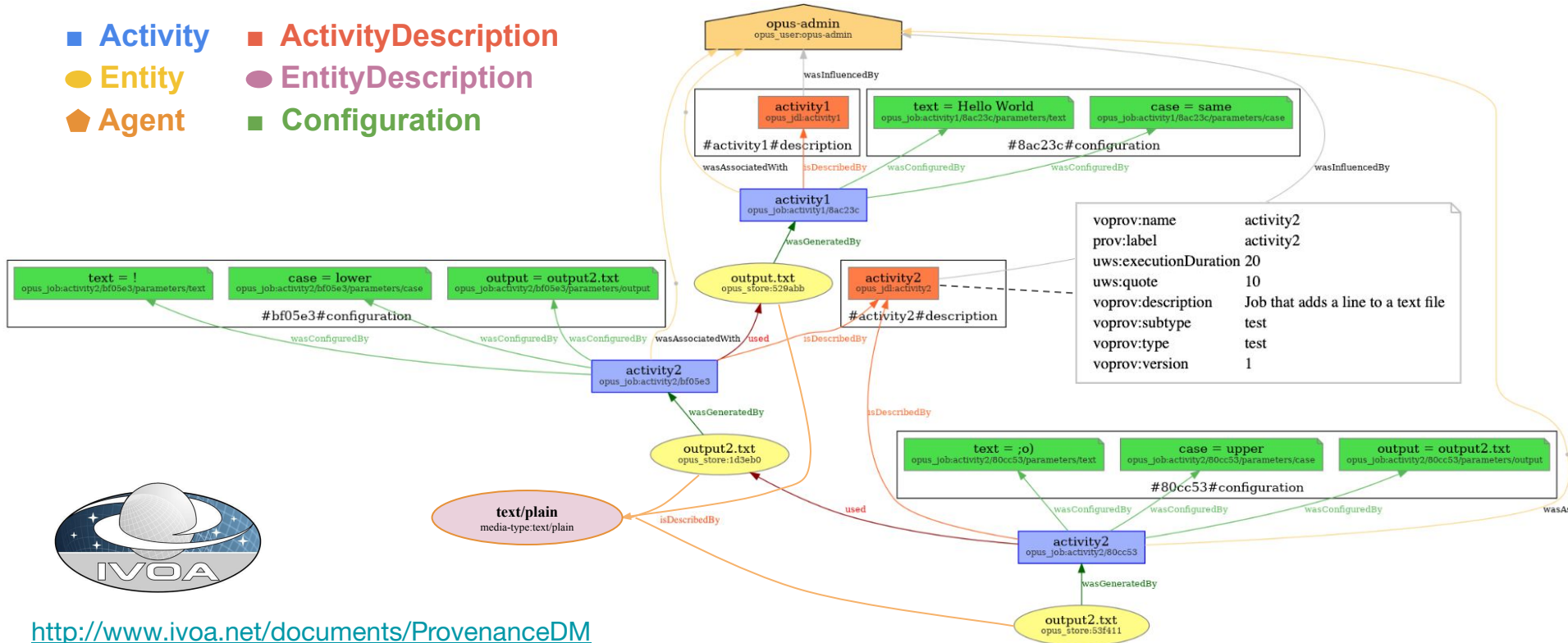


Word Wide Web Consortium

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

# Full IVOA Provenance graph

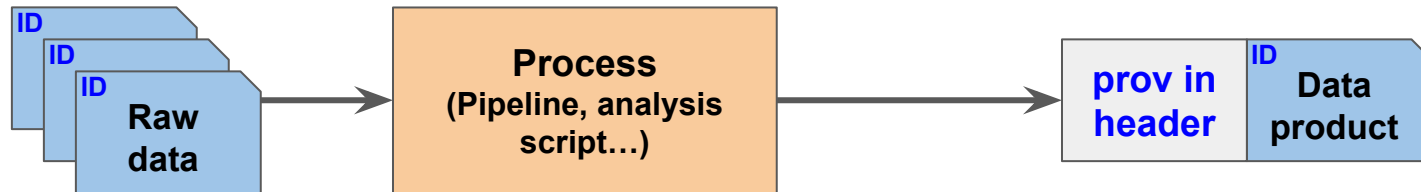
- Activity
- ActivityDescription
- Entity
- EntityDescription
- ⬠ Agent
- Configuration



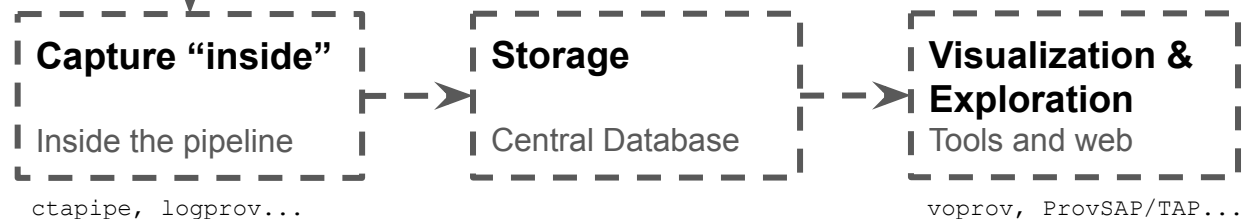
<http://www.ivoa.net/documents/ProvenanceDM>

# A provenance management system

- What scientists generally have in mind:



- But need for advanced provenance management:



Provenance Week 2021 proceedings : <https://arxiv.org/abs/2109.07751>

# Some terminology

- **full provenance**: graph/tree/chain that **traces** activities and entities up to the raw data. This information is not hosted by the entities themselves, it should be stored in a central database, or as separate files.
- **end-user/specific “provenance”**: can be embedded into an entity, keywords or data that provides project specific **key information to use/analyse** the entity (e.g. for CTA: event class/type, telescope configuration, sky conditions, reco method...)
- **last-step provenance**: embedded into an entity as a list of keywords that gives some context and info on **last activity** (general workflow, software, versions, contact...), including the list of generated and used entity ids, so that a full provenance may be reconstructed from this minimum provenance.

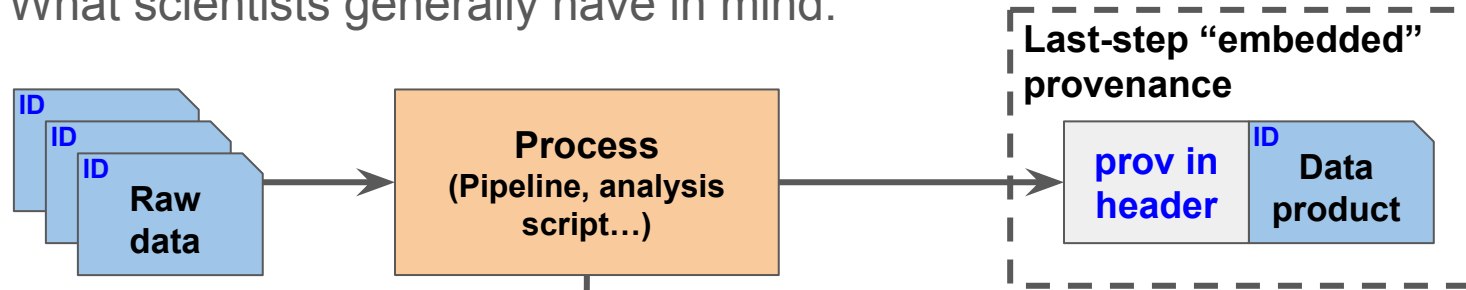
See ADASS XXX BoF proceedings : <https://arxiv.org/abs/2101.08691>

ESCAPE workshop on provenance : <https://indico.in2p3.fr/event/21913/page/2641-summary>

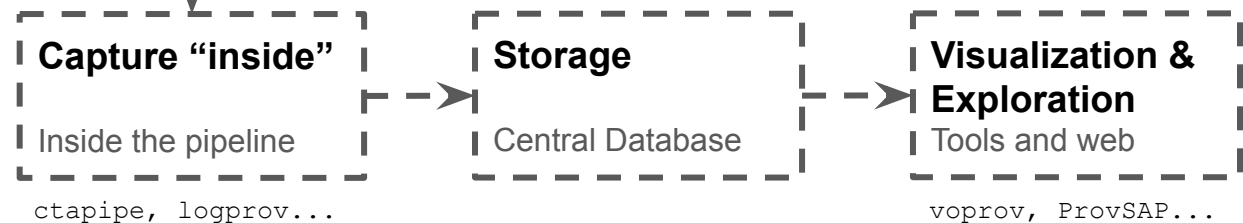
ASOV workshop on provenance : <https://indico.obspm.fr/event/1267>

# A provenance management system

- What scientists generally have in mind:



- But need for advanced provenance management:



Provenance Week 2021 proceedings : <https://arxiv.org/abs/2109.07751>



# Last-step provenance - IVOA Note

[https://wiki.ivoa.net/internal/IVOA/InterOpApr2022DM/2022-04-14\\_IVOA\\_Note\\_Last\\_step\\_provenance.pdf](https://wiki.ivoa.net/internal/IVOA/InterOpApr2022DM/2022-04-14_IVOA_Note_Last_step_provenance.pdf)

- Problematic

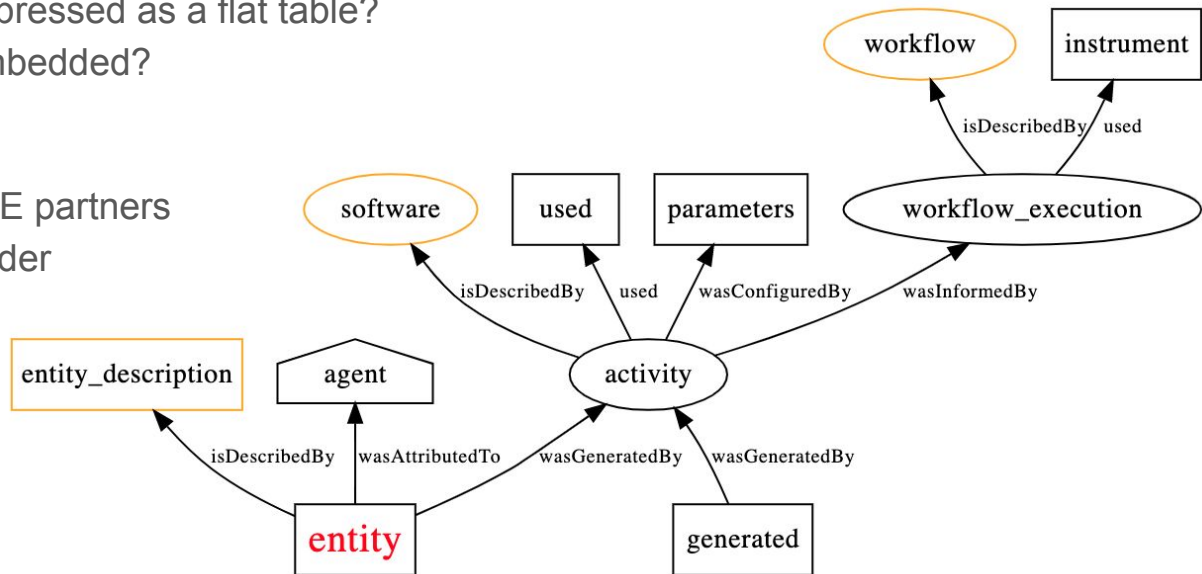
- Provenance graphs are complex, cannot be embedded in entities
- Is there a minimum provenance?
- Can provenance be expressed as a flat table?
- Can provenance be embedded?

- Use cases

- Workshop with ESCAPE partners
- CTA data products header

- Content

- 1 chain link
- subgraph
- keyword list
- FITS keywords



# Last-step provenance - keyword list - entity/agent

| keyword                | UType                                  | FITS keyword | Alternative |
|------------------------|--|--------------|-------------|
| entity_id              | <a href="#">Entity.id</a>              | ENT_ID       |             |
| entity_location        | Entity.location                        | ENT_LOC      |             |
| entity_generatedAtTime | Entity.generatedAtTime                 | ENT_GTIM     |             |
| entity_name            | <a href="#">EntityDescription.name</a> | ENT_NAME     |             |
| entity_type            | EntityDescription.type                 | ENT_TYPE     |             |
| entity_content_type    | EntityDescription.content_type         | ENT_CTYP     |             |
| entity_docurl          | EntityDescription.docurl               | ENT_DURL     |             |
| entity_comment         | Entity.comment                         | ENT_COMM     |             |
| agent_id               | <a href="#">Agent.id</a>               | AGT_ID       |             |
| agent_name             | <a href="#">Agent.name</a>             | AGT_NAME     |             |
| agent_type             | Agent.type                             | AGT_TYPE     |             |
| agent_email            | Agent.email                            | AGT_MAIL     |             |

# Last-step provenance - keyword list - context

| keyword             | UType                                    | FITS keyword | Alternative |
|---------------------|--|--------------|-------------|
| workflow_id         | <a href="#">Activity.id</a>              | WKF_ID       |             |
| workflow_name       | <a href="#">ActivityDescription.name</a> | WKF_NAME     |             |
| workflow_type       | ActivityDescription.type                 | WKF_TYPE     |             |
| workflow_comment    | Activity.comment                         | WKF_COMM     |             |
| instrument_id       | <a href="#">Entity.id</a>                | INS_ID       |             |
| instrument_location | Entity.location                          | INS_LOC      |             |
| instrument_name     | <a href="#">EntityDescription.name</a>   | INS_NAME     |             |
| instrument_type     | EntityDescription.type                   | INS_TYPE     |             |
| instrument_docurl   | EntityDescription.docurl                 | INS_DURL     |             |
| instrument_comment  | EntityDescription.comment                | INS_COMM     |             |

# Last-step provenance - keyword list - activity

| keyword              | UType                                    | FITS keyword | Alternative |
|----------------------|--|--------------|-------------|
| software_name        | <a href="#">ActivityDescription.name</a> | SFW_NAME     |             |
| software_version     | ActivityDescription.version              | SFW_VERS     |             |
| software_description | ActivityDescription.description          | SFW_DESC     |             |
| software_type        | ActivityDescription.type                 | SFW_TYPE     |             |
| software_docurl      | ActivityDescription.docurl               | SFW_DURL     |             |
|                      |  |              |             |
| activity_id          | <a href="#">Activity.id</a>              | ACT_ID       |             |
| activity_name        | <a href="#">Activity.name</a>            | ACT_NAME     |             |
| activity_start_time  | Activity.startTime                       | ACT_STIM     |             |
| activity_end_time    | Activity.endTime                         | ACT_ETIM     |             |
| activity_comment     | Activity.comment                         | ACT_COMM     |             |

# Last-step provenance - keyword list

| keyword             | UType                             | FITS keyword                                       | Alternative |
|---------------------|-----------------------------------|--|-------------|
| activity_parameters | List of Parameter.name=value      | PARN_001 to<br>PARN_999<br>PARV_001 to<br>PARV_999 |             |
| used_ids            | List of <a href="#">Entity.id</a> | USD_001 to<br>USD_999                              |             |
| generated_ids       | List of <a href="#">Entity.id</a> | GEN_001 to<br>GEN_999                              |             |

- Entity.id for used and generated entities
  - Link to next part of the chain
  - Can be used in queries (ProvSAP, ProvTAP)
  - Potentiel reconstruction of a full provenance