IVOA InterOp Meeting

Garching, 09/11/09

The Case for Provenance

Juande Santander-Vela, Arancha Delgado
European Southern Observatory, Archive Department

Talk Outline

- What is Provenance?
- How does Provenance fit in the VO?
- The case(s) for Provenance
- Conclusions

What is Provenance?

What is Provenance?

provenance |ˈprävənəns|

noun
the place of origin or earliest known history of something *: an orange rug of Iranian provenance.*
• the beginning of something's existence; something's origin *: they try to understand the whole universe, its provenance and fate.*
See note at **origin** .
• a record of ownership of a work of art or an antique, used as a guide to authenticity or quality *: the manuscript has a distinguished provenance.*
ORIGIN late 18th cent.: from French, from the verb ***provenir 'come or stem from,'*** from Latin ***provenire***, from ***pro- 'forth'*** + ***venire 'come.'***

What is Provenance?

provenance |ˈprävənəns|

noun
the place of origin or earliest known history of something *: an orange rug of Iranian provenance.*
• the beginning of something's existence; something's origin *: they try to understand the whole universe, its provenance and fate.*
See note at **origin** .
• a record of ownership of a work of art or an antique, used as a guide to authenticity or quality *: the manuscript has a distinguished provenance.*
ORIGIN late 18th cent.: from French, from the verb ***provenir 'come or stem from,'*** from Latin ***provenire***, from ***pro- 'forth'*** + ***venire 'come.'***

What is Provenance?

provenance |ˈprävənəns|

noun
the place of origin or earliest known history of something *: an orange rug of Iranian provenance.*
• the beginning of something's existence; something's origin *: they try to understand the whole universe, its provenance and fate.*
See note at **origin** .
• a record of ownership of a work of art or an antique, used as a guide to authenticity or quality *: the manuscript has a distinguished provenance.*
ORIGIN late 18th cent.: from French, from the verb ***provenir 'come or stem from,'*** from Latin ***provenire***, from ***pro- 'forth'*** + ***venire 'come.'***

What is Provenance?

provenance |ˈprävənəns|

noun
the place of origin or earliest known history of something *: an orange rug of Iranian provenance.*
• the beginning of something's existence; something's origin *: they try to understand the whole universe, its provenance and fate.*
See note at **origin** .
• a record of ownership of a work of art or an antique, used as a guide to authenticity or quality *: the manuscript has a distinguished provenance.*
ORIGIN late 18th cent.: from French, from the verb ***provenir 'come or stem from,'*** from Latin ***provenire***, from ***pro- 'forth'*** + ***venire 'come.'***

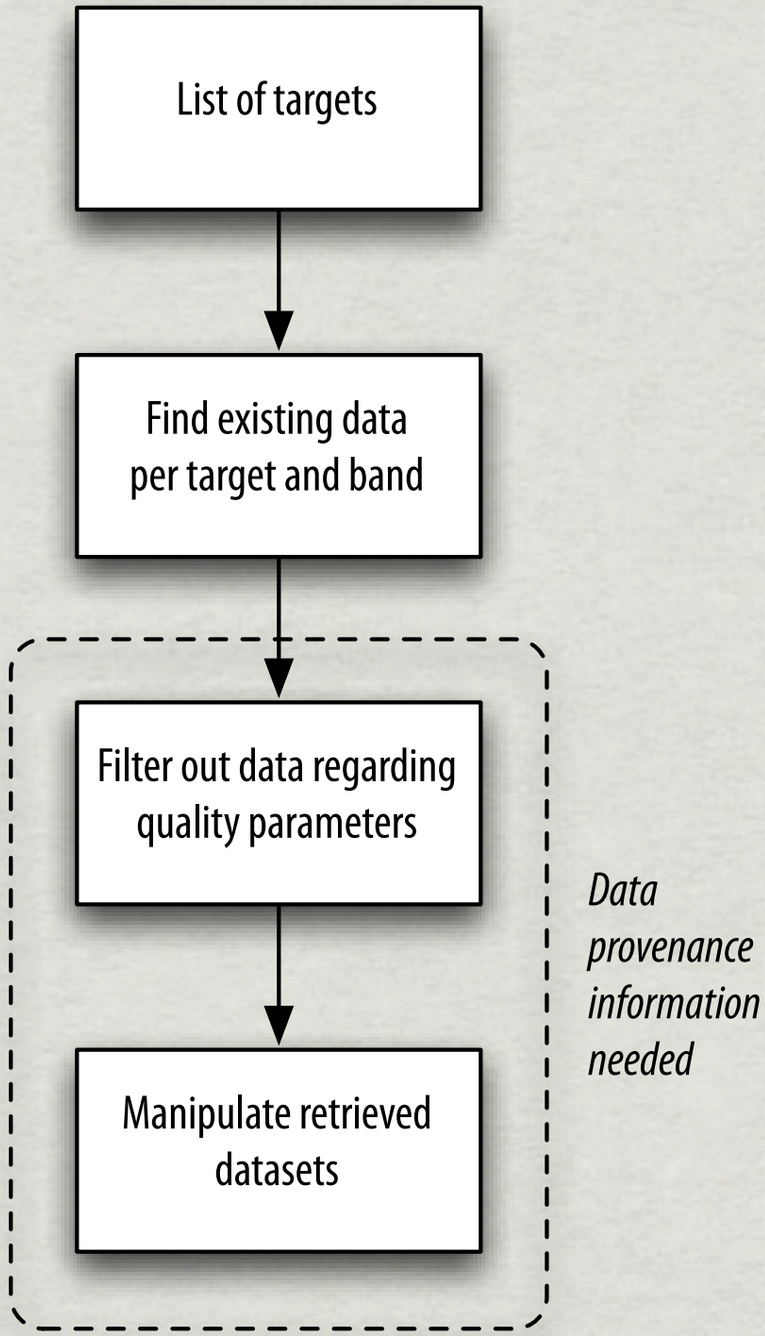
What is Provenance?

- Definition points out to:
  - **History** of astronomical data products
  - **Ownership/Attribution** (Observer, Proposal, Telescope/Instrument)…
  - **Quality**

What is Provenance?

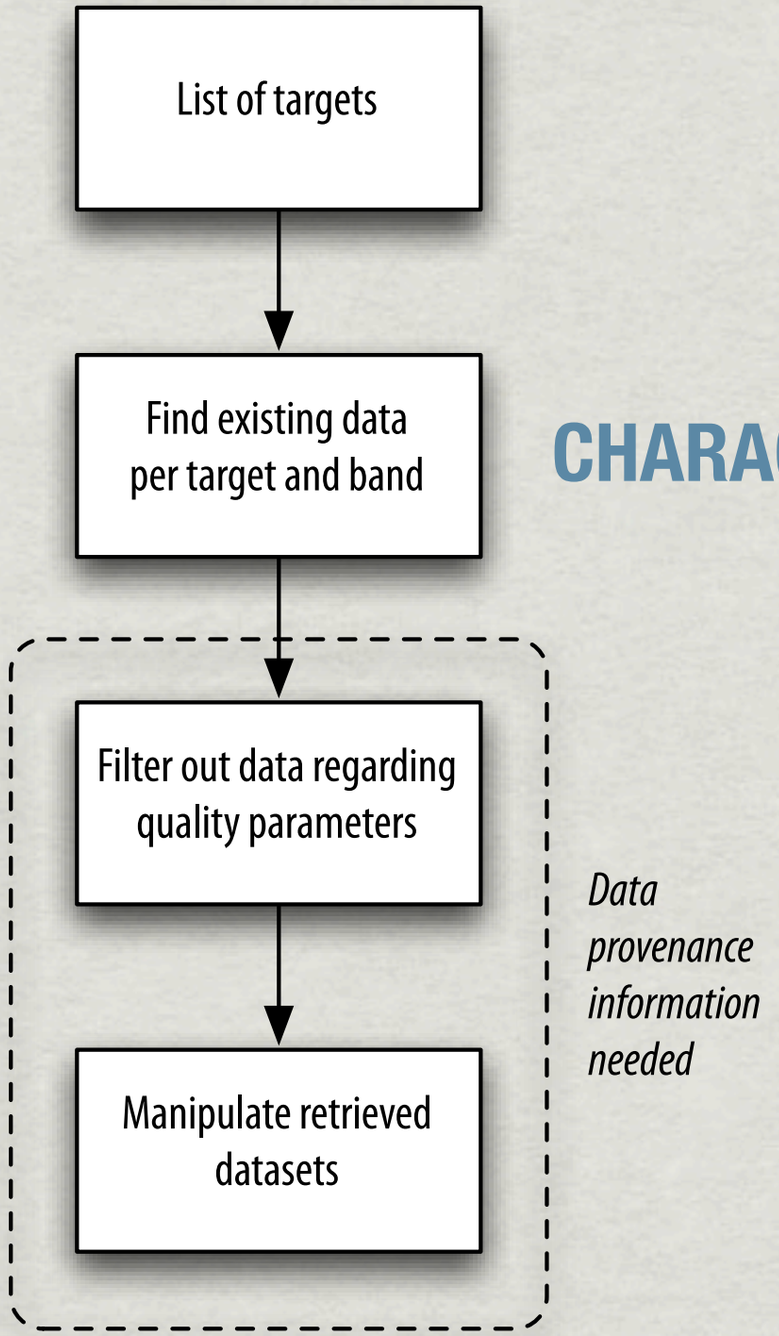
What is Provenance?

TYPICAL ASTRONOMICAL WORKFLOW



What is Provenance?

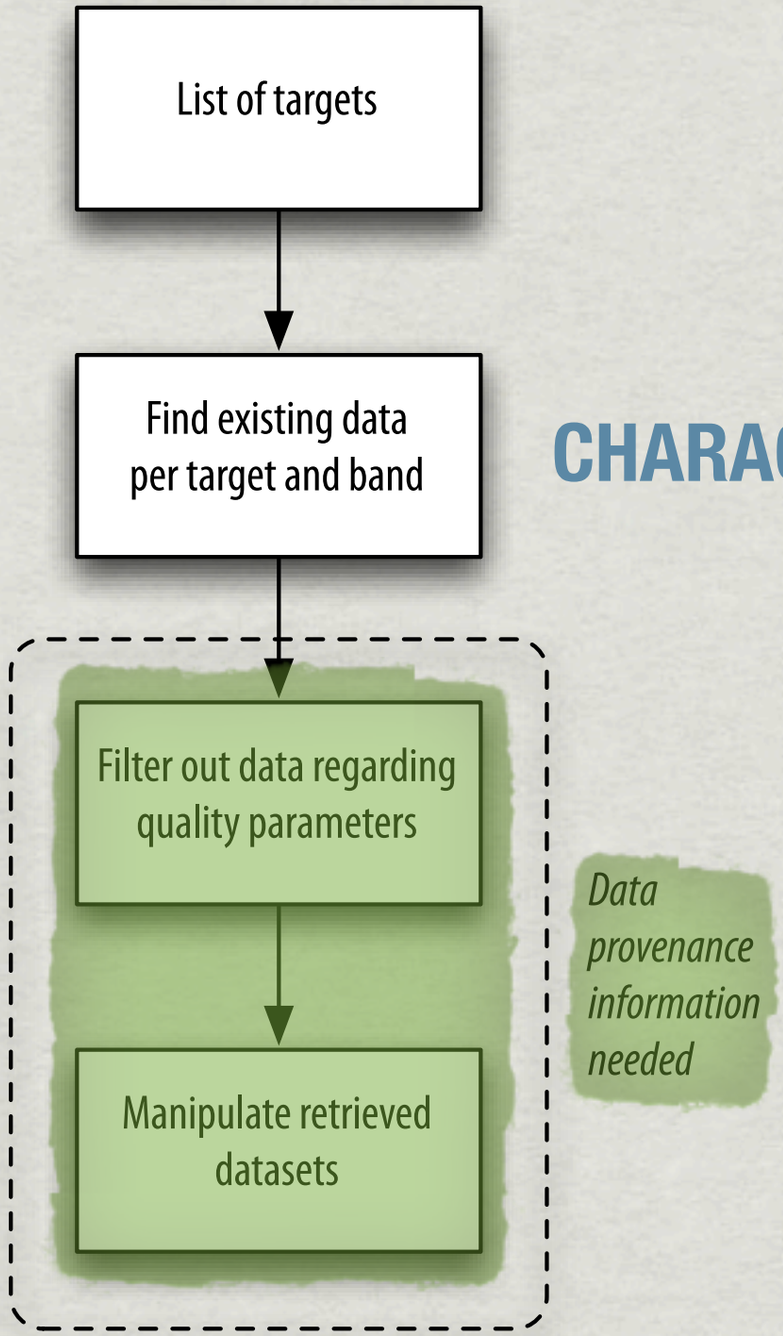
TYPICAL ASTRONOMICAL WORKFLOW



CHARACTERISATION

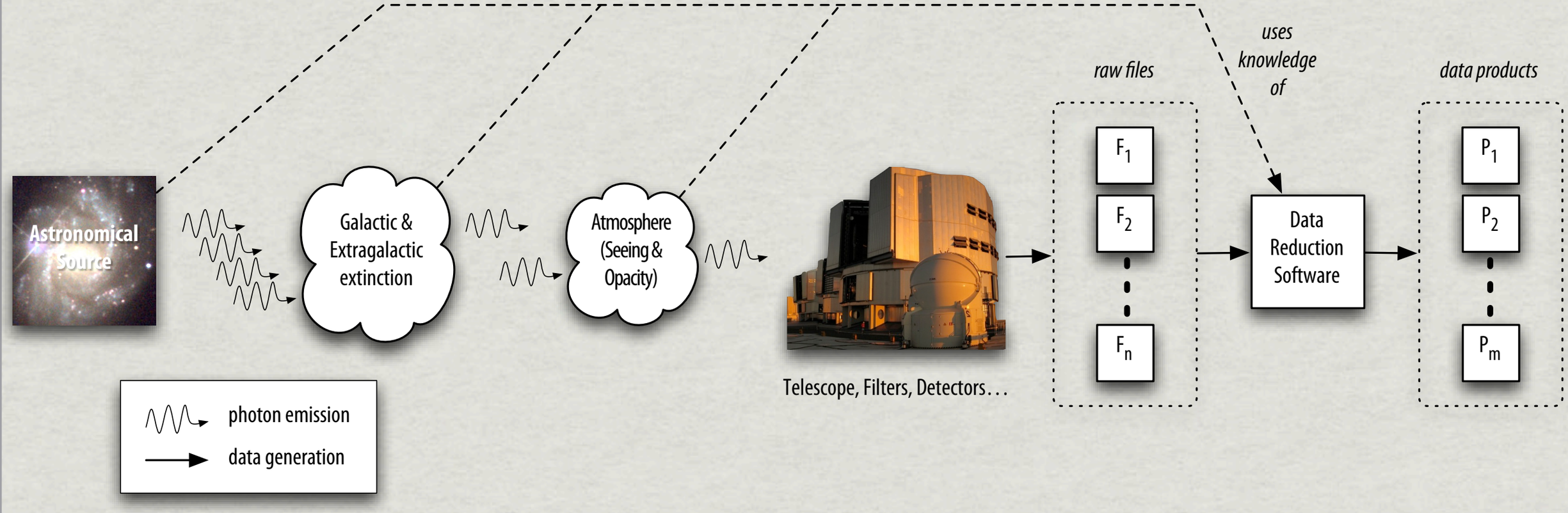
What is Provenance?

TYPICAL ASTRONOMICAL WORKFLOW

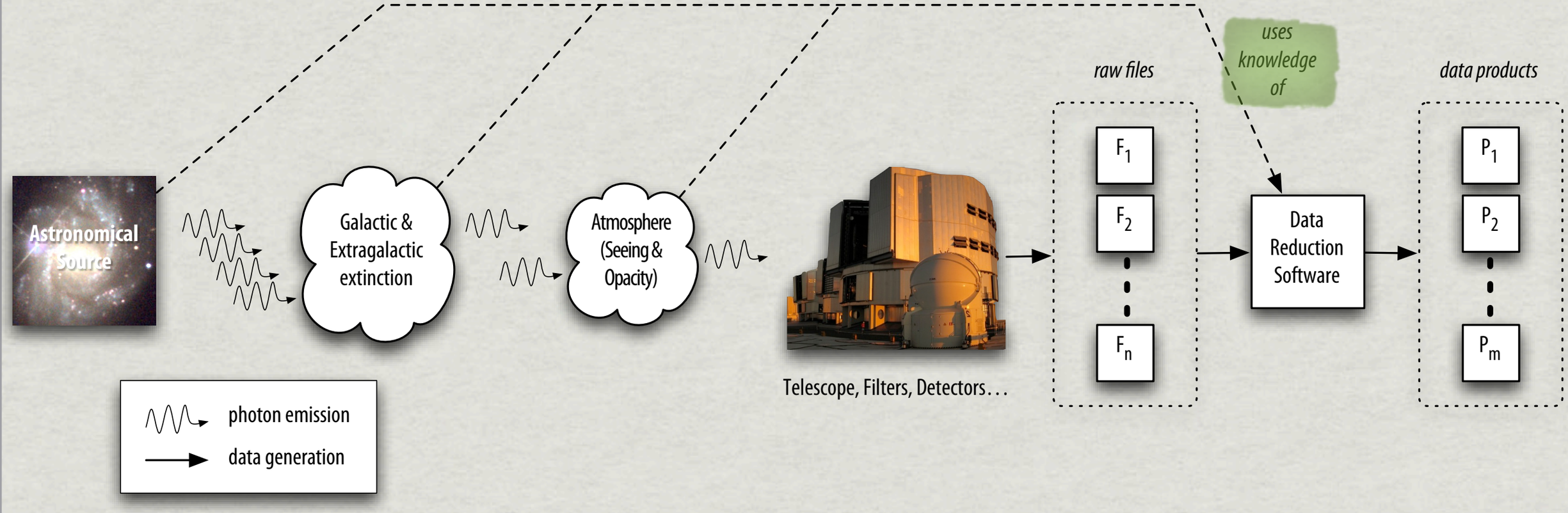


CHARACTERISATION

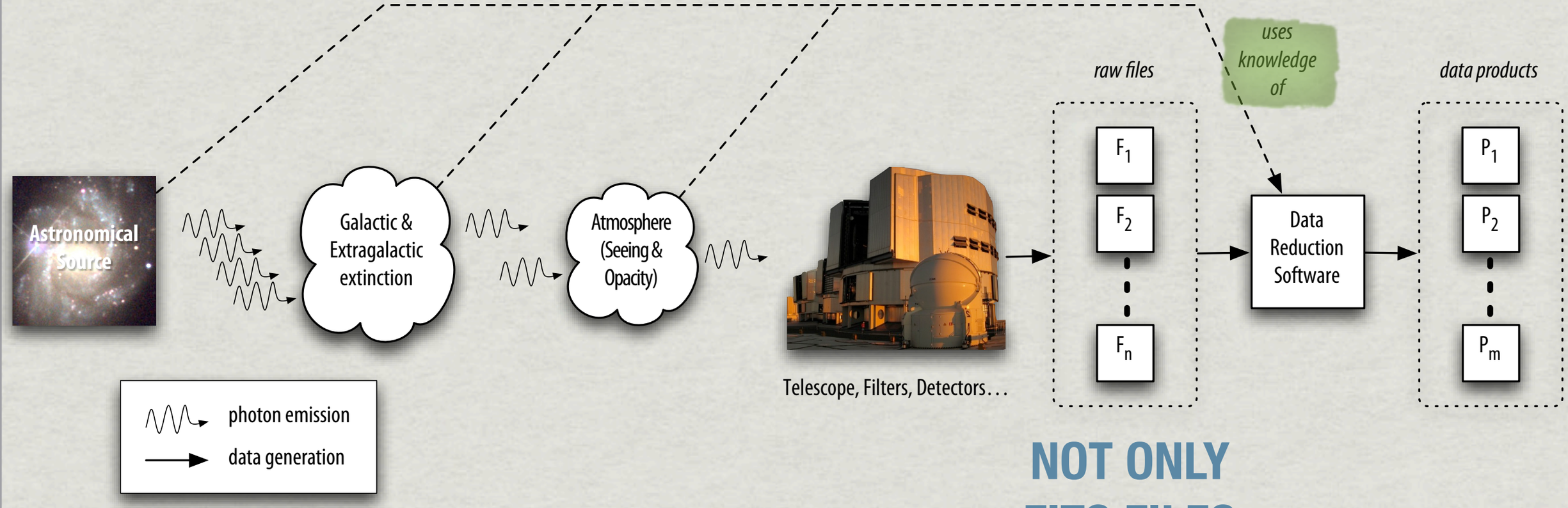
What is Provenance?



What is Provenance?

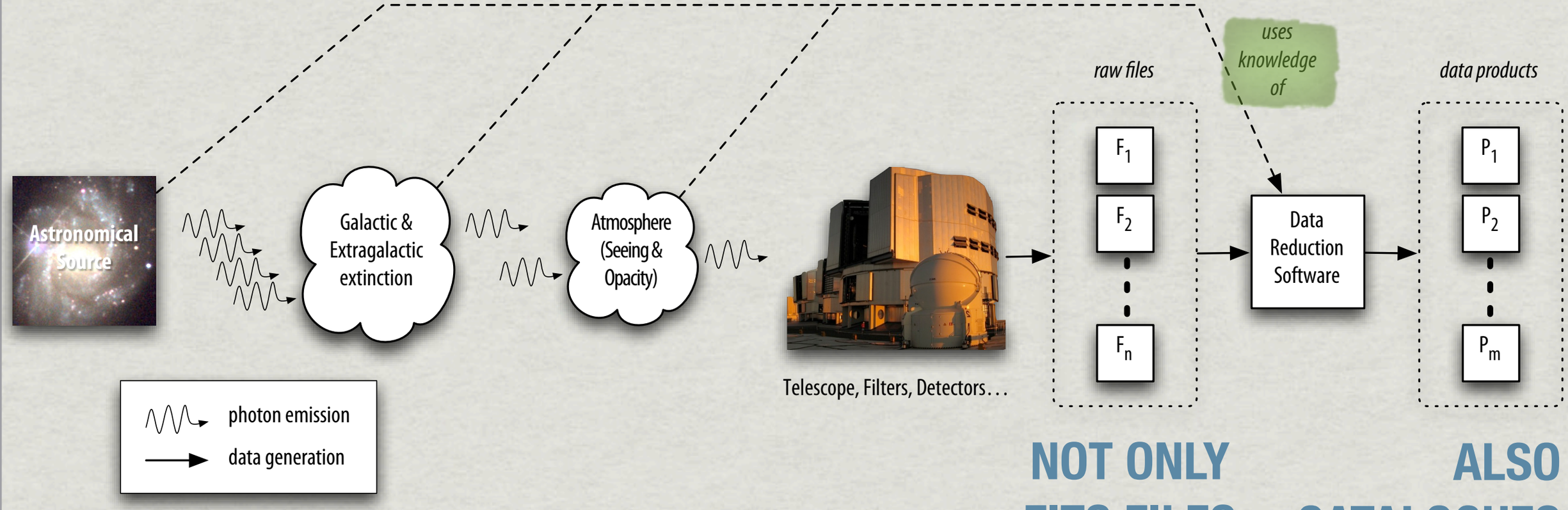


What is Provenance?



NOT ONLY FITS FILES

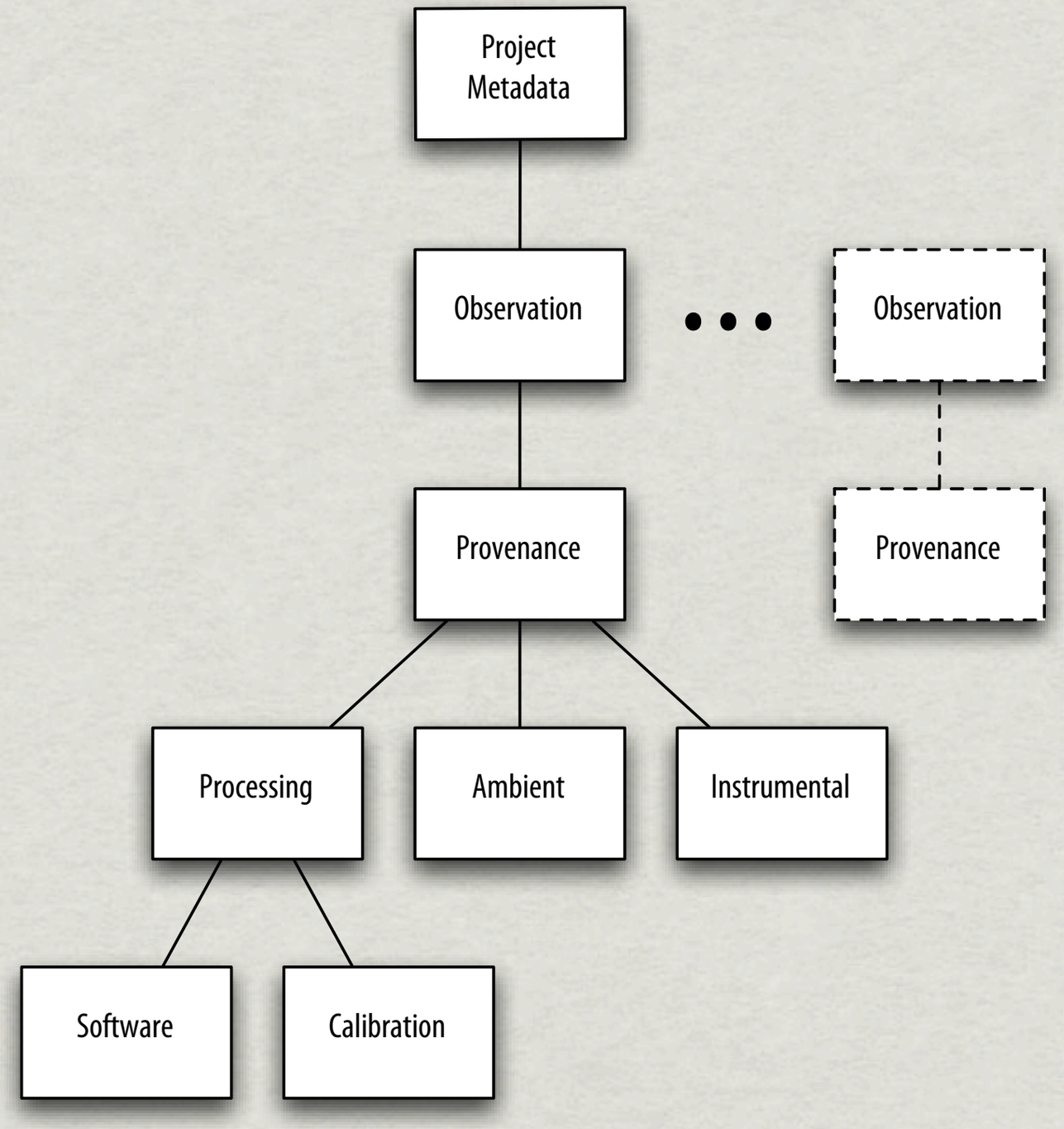
What is Provenance?



NOT ONLY FITS FILES

ALSO CATALOGUES

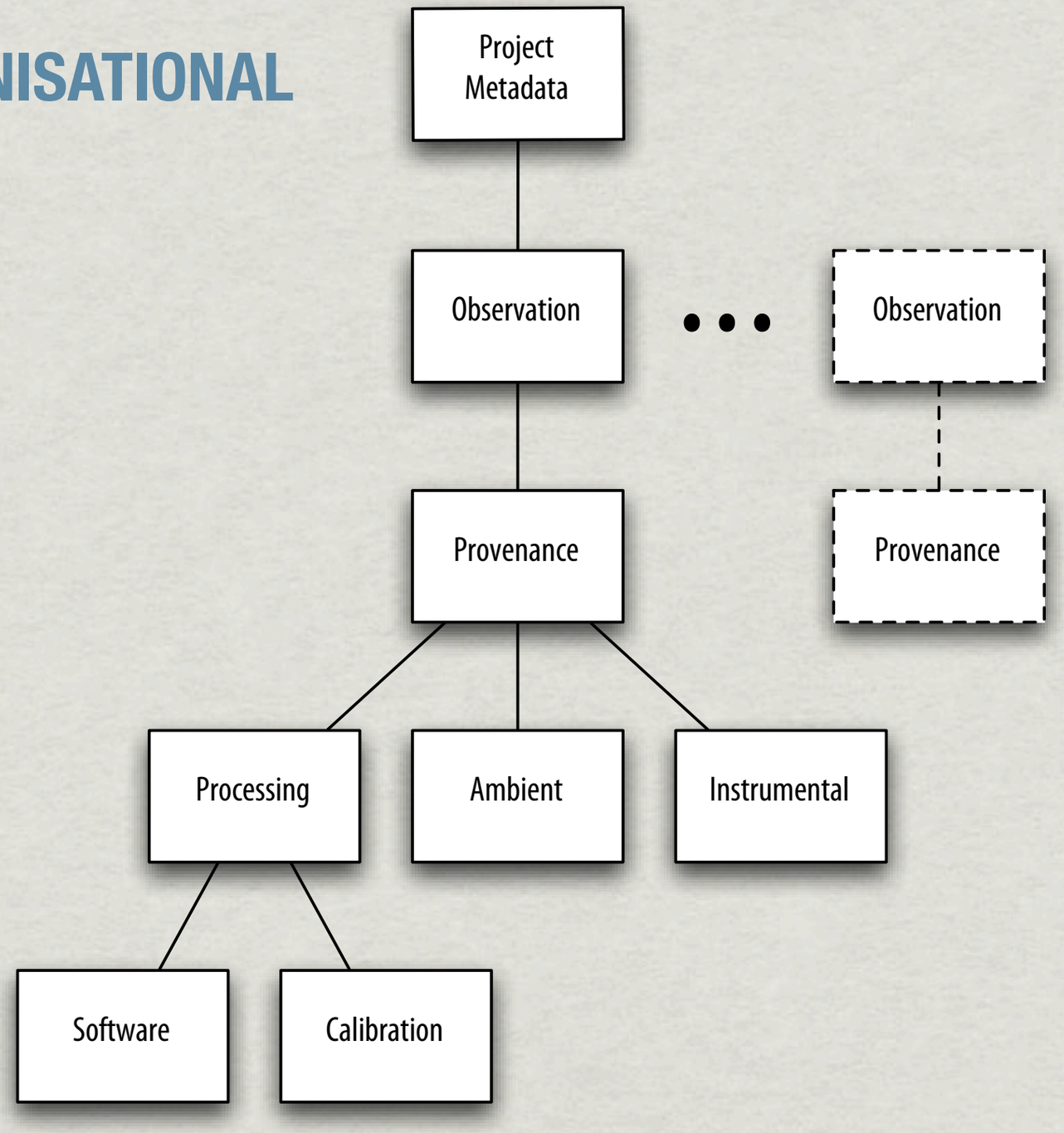
What is Provenance?



INTERNAL PROVENANCE WHITEPAPER BY N. DELMOTTE

What is Provenance?

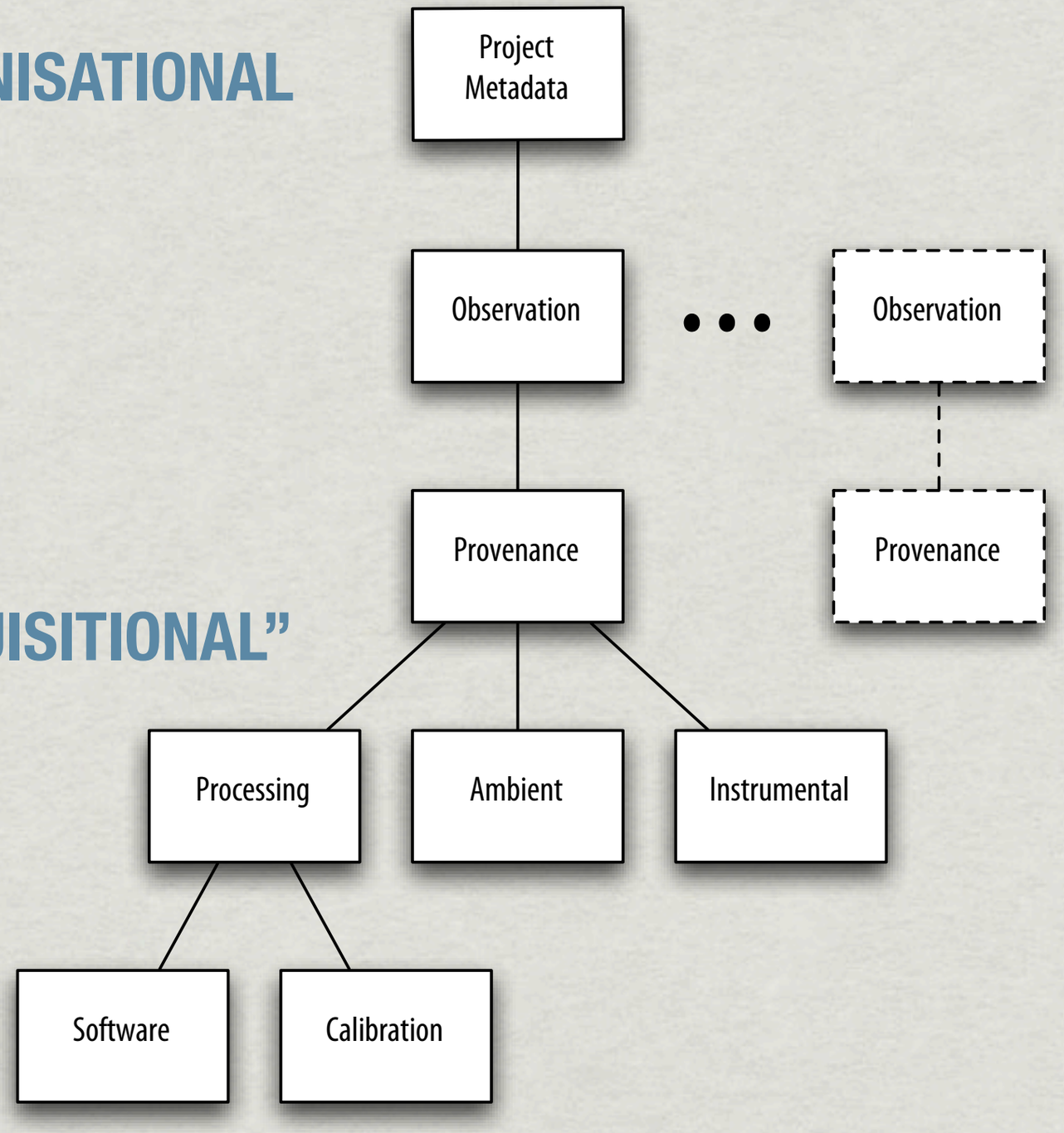
ORGANISATIONAL



INTERNAL PROVENANCE WHITEPAPER BY N. DELMOTTE

What is Provenance?

ORGANISATIONAL

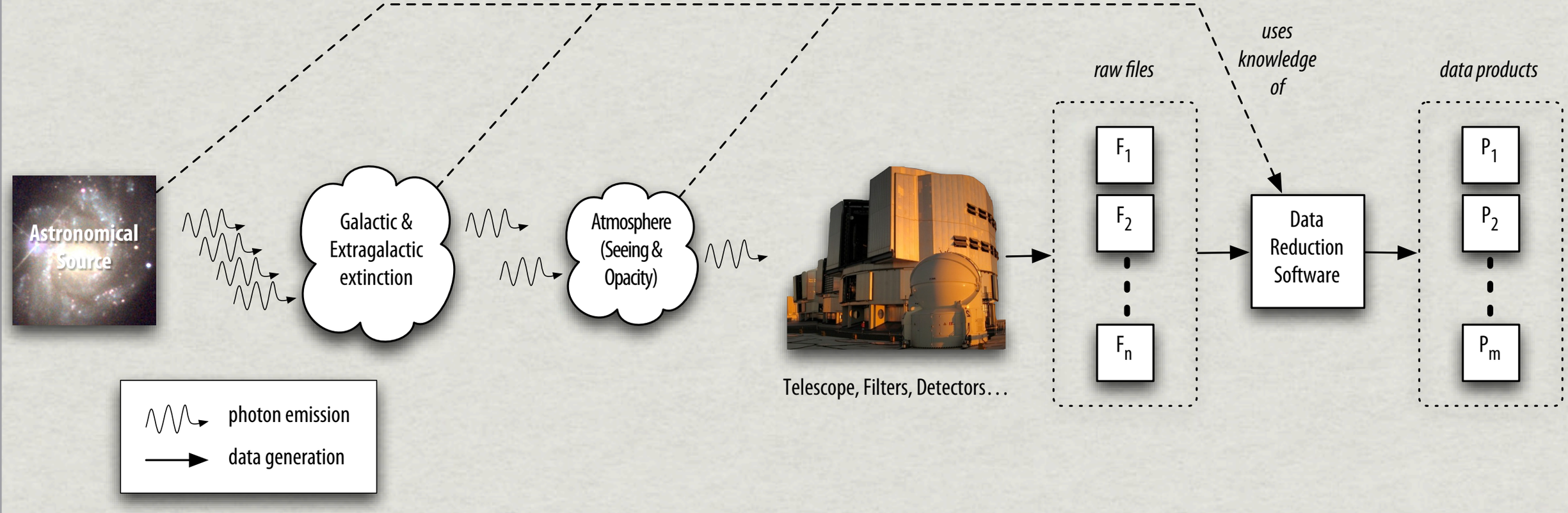


“ACQUISITIONAL”

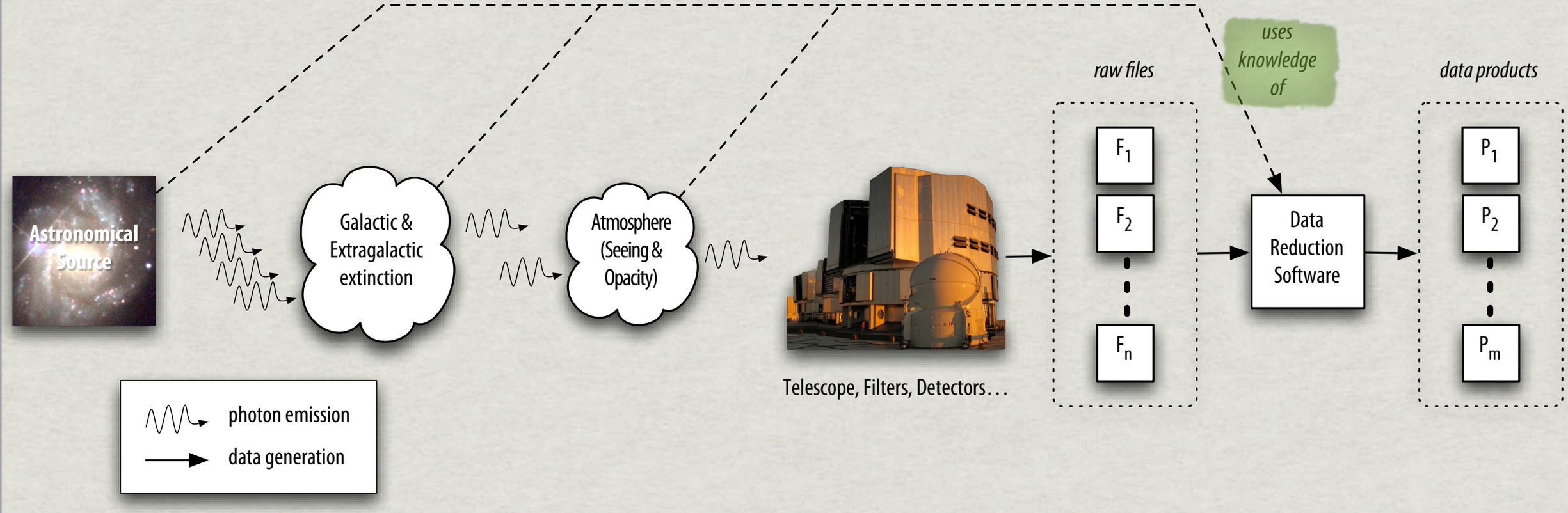
INTERNAL PROVENANCE WHITEPAPER BY N. DELMOTTE

How does Provenance fit in the VO?

How does Provenance fit in the VO?

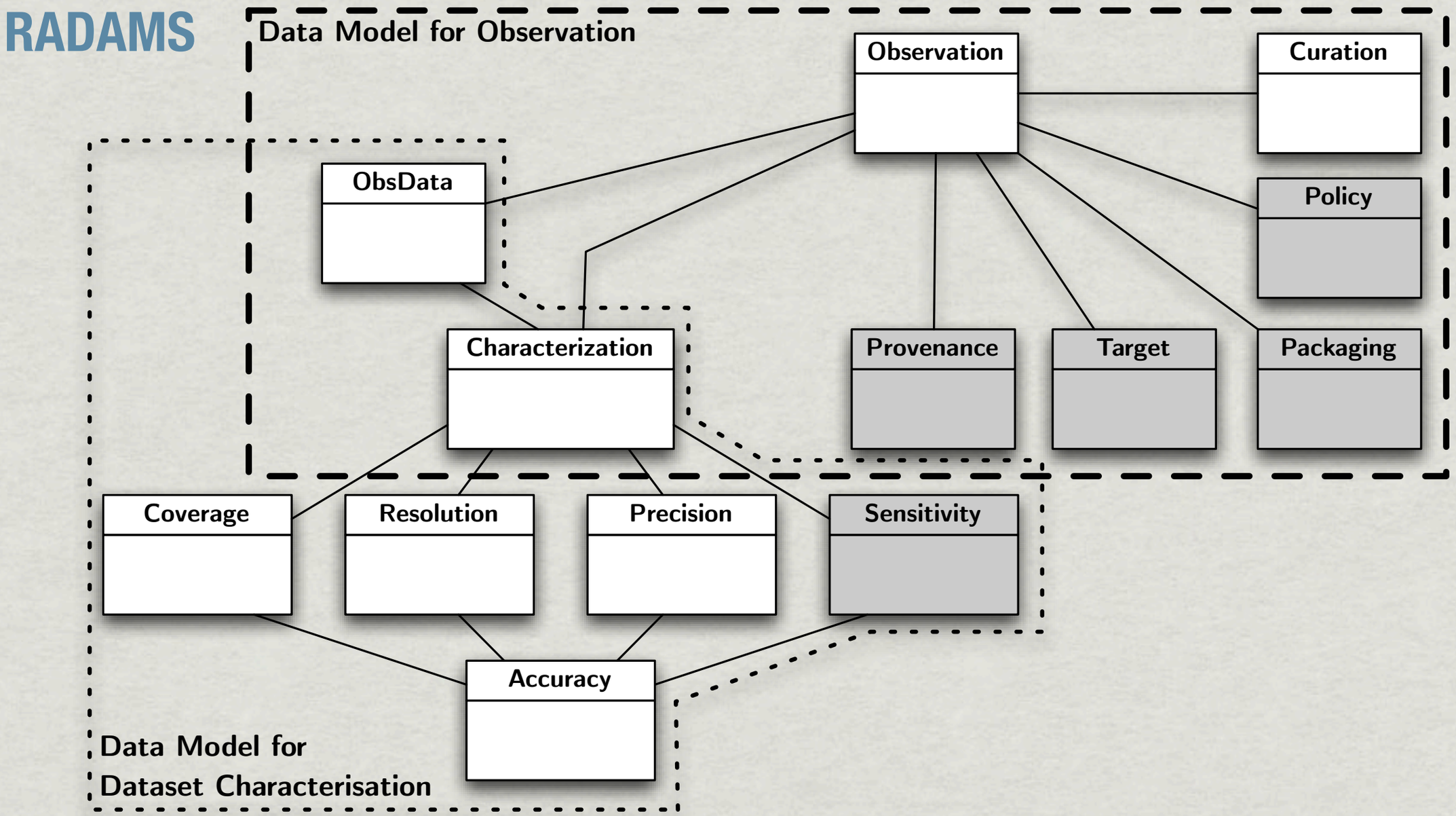


How does Provenance fit in the VO?



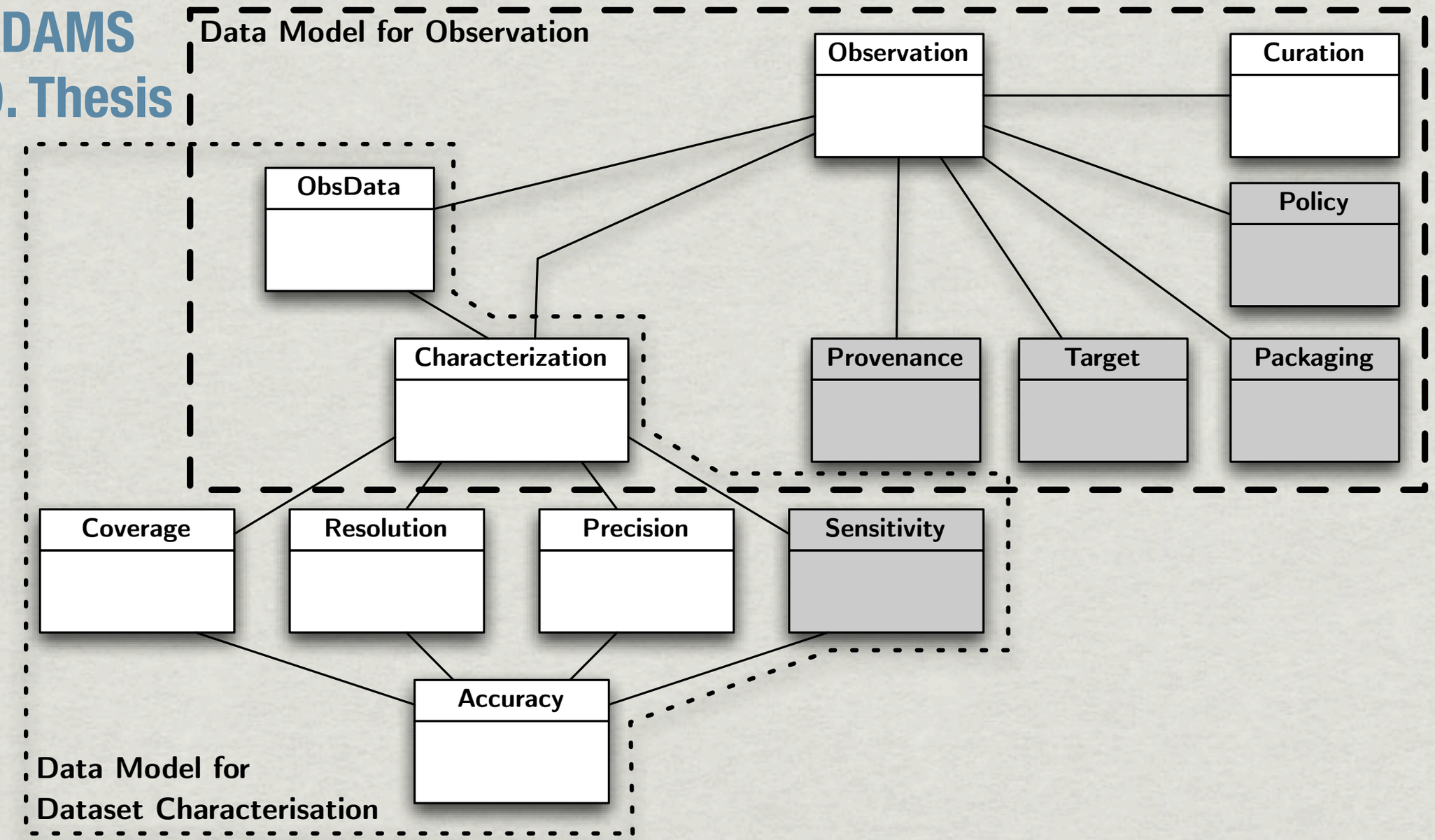
How does Provenance fit in the VO?

How does Provenance fit in the VO?

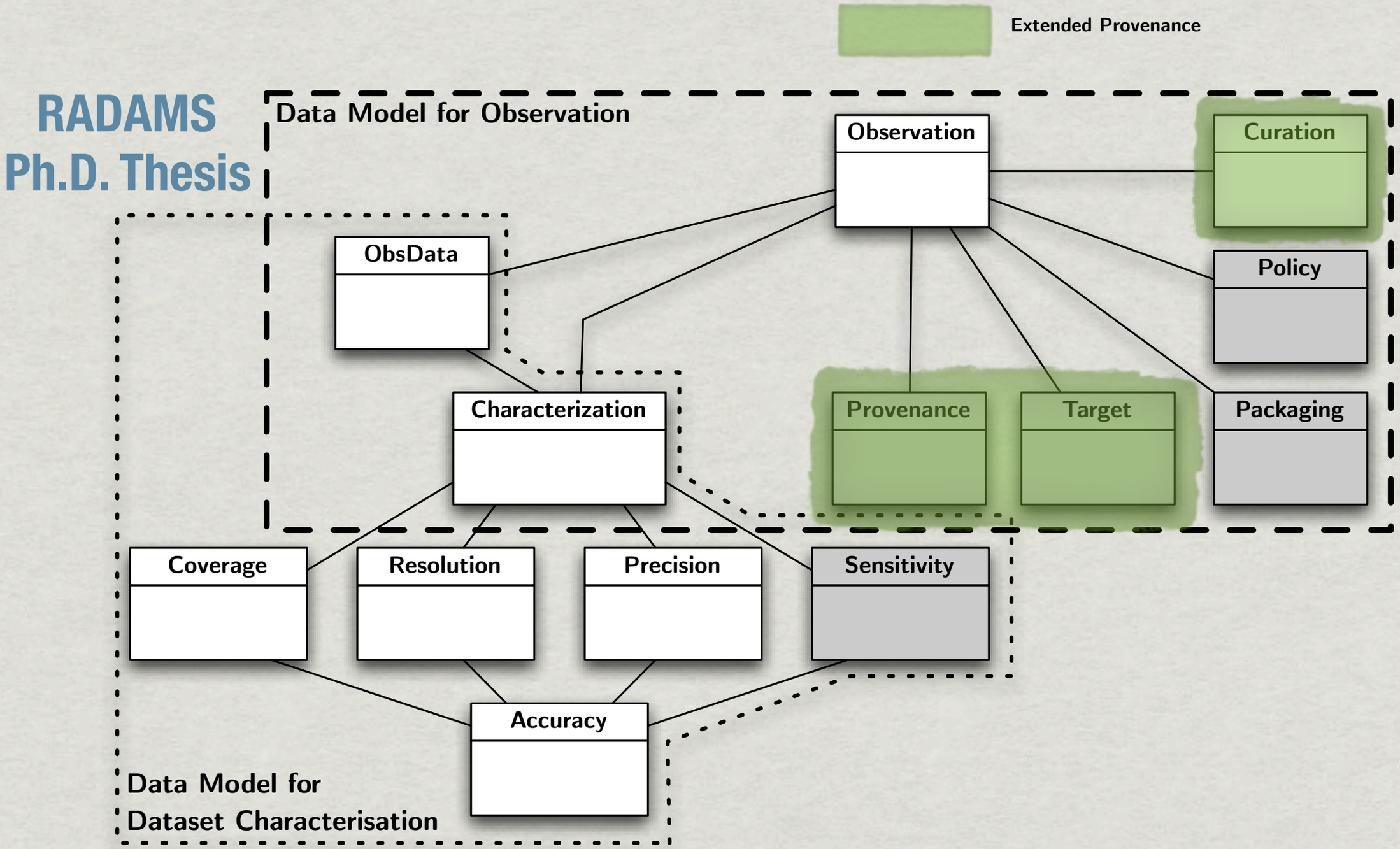


How does Provenance fit in the VO?

RADAMS
Ph.D. Thesis

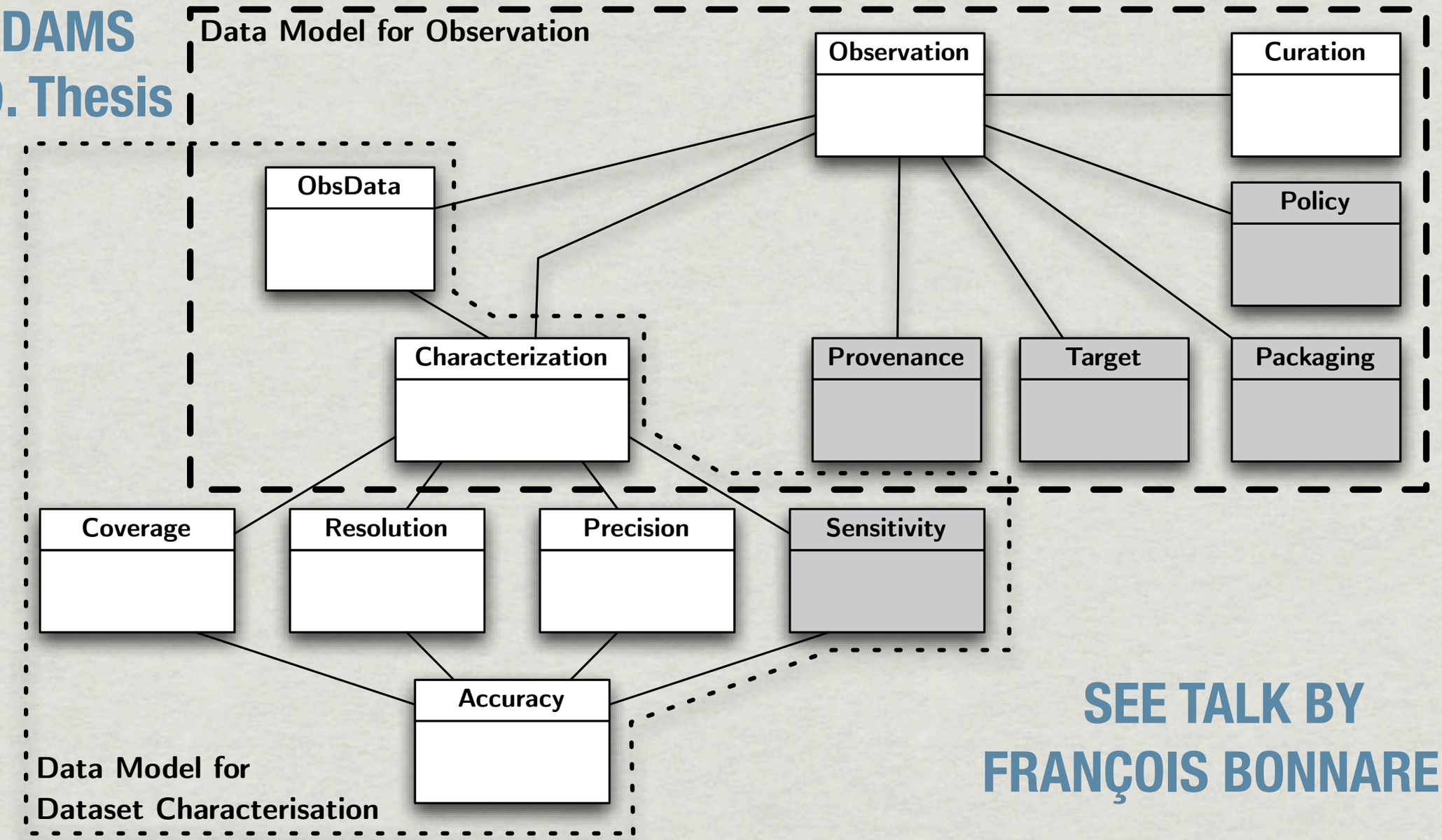


How does Provenance fit in the VO?



How does Provenance fit in the VO?

RADAMS
Ph.D. Thesis



SEE TALK BY FRANÇOIS BONNAREL

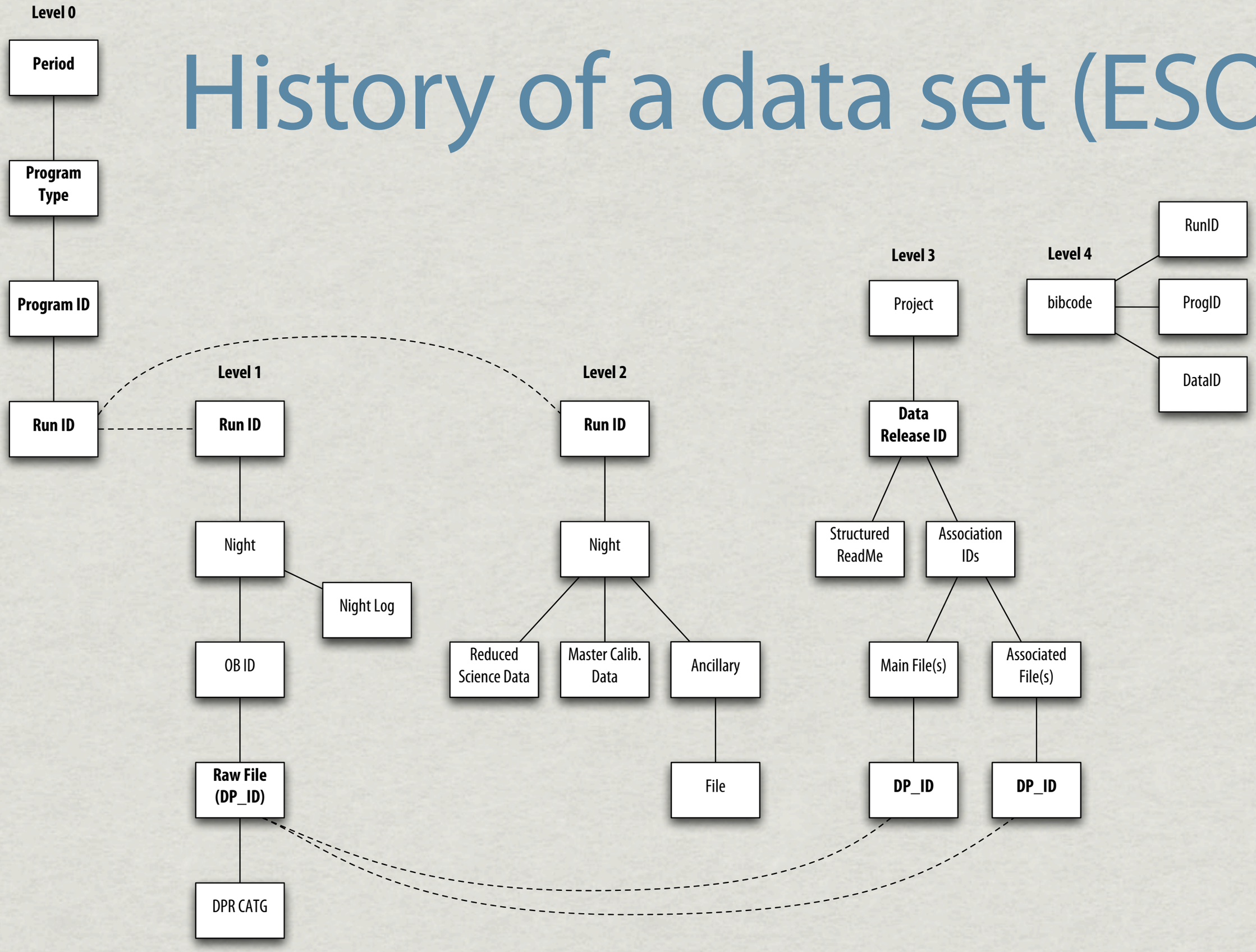
The case(s) for Provenance

The case(s) for Provenance

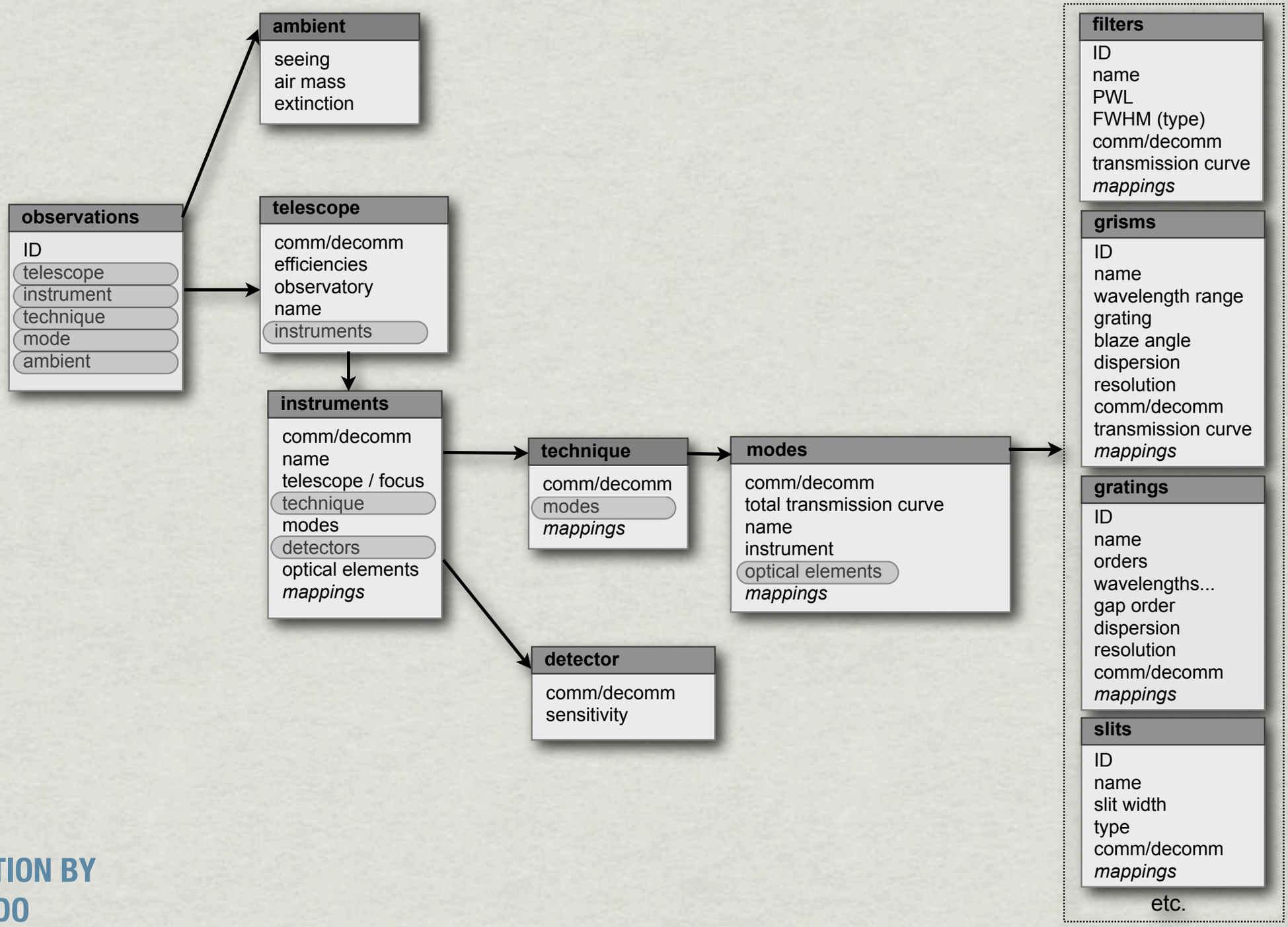
- Use cases for Provenance, centred around the core concepts:
  - **History** of a data set (**documentation**)
  - Establishing **ownership**/providing **attribution**
  - **Quality** assessment

History of a data set (ESO)

History of a data set (ESO)

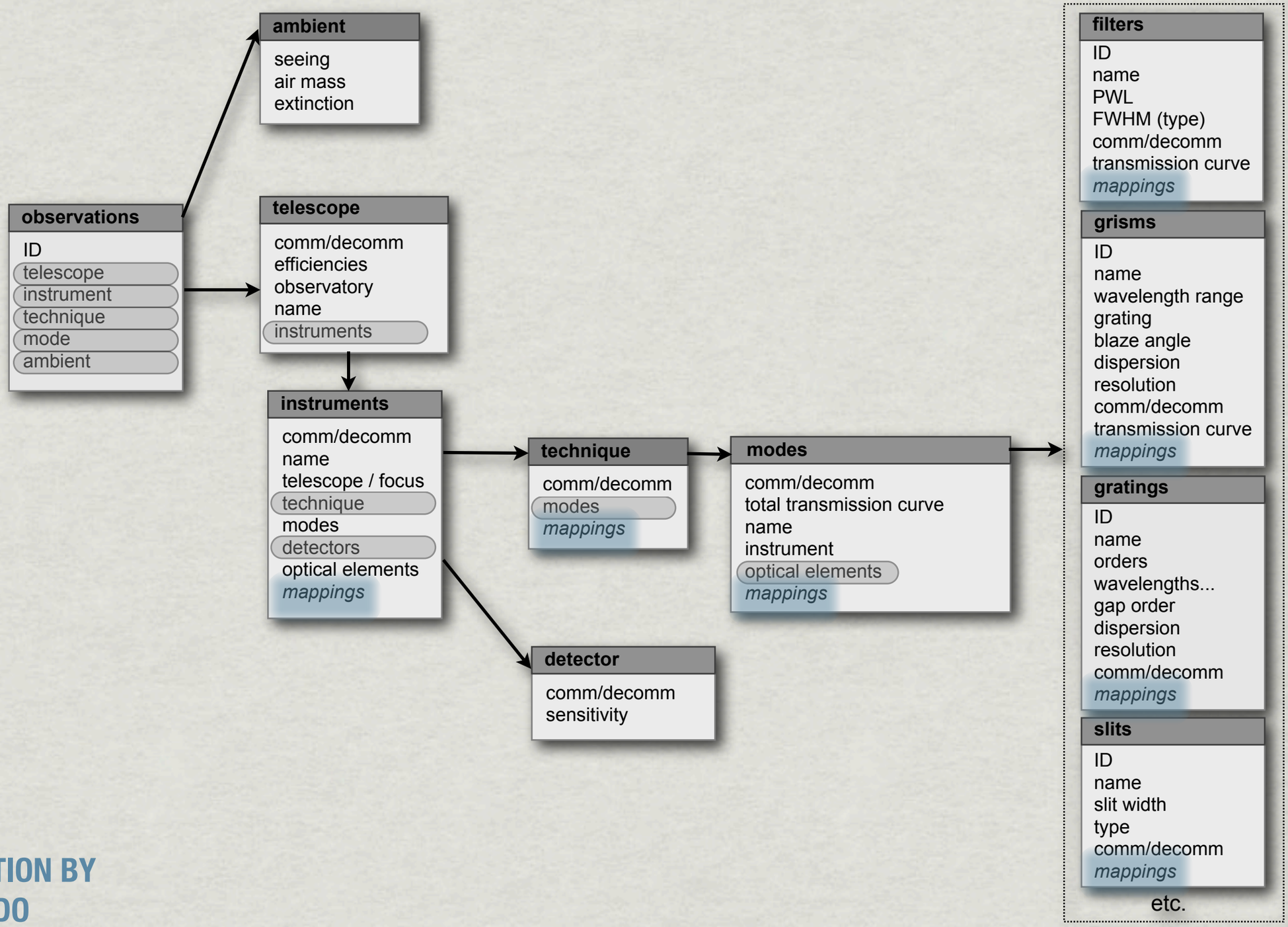


History of a data set (ESO)



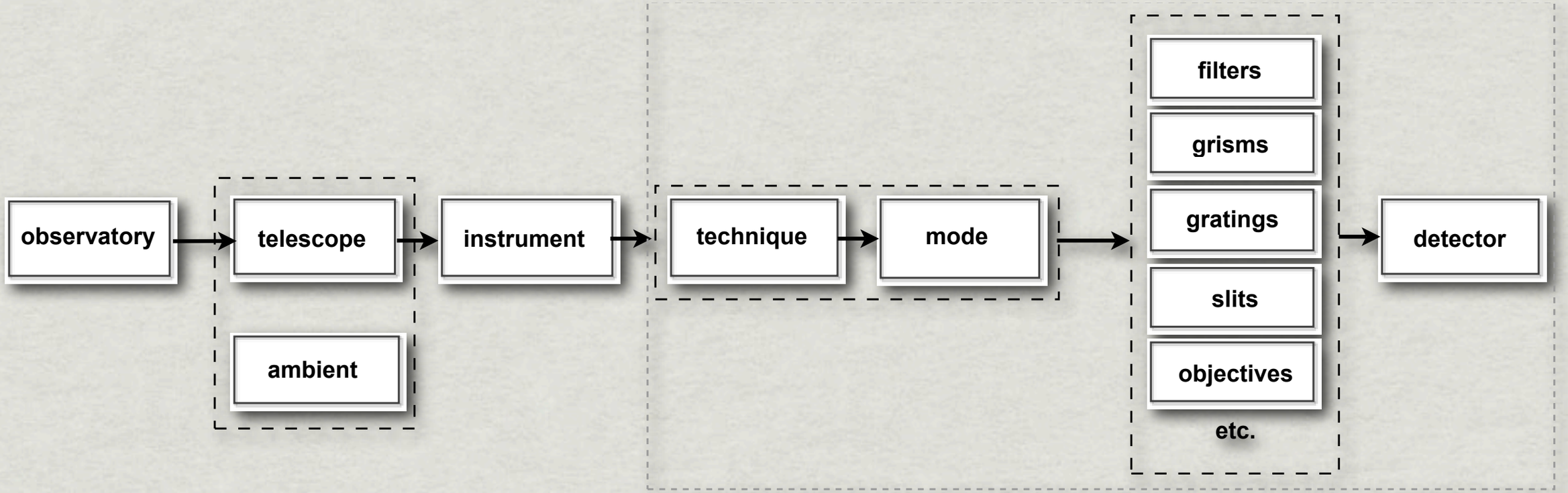
COMPILATION BY
A. DELGADO

History of a data set (ESO)



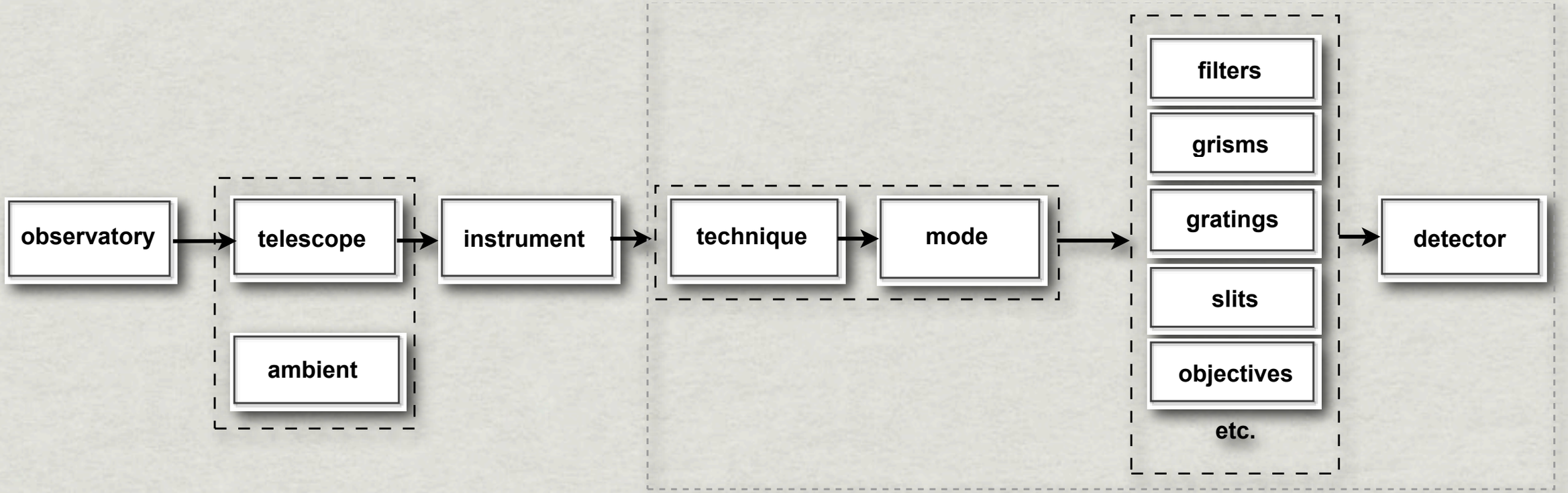
COMPILATION BY
A. DELGADO

History of a data set (ESO)

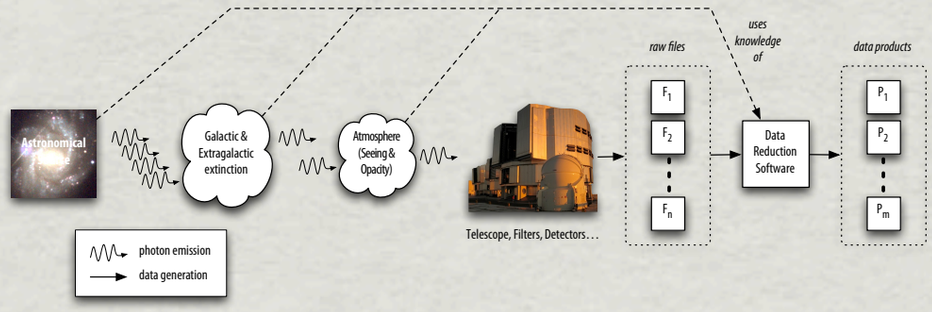


COMPILATION BY
A. DELGADO

History of a data set (ESO)



COMPILATION BY
A. DELGADO



History of a data set (ESO)

- **Telescope Information**
  - Site
  - Dates: First Light, Commissioning…
  - Instruments & Optical Elements
- **Instruments**
  - Technologies
  - Configurations
- **Optical Elements**
  - Installation dates
  - Kind (Filter, Grisms, Gratings…)
  - Names ↔ IDs ↔ Transmission Curves

COMPILATION BY
A. DELGADO

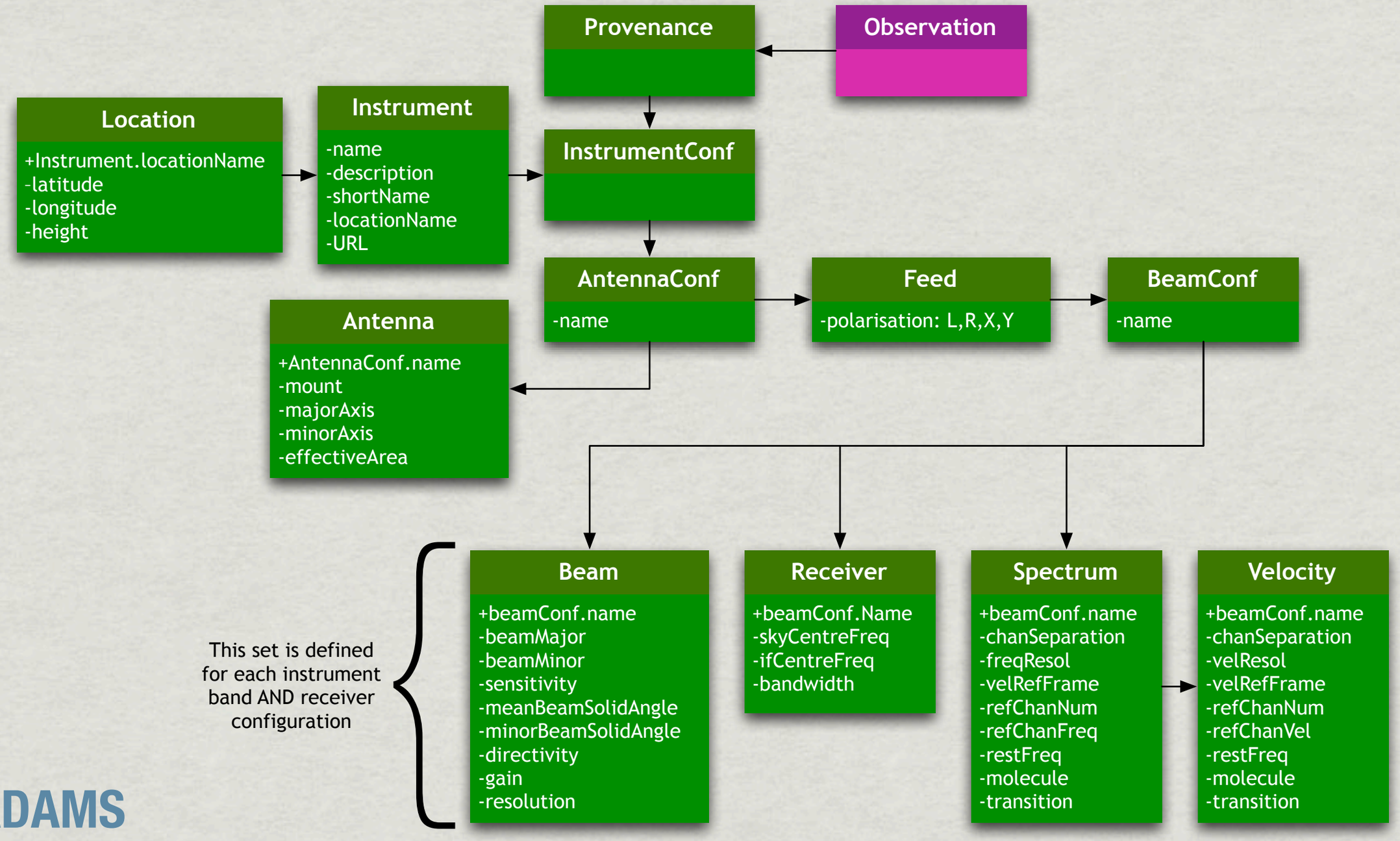
History of a data set (ESO)

- **Telescope Information**
  - Site
  - Dates: First Light, Commissioning…
  - Instruments & Optical Elements
- **Instruments**
  - Technologies
  - Configurations
- **Optical Elements**
  - Installation dates
  - Kind (Filter, Grisms, Gratings…)
  - Names ↔ IDs ↔ Transmission Curves

COMPILATION BY A. DELGADO

TIME CHANGING

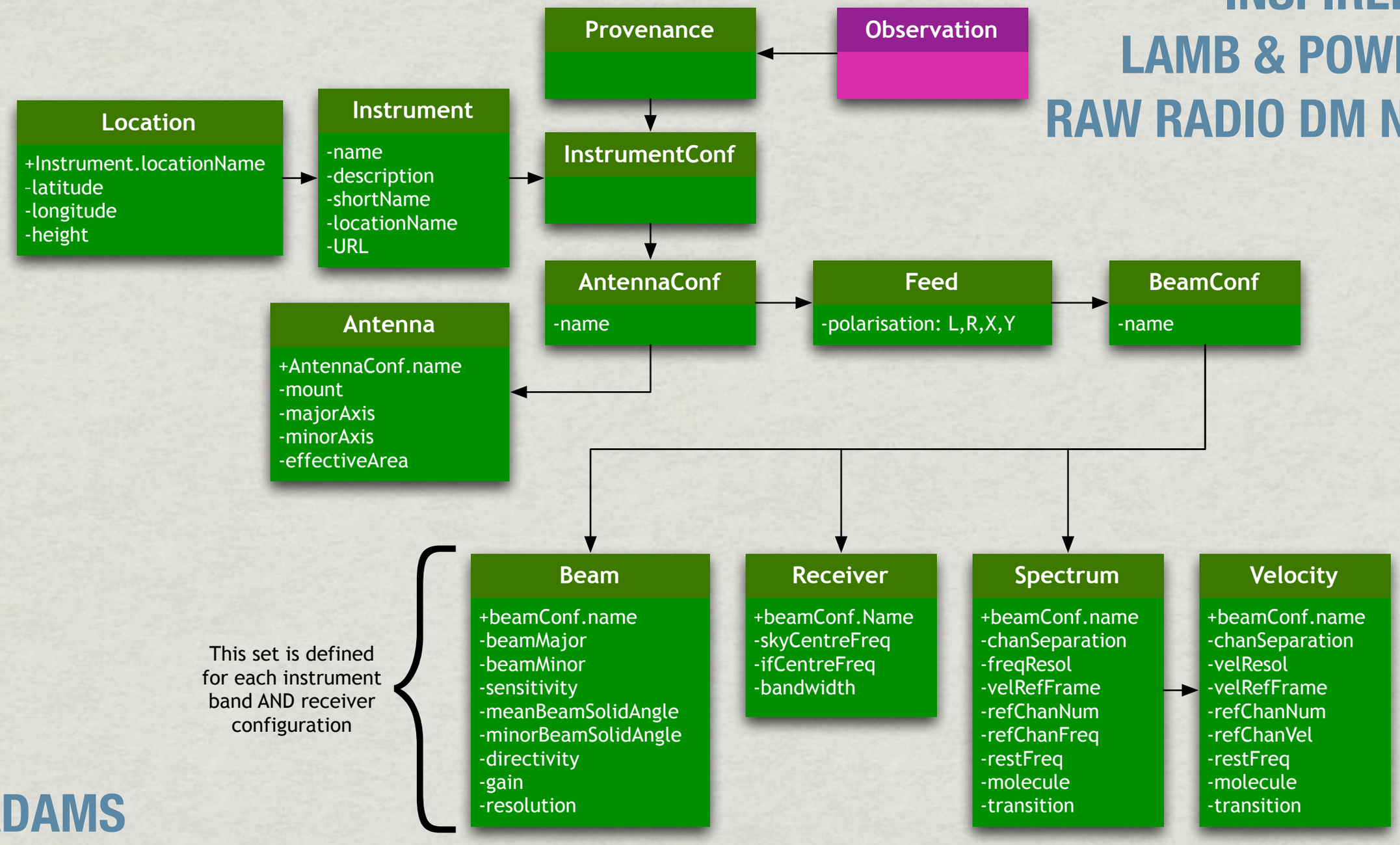
History of a data set



RADAMS

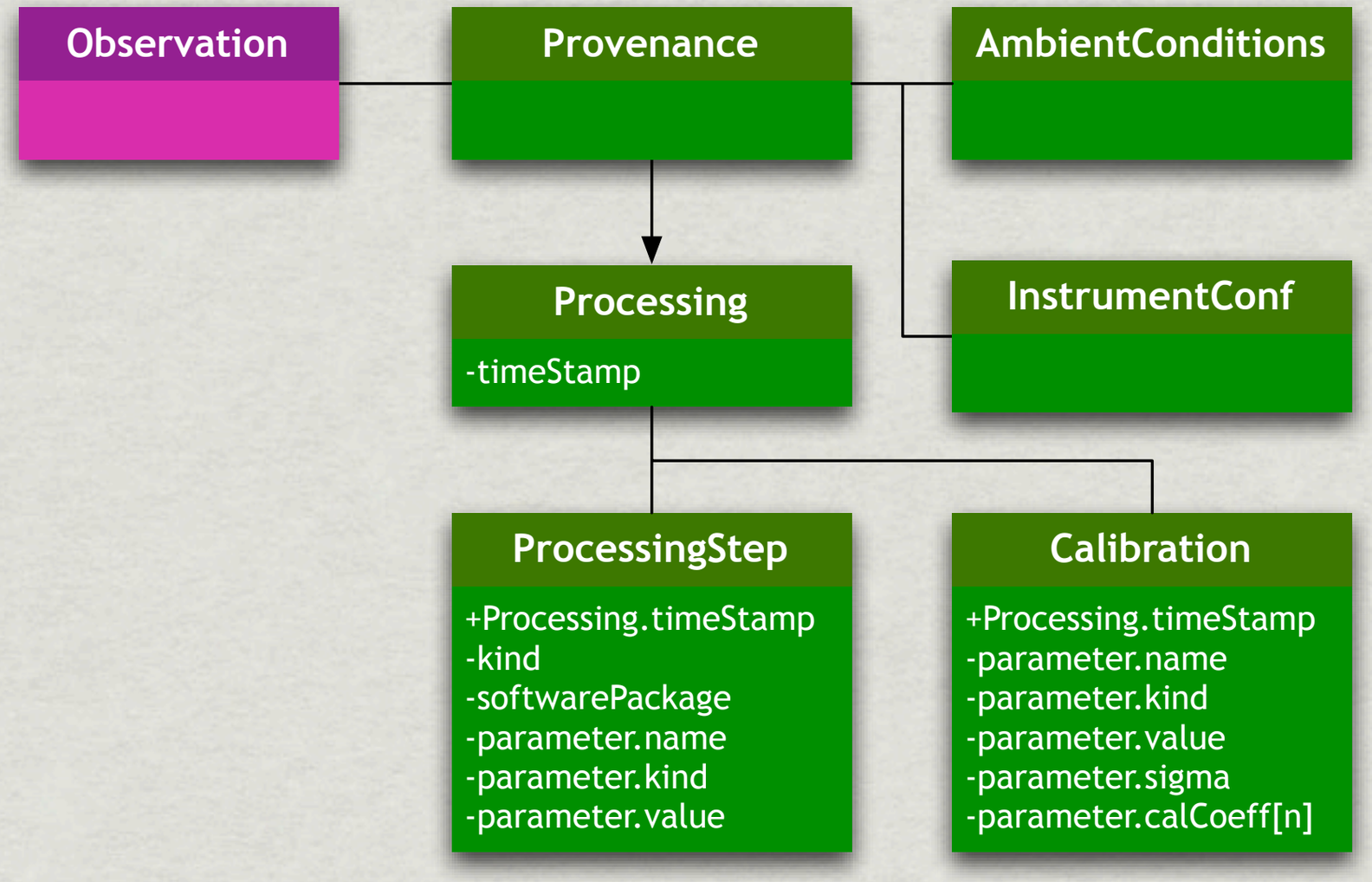
History of a data set

INSPIRED BY LAMB & POWER'S RAW RADIO DM NOTE



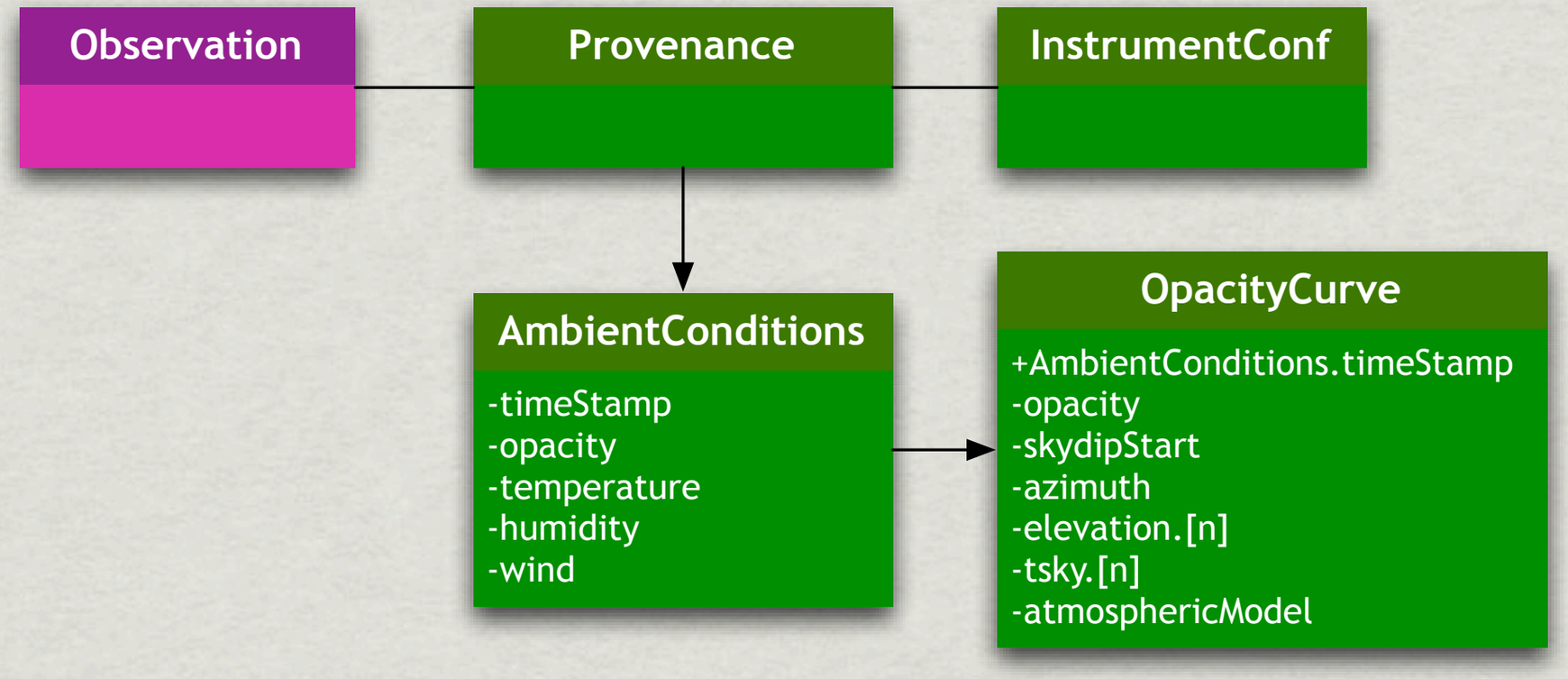
RADAMS

History of a data set



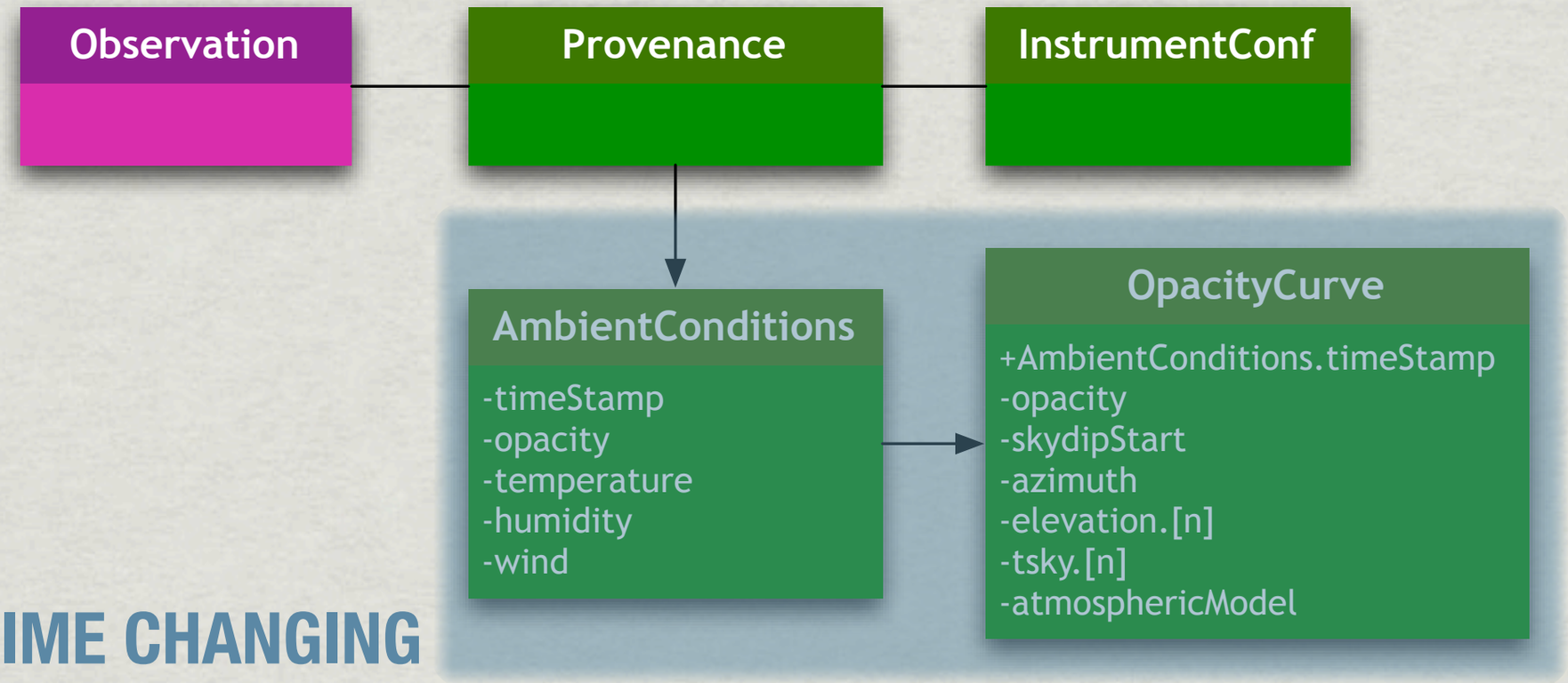
RADAMS

History of a data set



RADAMS

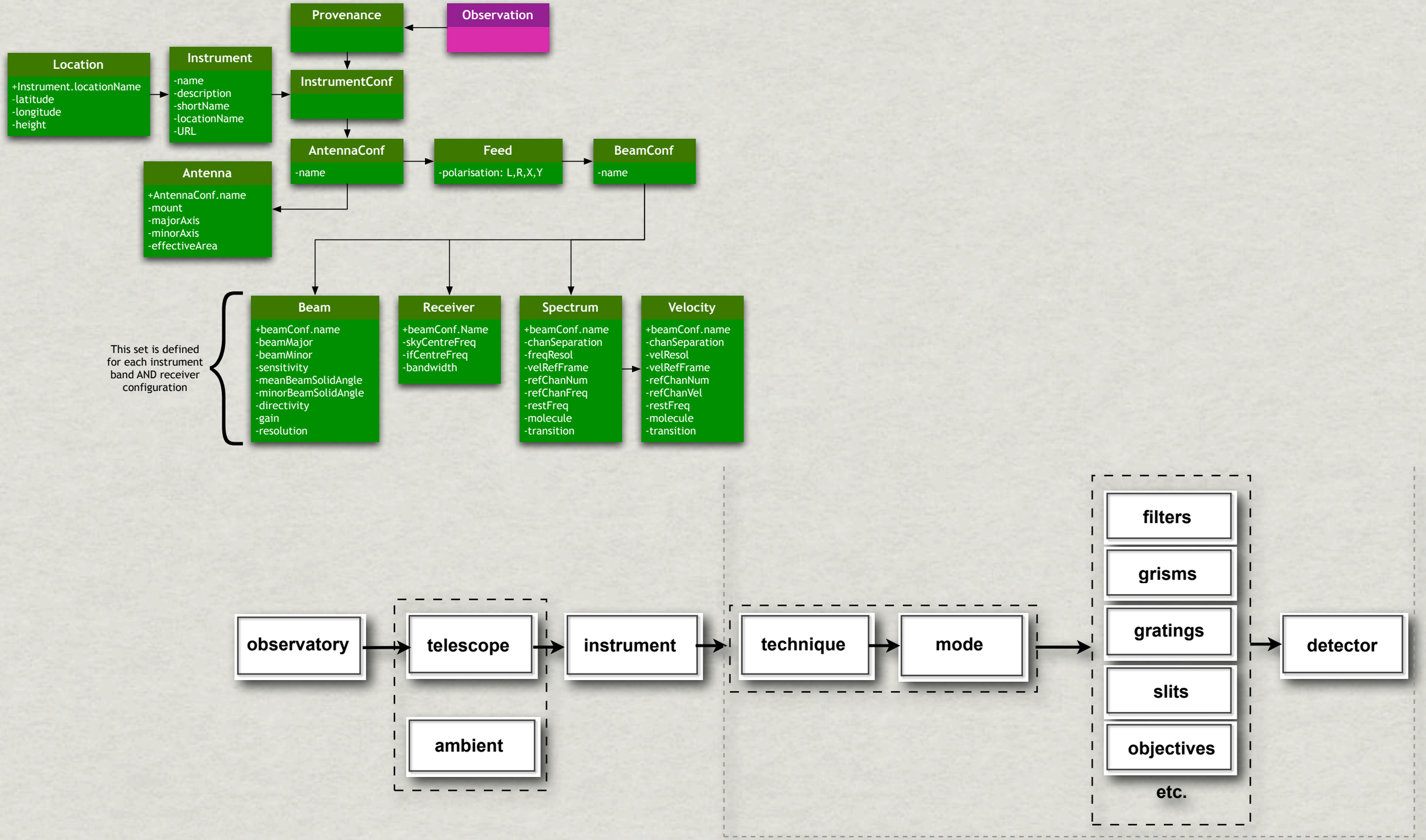
History of a data set



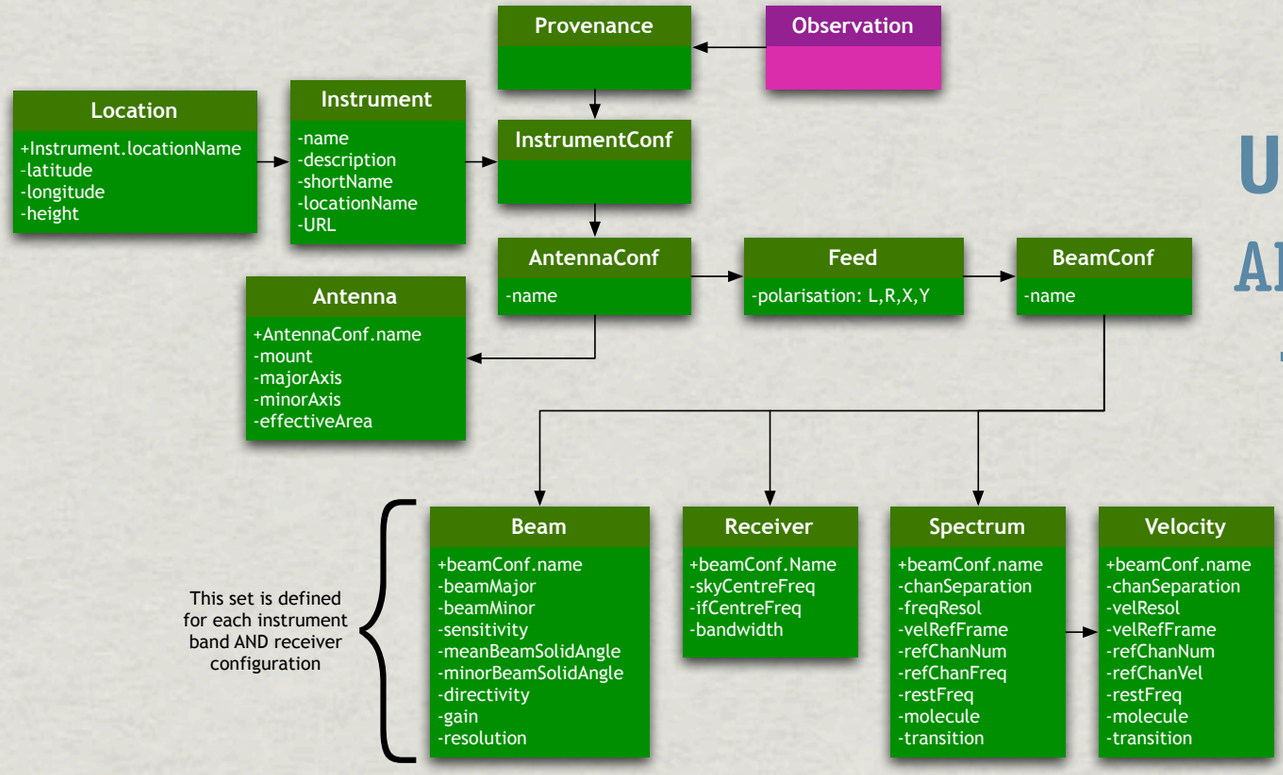
TIME CHANGING

RADAMS

History of a data set

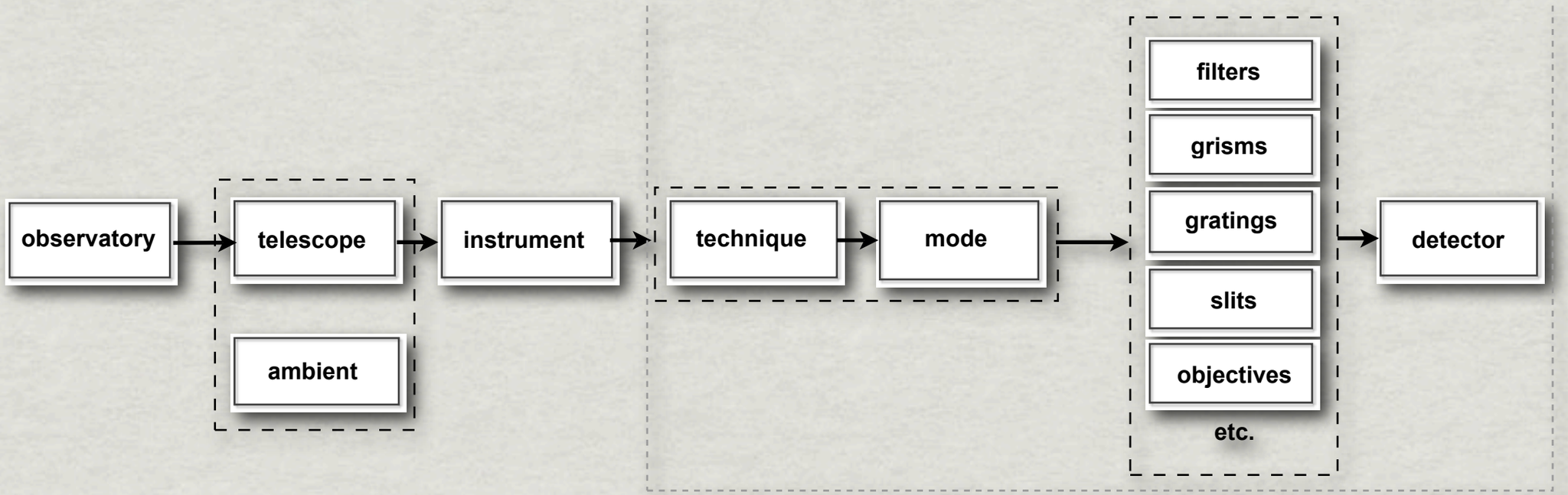


History of a data set



UNIFICATION OF PROVENANCE AT THE `AMBIENT`, `TELESCOPE_CONF` LEVELS:

THE REST IS VERY HETEROGENOUS



History of a data set

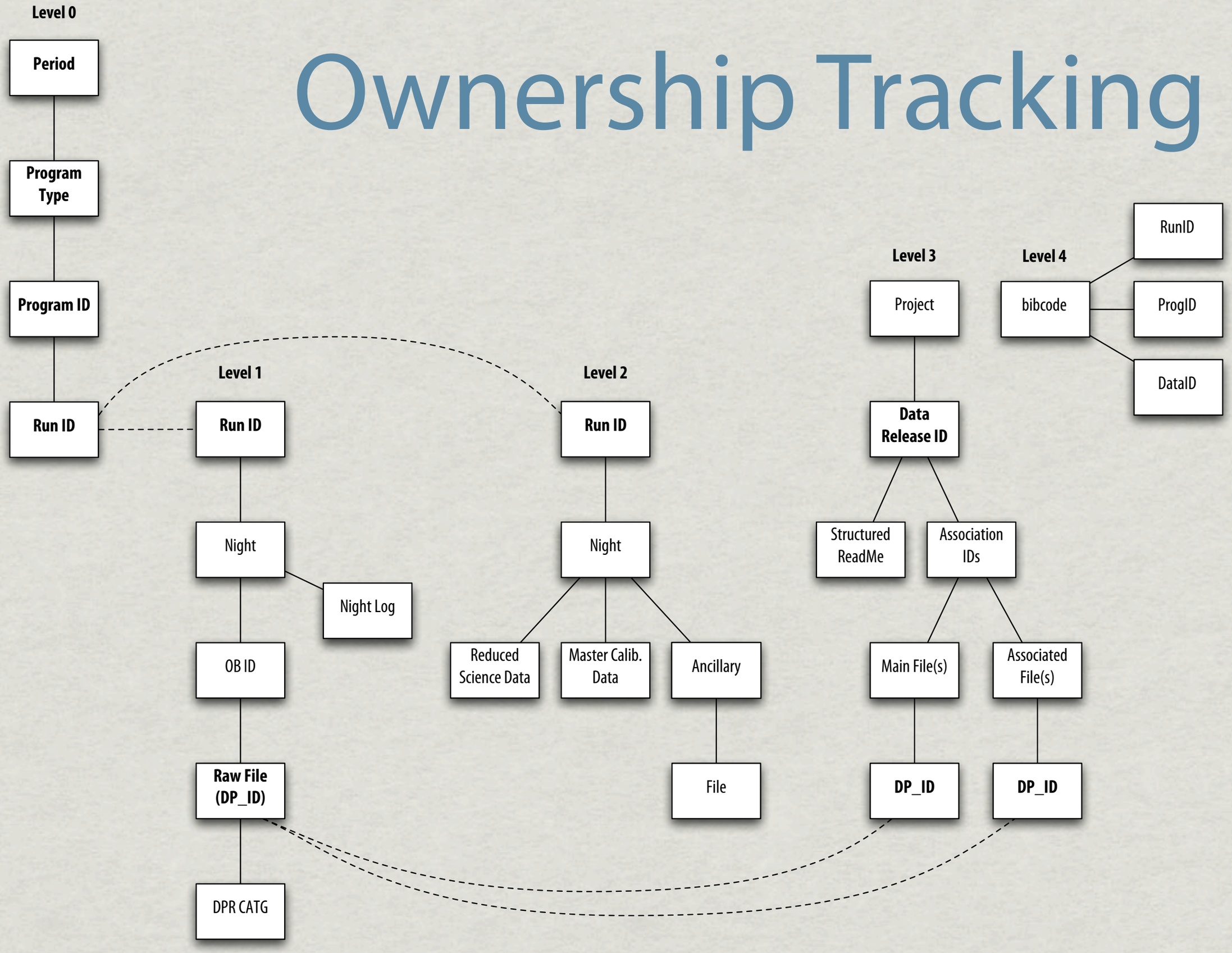
- Can only be done with:
  - Historical archive of:
    - optical elements
    - software (possibly with virtualization)
    - configuration values / fudge factors
    - ambient information
      - DIMM seeing, opacity, conductivity…

History of a data set

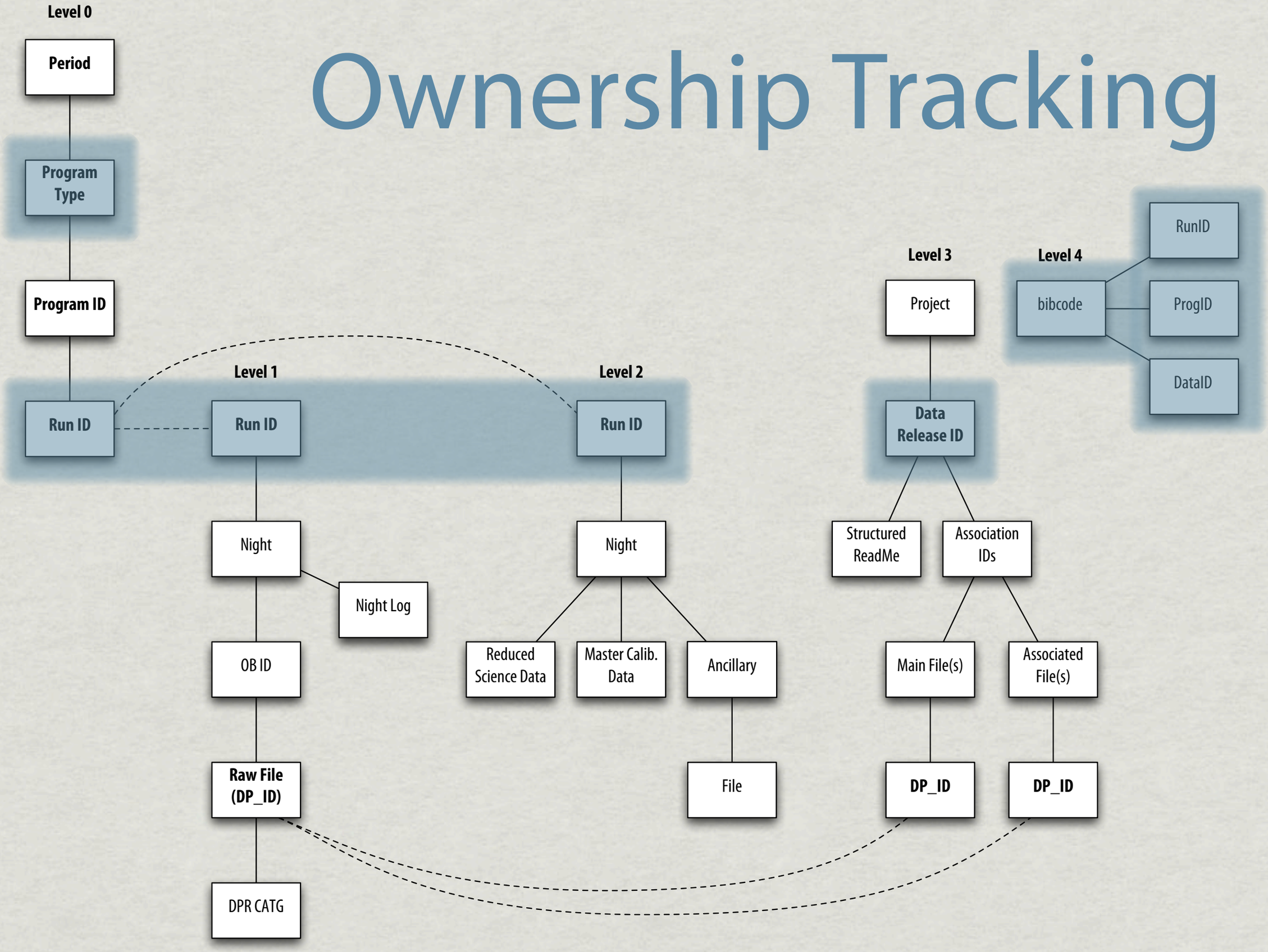
- Can only be done with:
  - Historical archive of:
    - optical elements
    - software (possibly with virtualization)
    - configuration values / fudge factors
    - ambient information
      - DIMM seeing, opacity, conductivity…

+ SYSTEMATIC OBSERVING LOGS!

Ownership Tracking



Ownership Tracking



Ownership Tracking

Ownership Tracking

- Can only **rely** on **unique identifiers** being **maintained**
- Other **VO manipulations** (SAMP messages, cross-matching) can **lose associations**
- Need to provide **services** for identification of key IDs
  - Increase of the role of the IVOA Registry?
- This is **stopping** some **small publishers**!

Ownership Tracking

- Can only **rely** on **unique identifiers** being **maintained**
- Other **VO manipulations** (SAMP messages, cross-matching) can **lose associations**
- Need to provide **services** for identification of key IDs
  - Increase of the role of the IVOA Registry?
- This is **stopping** some **small publishers**!

THIS TRACKING CAN BE USED TO MEASURE PUBLICATION RATIOS

Quality Assessment

Quality Assessment

- Linked to:
  - Acquisition **configuration**
  - **Actual problems** with instruments/telescope
  - **Weather**
  - **Intended usage** of the dataset(s)!

Quality Assessment

- Linked to:
  - Acquisition **configuration**
  - **Actual problems** with instruments/telescope
  - **Weather**
  - **Intended usage** of the dataset(s)!

NO OBJECTIVE, ONE-SIZE-FITS-ALL, QUALITY ASSESSMENT METRIC

Quality Assessment

- Linked to:
  - Acquisition **configuration**
  - **Actual problems** with instruments/telescope
  - **Weather**
  - **Intended usage** of the dataset(s)!

Quality Assessment

- Linked to:
  - Acquisition **configuration**
  - **Actual problems** with instruments/telescope
  - **Weather**
  - **Intended usage** of the dataset(s)!

TIME CHANGING

Queriable Provenance?

- Most queries, on **Characterisation/Target/Curation**
- **Instrument-specific forms** for instrument configurations
  - Allow UType/UCD/IVOAT **key-value pairs**?
  - Instrument-specific *a priori* quality assessment
  - *A posteriori*, usage-specific quality assessment, offline > usage-specific query to data Provenance

Conclusions

Conclusions

- Provenance is an **integral part** of the Observation DM
- Provenance comes with **discipline**, but allows for **quality science**
- Different approach for different kinds of instruments, but all **under the same general framework**
- Provenance should be accessible for any item; **specialised TAP** version?

Conclusions

- Data centres: **consistent naming/coding**, plus mappings for existing data
- **History, history, history**!
- Should we **forget** about **past data**, and focus in the **future**?

Vielen Dank!

STARTING WITH MY GERMAN CLASSES ;-)