



# Observatoire Virtuel et Provenance des données

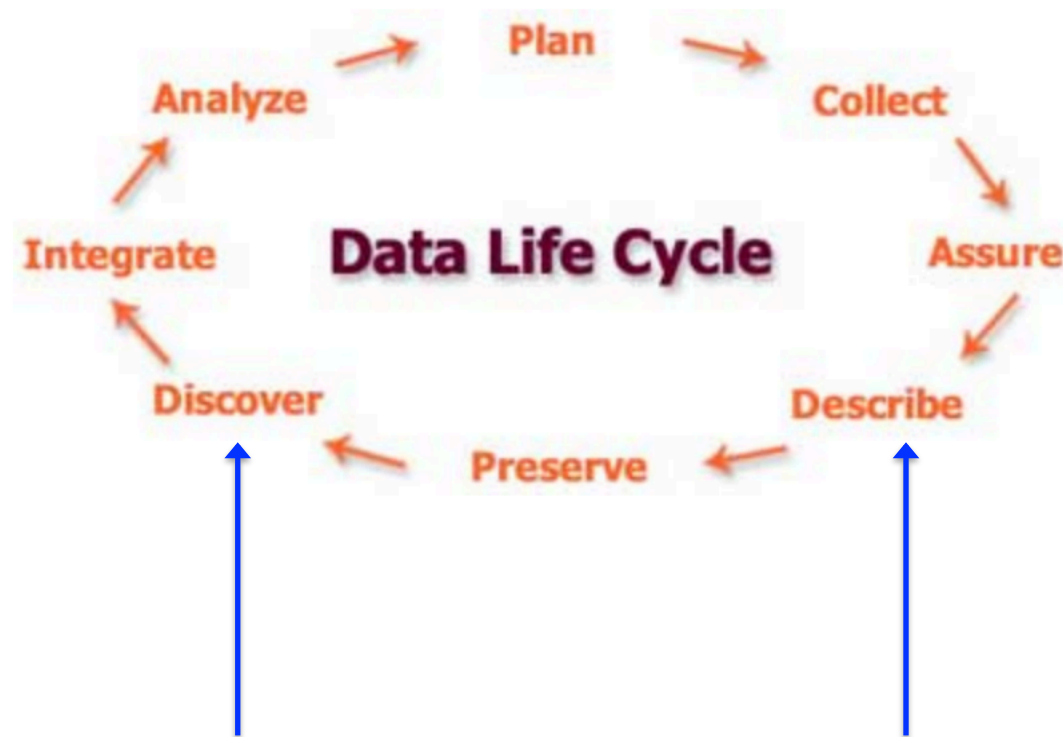
Catherine Boisson<sup>(2)</sup>, François Bonnarel<sup>(3)</sup>, Johan Bregeon<sup>(1)</sup>,  
Julien LeFaucheur<sup>(2)</sup>, Pierre Le Sidaner<sup>(4)</sup>, Mireille Louys<sup>(3)</sup>,  
Laurent Michel<sup>(5)</sup>, Markus Nullmeier<sup>(7)</sup>, Ana Palacios<sup>(1)</sup>, Kristin  
Riebe<sup>(6)</sup>, Michèle Sanguillon<sup>(1)</sup>, Mathieu Servillat<sup>(2)</sup>.

(1) LUPM, Montpellier, France – (2) LUTH, Meudon, France – (3) CDS, Strasbourg, France

(4) DIO, Paris, France – (5) OAS, Strasbourg, France – (6) – AIP, Pstam, Allemagne – (7) ARI, Heidelberg, Allemagne

# Data Management Plan

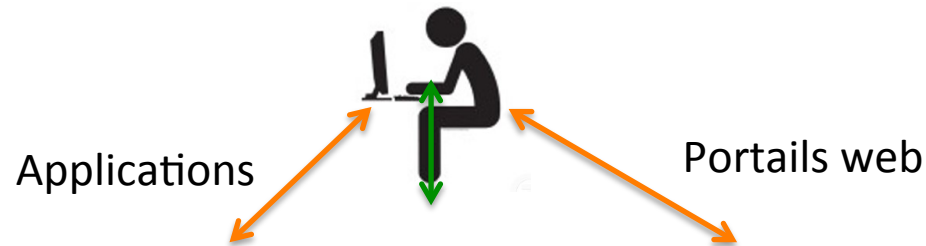
---



# Observatoire Virtuel en Astronomie

La notion d'Observatoire Virtuel repose sur la **mise à disposition** et **l'interconnexion**, à l'échelle internationale, de **banques de données numériques** et **d'outils** (analyse, visualisation, ...) développés dans les domaines de **Astronomie** et de **Astrophysique**.

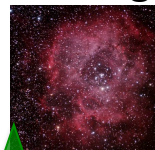
Les données constituent un « **ciel virtuel** » auquel l'astronome accède de façon transparente et continue avec son « télescope » (son ordinateur).



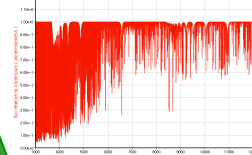
Annuaire



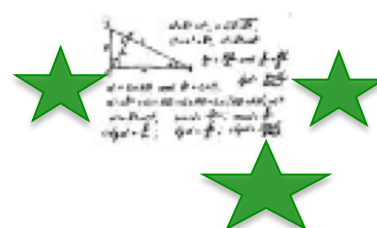
BD images



BD spectres



Catalogues



Stockage / Calcul



# International Virtual Observatory Alliance

L'IVOA est une alliance de projets d'observatoires virtuels du monde entier qui développent les normes requises à l'interopérabilité des données astronomiques.



@M. Louys  
RDA 6th plenary,  
septembre 2015

**If you want to go quickly go alone, if you want to go far go together**

# www.ivoa.net

- Objectif : permettre de faire de la **nouvelle science**
- Coordination des efforts pour l'**interopérabilité**
- Définition des **standards** de l'OV
- Fonctionnement :
  - groupes de travail, groupes d'intérêt (2 ateliers par an)
  - comité exécutif, comité sur les priorités scientifiques, groupe de coordination technique.

IVOA Document Standards Process

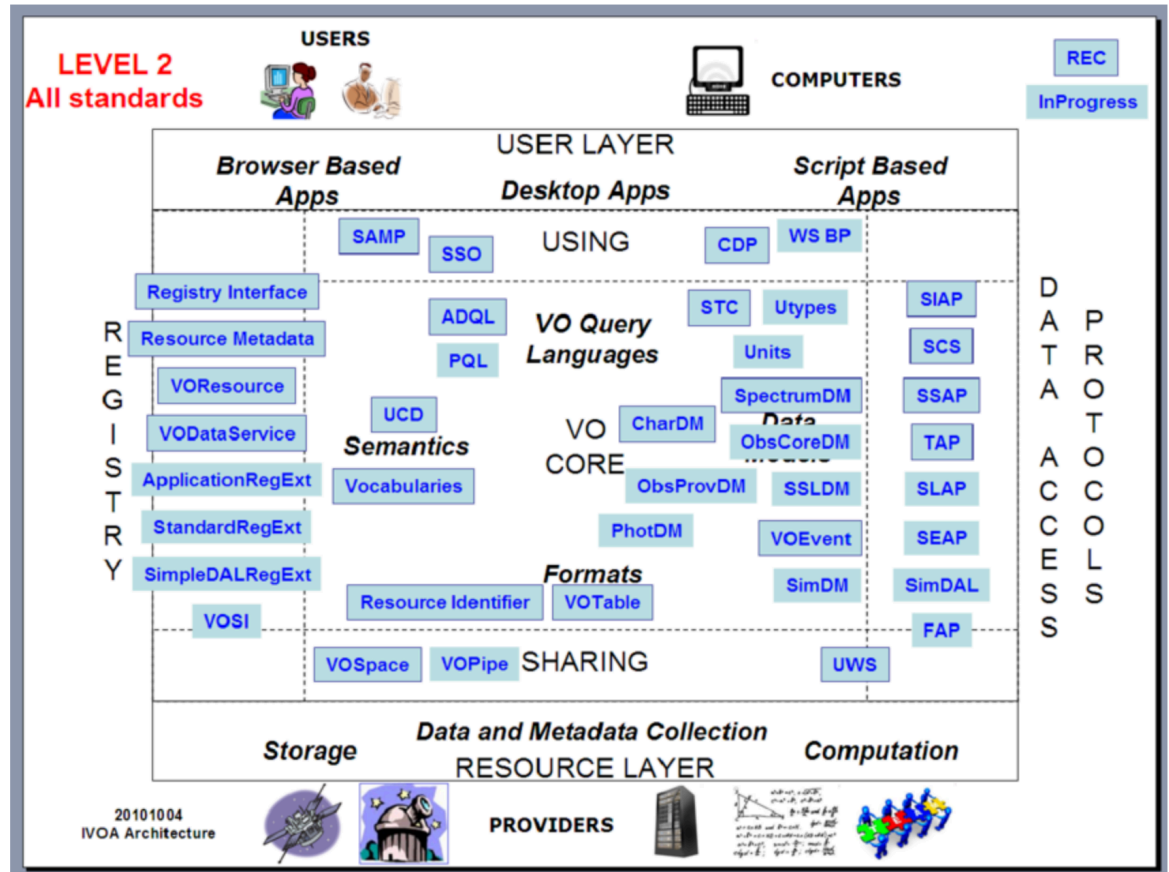


# www.ivoa.net

---

- **Groupes de travail** : ils travaillent pour un consensus et un agrément du travail technique  
(Data Model, Data Access Layer, Applications, Grid & Web services, Semantics, Registry)
- **Groupes d'intérêt** : ils agissent comme points de contact entre la communauté et les groupes de travail et vice versa  
(Data curation & Preservation, Theory, Education, Time Domain, Operations, Knowledge discovery in Databases)

# L'Observatoire Virtuel en Astronomie



# L'Observatoire Virtuel et l'IN2P3

---

- Utilisent l'OV :
  - CTA
  - LIGO/VIRGO
  - Pollux
- Intéressés par l'OV :
  - LSST
  - SVOM



# Qu'est-ce que la provenance ?

---

- Le principal objectif de tracer la provenance d'un objet ou d'une entité est de fournir une preuve contextuelle et circonstancielle de sa production d'origine ou de sa découverte, et de retracer, autant que possible, son histoire.
- Ce terme est utilisé dans les domaines tels que :
  - L'art
  - L'archéologie
  - La paléontologie
  - L'industrie alimentaire (viande, ...)
  - Les données numériques (web, données scientifiques)

# Provenance de nos données

---

- Pour un jeu de données, la provenance devrait aider à ...
  - Découvrir les étapes de la production
  - Savoir qui fournit ces données
  - Effectuer du reprocessing
  - Faire du debugging
  - Permettre d'évaluer la qualité des données
  - Rechercher dans les métadonnées structurées de provenance

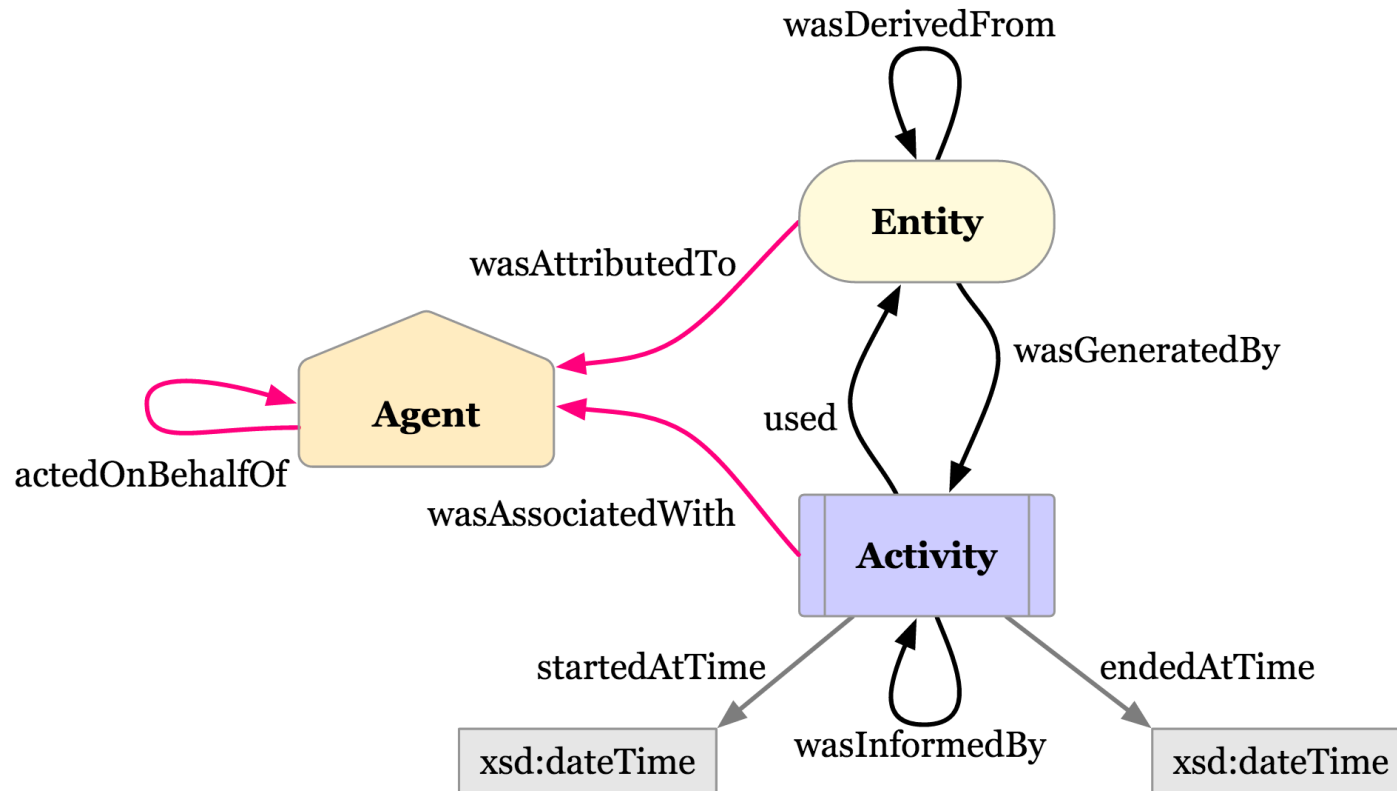
@K. Riebe -IVOA InterOp - May 2016, Cape Town

# Définition W3C

---

Provenance is information about entities, activities, and people involved in producing a piece of data or thing, which can be used to form assessments about its quality, reliability or trustworthiness.

# Le modèle Prov-DM du W3C

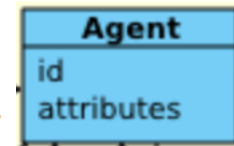
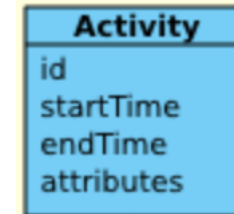
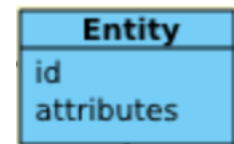


PROV-O: The PROV Ontology - W3C Recommendation 30 April 2013

# Prov-DM dans notre contexte

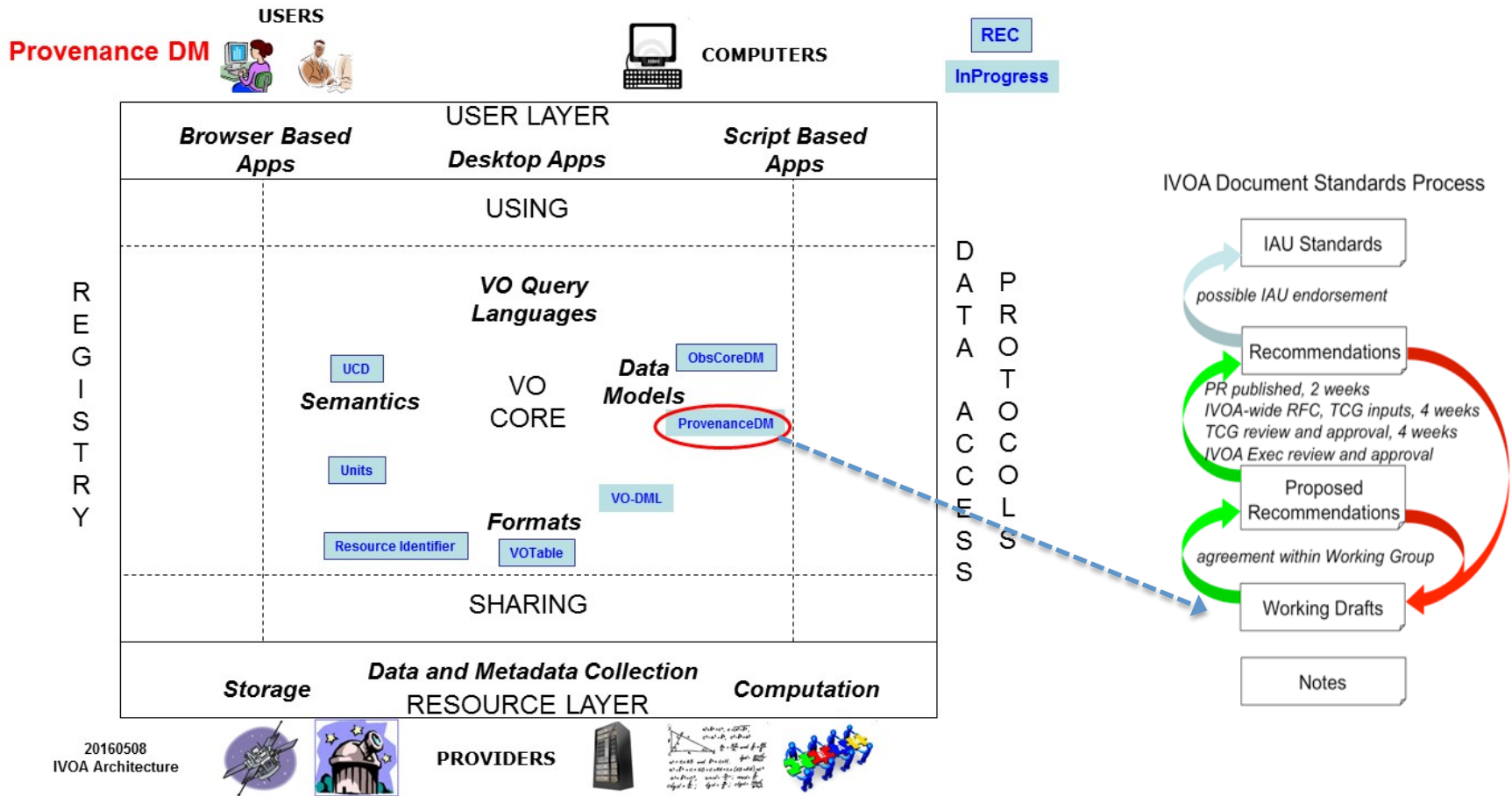
- Entity = Fichier ou collection de données ou paramètres (spectres, images, courbes de lumière, listes d'évènements, raw data, données calibrées, paramètres, ...)
- Activity = Instance de programme (acquisition, calibration, simulation, transformation, ...)
- Agent = Personne ou Organisation

Définition W3C

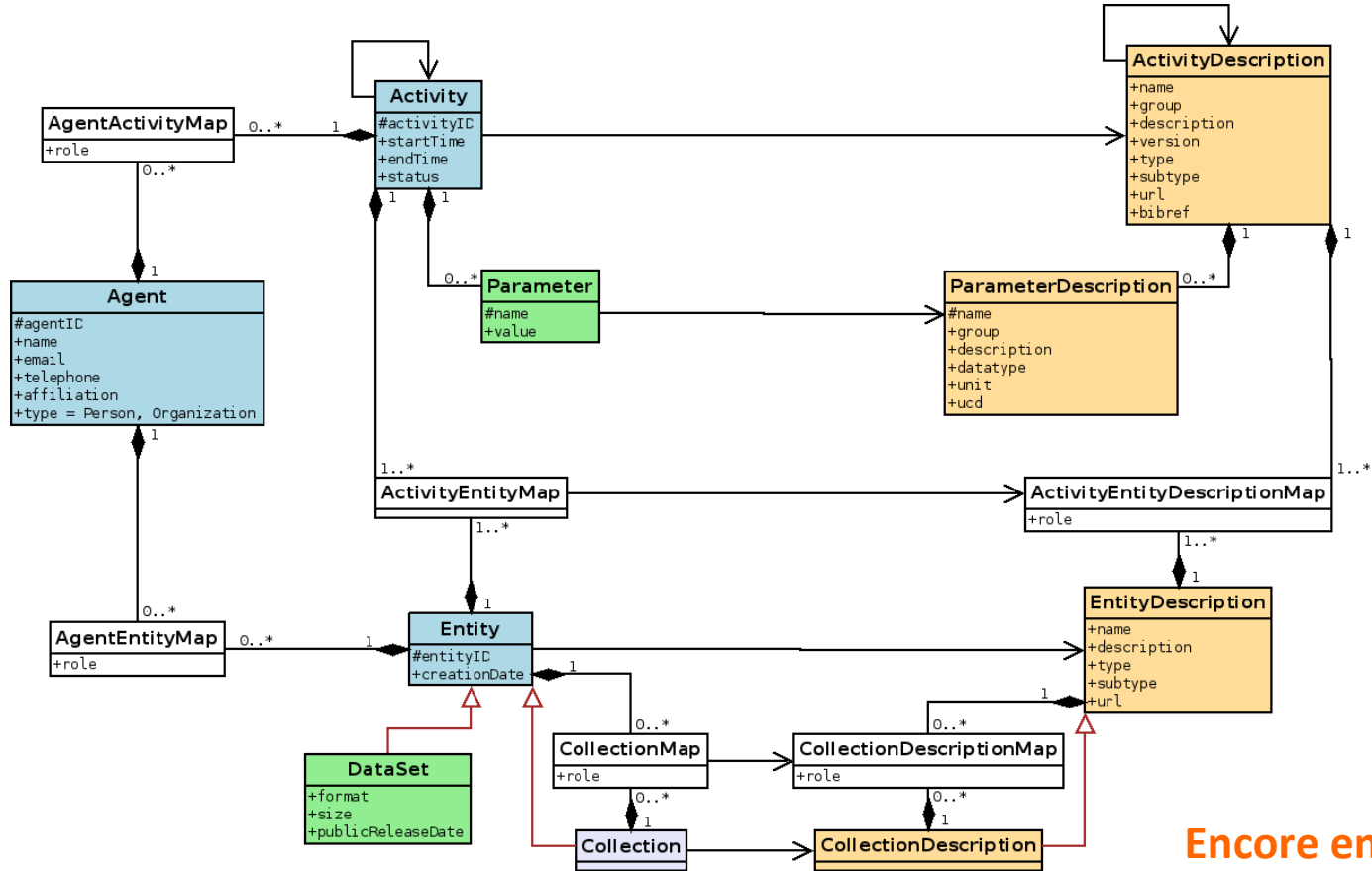


**Besoin d'enrichir le modèle**

# Le modèle ProvenanceDM de l'IVOA



# ProvenanceDM



Encore en discussion



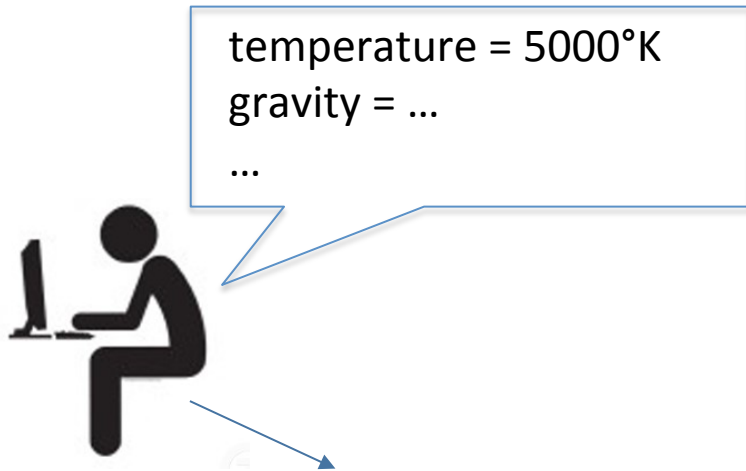
# Description de la provenance



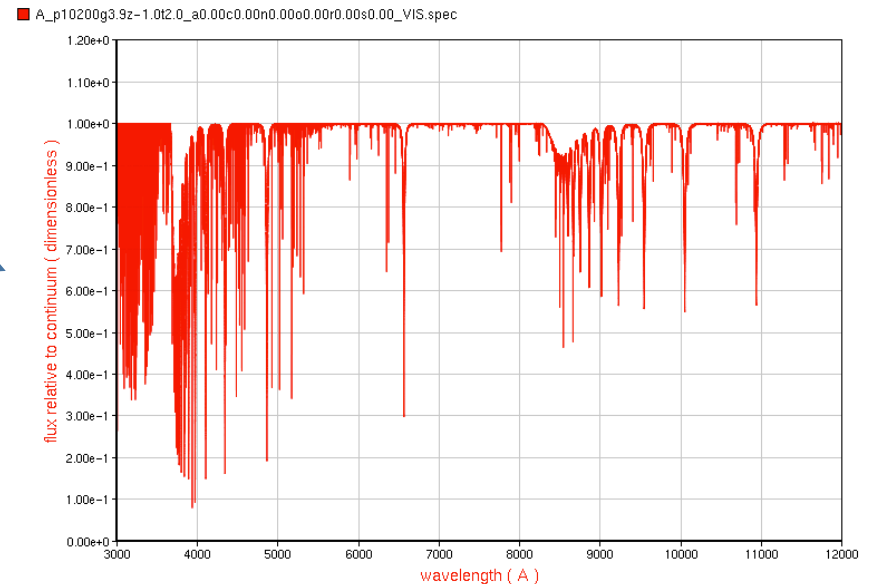
- Code : Utilisation du package python prov 1.4.0 (développé à l'Université de Southampton)
  - Bibliothèque qui implémente le Modèle de Données du W3C
  - Qui peut générer différents formats:
    - Formats texte : **PROVN, JSON, XML**
    - Formats graphiques : **PNG, SVG, PDF**
  - Possibilité d'utiliser différents espaces de nom :
    - Celui définit par le W3C :  
<http://www.w3.org/ns/prov#>
    - Celui que définira l'IVOA :  
<http://www.ivoa.net/documents/dm/provdm/voprov/>
    - Celui que définira une expérience
- Validation: utilisation de **Southampton Provenance Suite**  
<https://provenance.ecs.soton.ac.uk/>



# Exemple 1 : Pollux - Contexte



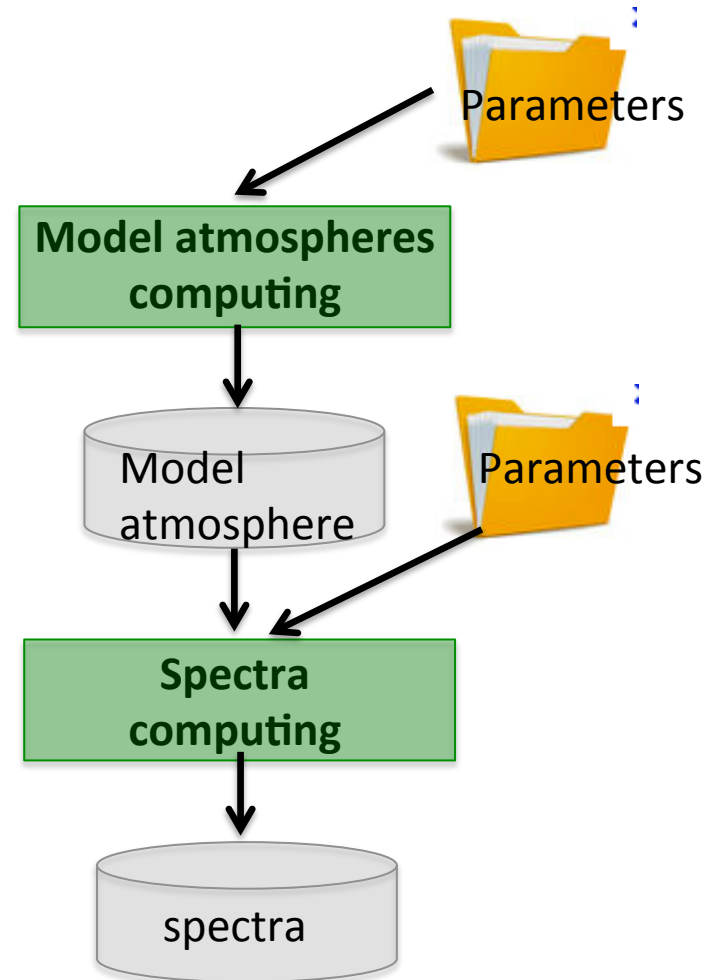
La base de données POLLUX collecte et met à disposition des spectres synthétiques haute résolution\*.



\* Le calcul de ces spectres sont faits par les fournisseurs de ceux-ci et non à la volée.

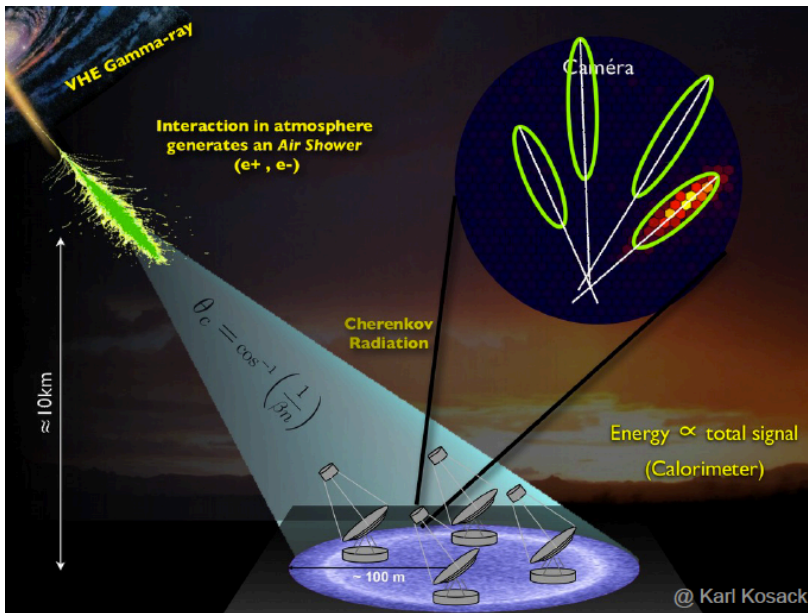
# Exemple 1 : Pollux - Données

Seuls les spectres (données + métadonnées) sont mis à la disposition des utilisateurs de l'Observatoire Virtuel

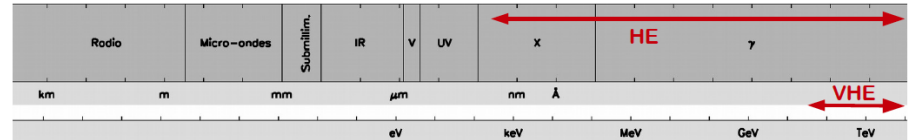




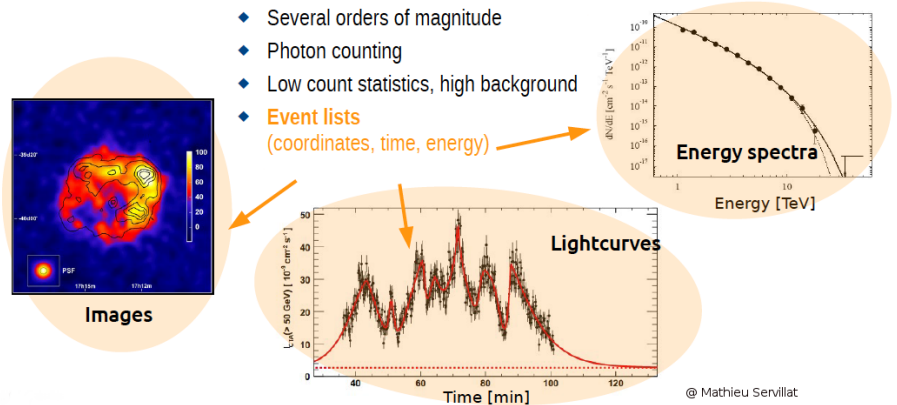
# Exemple 2 : CTA - Contexte



- Instrument pour « détecter » les rayons gamma de très haute énergie
- Données complexes :
  - Détection indirecte
  - Besoin de simulations pour comparer les données acquises aux données simulées



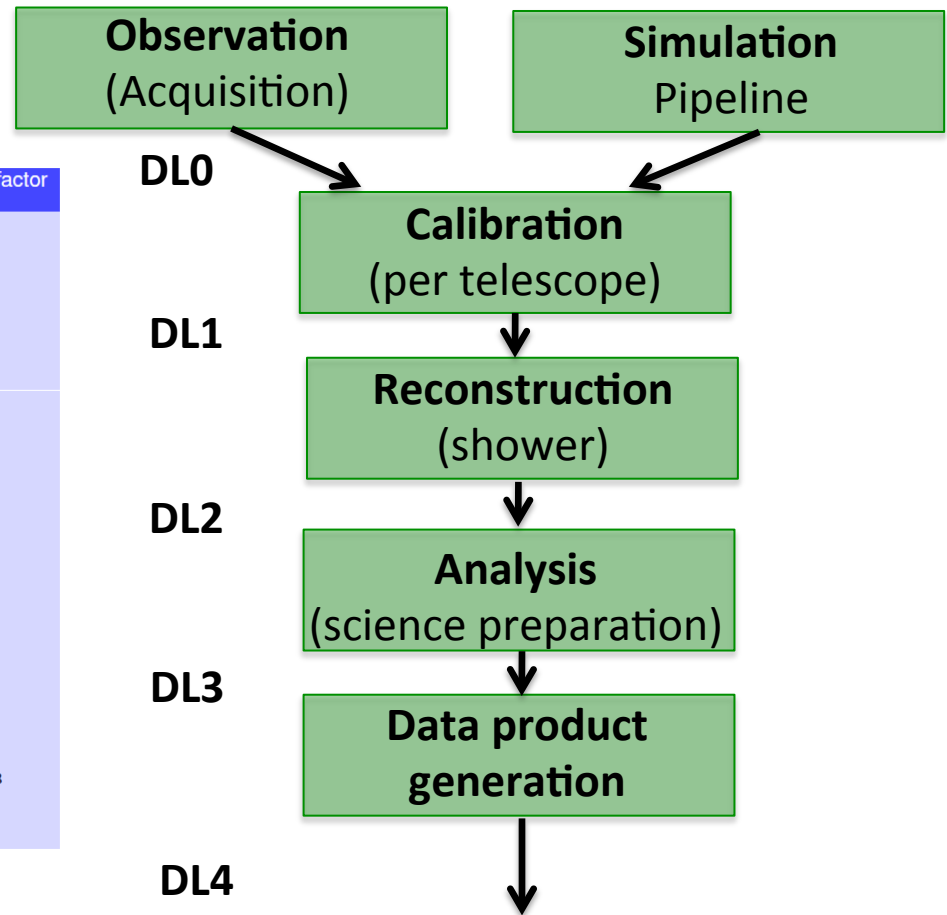
CTA sera le premier Observatoire Cherenkov mettant à disposition ses données de haut niveau (listes d'évènements, spectres, images, courbes de lumière) pour les utilisateurs de l'Observatoire Virtuel.



# Exemple 2 : CTA - Données

Data Level	Short Name	Description	Data reduction factor
Level 0 (DL0)	DAQ-RAW	Data from the Data Acquisition hardware/software.	
Level 1 (DL1)	CALIBRATED	Physical quantities measured in each separate camera: photons, arrival times, etc., and per-telescope parameters derived from those quantities.	1-0.2
Level 2 (DL2)	RECONSTRUCTED	Reconstructed shower parameters (per event, no longer per-telescope) such as energy, direction, particle ID, and related signal discrimination parameters.	$10^{-1}$
Level 3 (DL3)	REDUCED	Sets of selected (e.g. gamma-ray-candidate) events, along with associated instrumental response characterizations and any technical data needed for science analysis.	$10^{-2}$
Level 4 (DL4)	SCIENCE	High Level binned data products like spectra, sky maps, or light curves.	$10^{-3}$
Level 5 (DL5)	OBSERVATORY	Legacy observatory data, such as CTA survey sky maps or the CTA source catalog.	$10^{-5} - 10^{-3}$

@M. Servillat -IVOA InterOp - May 2016, Cape Town



# Exemple 2 : CTA - Provenance

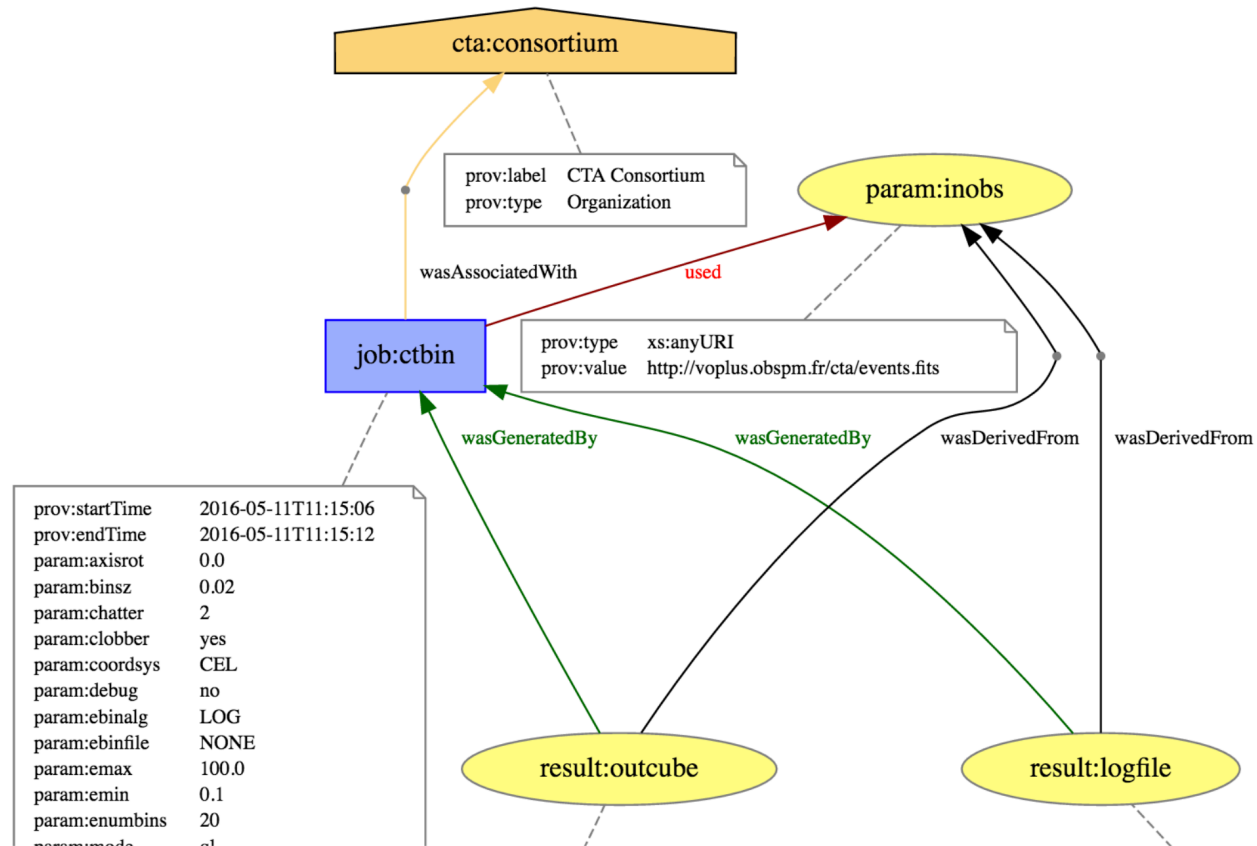
## PROV output (XML and JSON)

```
<prov:document xmlns:ctadata="ivo://vopdc.obspm/cta#" xmlns:ctajob
  <prov:activity prov:id="ctajobs:ctbin">
    <prov:startTime> 2016-03-13T23:44:46 </prov:startTime>
    <prov:endTime> 2016-03-13T23:44:56 </prov:endTime>
  </prov:activity>
  <prov:agent prov:id="cta:consortium">
    <prov:type xsi:type="xsd:string"> Organization </prov:type>
  </prov:agent>
  <prov:wasAssociatedWith>
    <prov:activity prov:ref="ctajobs:ctbin" />
    <prov:agent prov:ref="cta:consortium" />
  </prov:wasAssociatedWith>
  <prov:entity prov:id="uwsdata:parameters/inobs" />
  <prov:used>
    <prov:activity prov:ref="ctajobs:ctbin" />
    <prov:entity prov:ref="uwsdata:parameters/inobs" />
  </prov:used>
  <prov:entity prov:id="uwsdata:results/outcube" />
  <prov:wasGeneratedBy>
    <prov:entity prov:ref="uwsdata:results/outcube" />
    <prov:activity prov:ref="ctajobs:ctbin" />
  </prov:wasGeneratedBy>
  <prov:wasDerivedFrom>
    <prov:generatedEntity prov:ref="uwsdata:results/outcube" />
    <prov:usedEntity prov:ref="uwsdata:parameters/inobs" />
  </prov:wasDerivedFrom>
  <prov:entity prov:id="uwsdata:results/logfile" />
  <prov:wasGeneratedBy>
    <prov:entity prov:ref="uwsdata:results/logfile" />
    <prov:activity prov:ref="ctajobs:ctbin" />
  </prov:wasGeneratedBy>
  <prov:wasDerivedFrom>
    <prov:generatedEntity prov:ref="uwsdata:results/logfile" />
    <prov:usedEntity prov:ref="uwsdata:parameters/inobs" />
  </prov:wasDerivedFrom>
</prov:document>
```

```
{
  - wasAssociatedWith: {
    - _:id1: {
      prov:agent: "cta:consortium",
      prov:activity: "cta:anactools_v1.1"
    }
  },
  - agent: {
    - cta:consortium: {
      prov:type: "Organization"
    }
  },
  - entity: {
    uwsdata:results/fit_results: { },
    uwsdata:results/configfile: { },
    uwsdata:results/butterfly: { },
    uwsdata:results/spectrum_plot: { },
    uwsdata:results/spectrum: { }
  },
  - prefix: {
    uwsdata: "https://voparis-uws-test.obspm.fr/rest",
    cta: "http://www.cta-observatory.org#",
    voprov: "http://www.ivoa.net/ns/voprov#"
  },
  - activity: {
    - cta:anactools_v1.1: {
      prov:startTime: "2016-04-07T00:26:00",
      prov:endTime: "2016-04-07T00:27:15"
    }
  },
  - wasGeneratedBy: {
    - _:id5: {
      prov:entity: "uwsdata:results/butterfly",
      prov:activity: "cta:anactools_v1.1"
    },
    - _:id4: {
      prov:entity: "uwsdata:results/fit_results",
      prov:activity: "cta:anactools_v1.1"
    }
  },
  ...
}
```

@M. Servillat - IVOA InterOp - May 2016, Cape Town

# Exemple 2 : CTA - Provenance




@M. Servillat -IVOA InterOp - May 2016, Cape Town

# RDA

<https://rd-alliance.org/groups/research-data-provenance.html>

RESEARCH DATA SHARING WITHOUT BARRIERS

RDA EU RDA US CONTACT US LOGIN REGISTRATION

 RESEARCH DATA ALLIANCE

ABOUT RDA GET INVOLVED GROUPS RECOMMENDATIONS & OUTPUTS RDA FOR DISCIPLINES PLENARIES EVENTS NEWS & MEDIA

Home » Working And Interest Groups » Interest Group » Research Data Provenance

## Research Data Provenance

**IG** **Group details**

**Status:** Recognised & Endorsed  
**Chair(s):** David Dubin, Nicholas Car  
**Case Statement:** [Download](#)

**Research Data Provenance**

**Status:** Recognised & Endorsed  
**Chair(s):** David Dubin, Nicholas Car

**Group Email:**  
**Secretariat Liaison:**



**Je vous remercie de votre attention.**