Data Registration

INCLUDING ONTOLOGY

Presented by Ivan Galkin, BGD

Outline

- ◉ **Data Registration Status**
  - Promised: ~60 collections (datasets and models)
  - Completed: ~40% at best
  - Train the Trainers (TTT) group, was it helpful? Continue?
    - ➢ No follow-on requests to register more data..
- ◉ **Overview of PITHIA metadata system and motivation**
- ◉ **Status of Ontology Development**
  - *Activity Indicator* and its new dictionary
- ◉ **Where to go from here**
  - Registration review task force is needed

Definitions:

- ⦿ SCHEMA
  - Standard rules for building metadata: which elements and how they are related
    - ➢ ISO O&M is the foundation of our schema
    - ➢ 12 steps for data collections, 3 steps for Catalogues
    - ➢ a.k.a. "Data Model"
- ⦿ ONTOLOGY
  - Standard vocabularies for space physics terms
- ⦿ RESOURCES
  - 12 steps of XML files, prepared by data providers

Quick check of our understanding

these things are part of…

	SCHEMA	ONTOLOGY	RESOURCE
TEMPERATURE	☐	■	☐
OBSERVED PROPERTY	■	☐	☐
PLASMASPHERE DENSITY BY WHISPER	☐	☐	■

DATA REGISTRATION
is adding new information to...

|	YES	NO	MAYBE
SCHEMA	☐	■	☐
RESOURCES	■	☐	☐
ONTOLOGY	☐	☐	■

Why all this XML nonsense?

- Modern data infrastructures are FAIR
  - Findable, Accessible, Interoperable, Reusable
  - Next level of metadata fidelity
    - Full description of data content and input parameters
      - Do not confuse parameters with observed properties
      - Example: observed property of IRI is foF2, input parameter is sunspot number
  - ISO for Observations and Measurements is our OM matra
    - Simplified for PITHIA

Schema: ISO Standard "O&M"
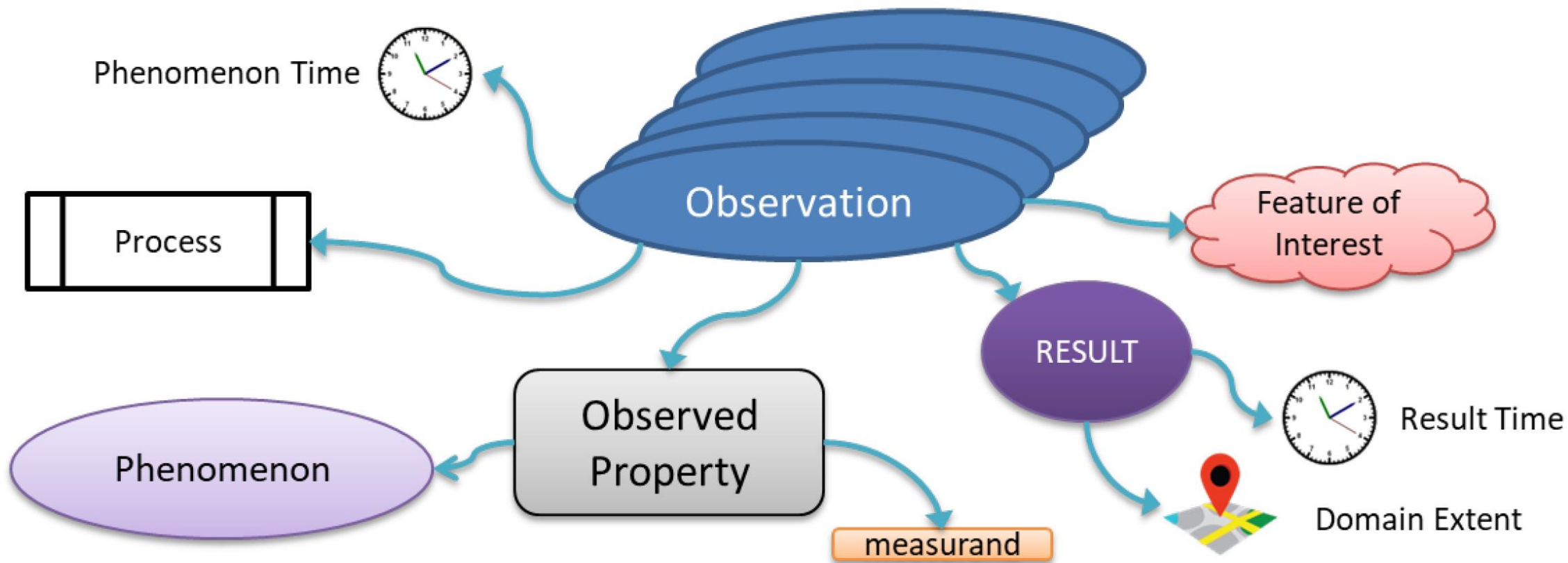
Observed Property of a **Feature of Interest** describes the **Phenomenon** for which the **Observation** **Result** provides an estimate of its **Measurand** value, using a **Process**

Schema for PITHIA Data Collections
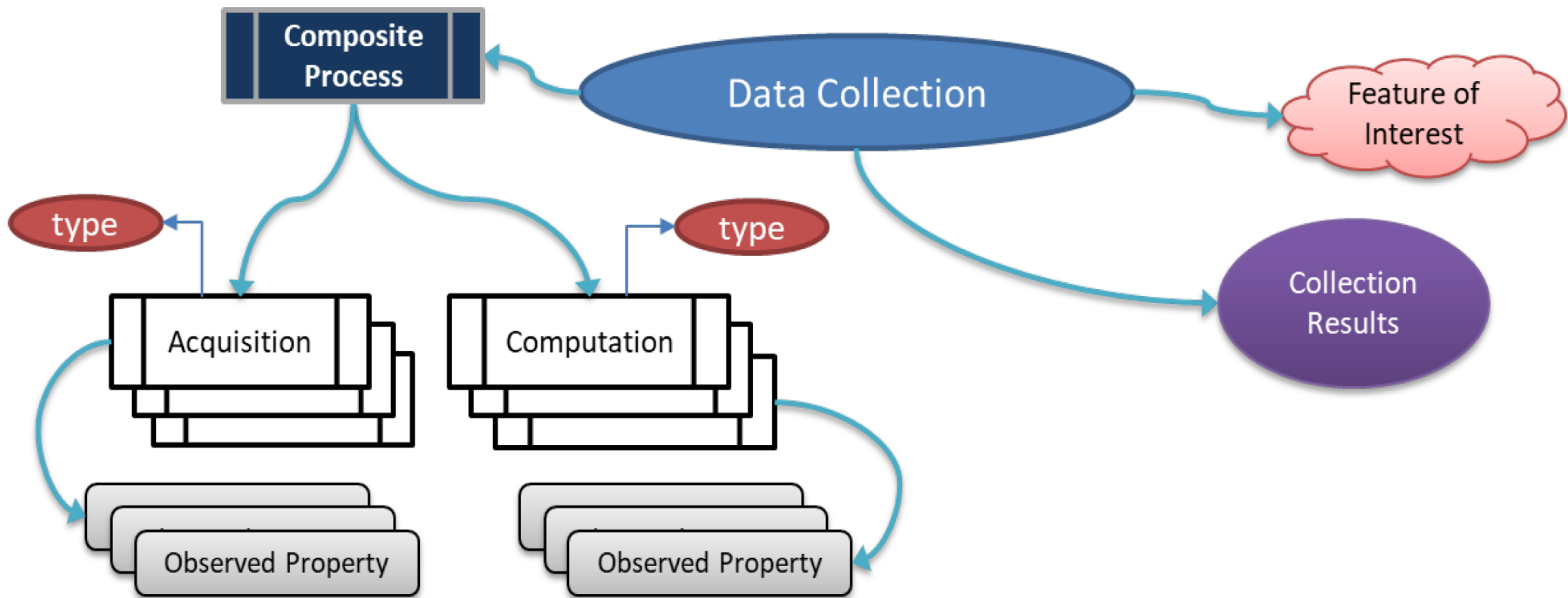
Type is Searchable   type

- Acquisition's Type is **Instrument Type**

- Computation's Type is **Computation Type**

- Search by Type is available in eSC
  - Looking for measurements – specify Instrument Type
  - Looking for model data – specify Computation Type

- Advanced topic: assimilative models have both Instrument Type and Computation Type defined
  - Instrument Type is given for the assimilated measurements

Schema for PITHIA: Catalogue
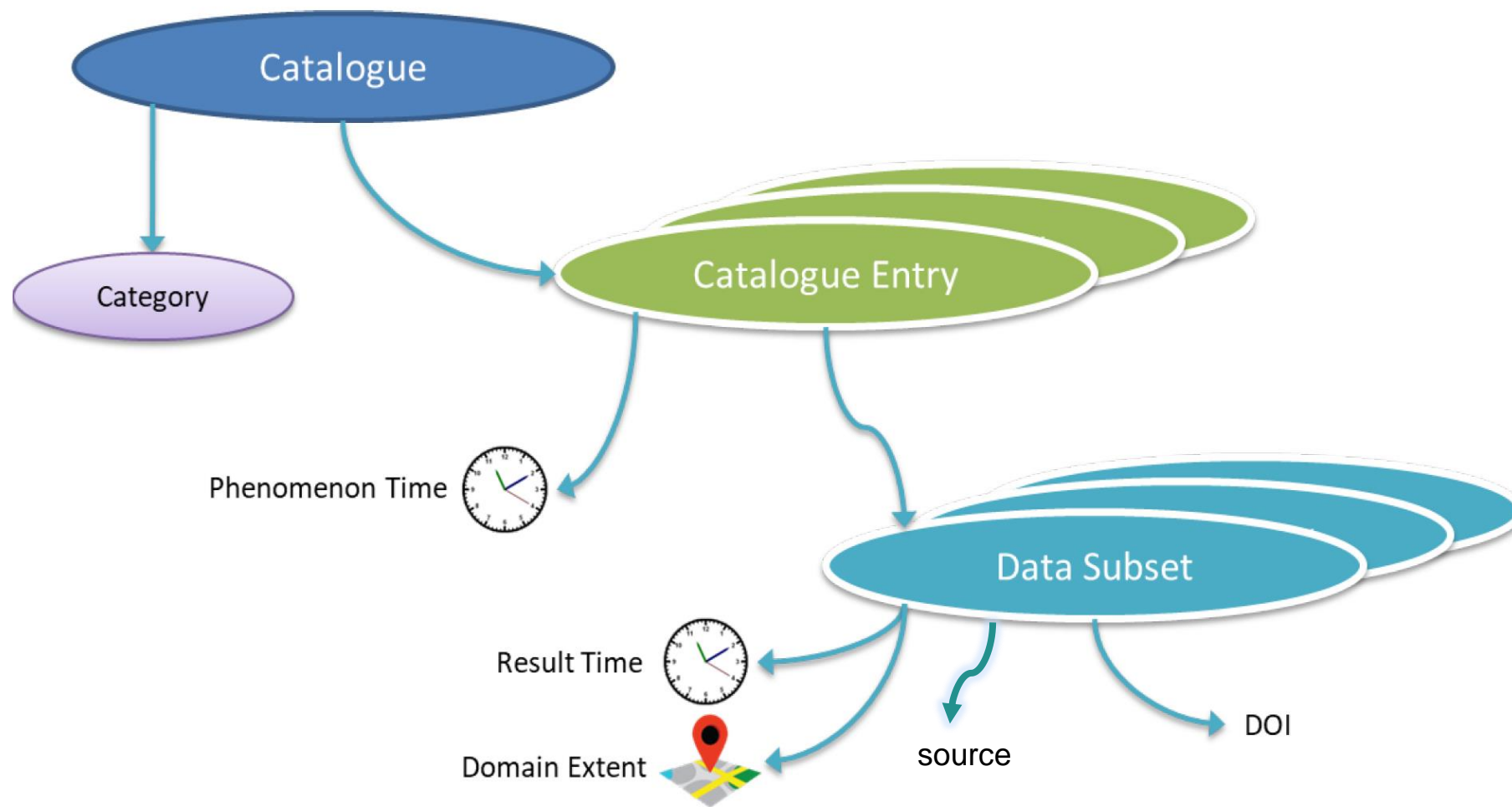
Search, selection, discovery

- ◉ **Search data collections by content**
  - **Feature of Interest and Observed Properties**
  - **Computation and Instrument Types**
- ◉ Search data availability by context
  - By coincidence
    - ➤ e.g., find data by helio- and geospace activity thresholds
      - We will be able to search by activity
        - This service would not be arranged by PITHIA providers, there are standard tools
      - Problem: we do not record data availability for PITHIA observational resources
      - Solution: provide new type of interaction with the Data Collection, "availability"
  - By conjunction
    - ➤ Matching time <u>and</u> location
      - Problem: we do not record S/C ephemeris, either
      - Solution: provide access to external orbit propagators

12 STEPS of Registration

12. Data Collection — FOI

11. Process

10. Computation

9. Computation Capabilities — OP

8. Acquisition

7. Acquisition Capabilities — OP

6. Instrument

5. Operation

4. Platform

3. Project

2. Individual

1. Organisation

OP and FOI are used in searching by content"
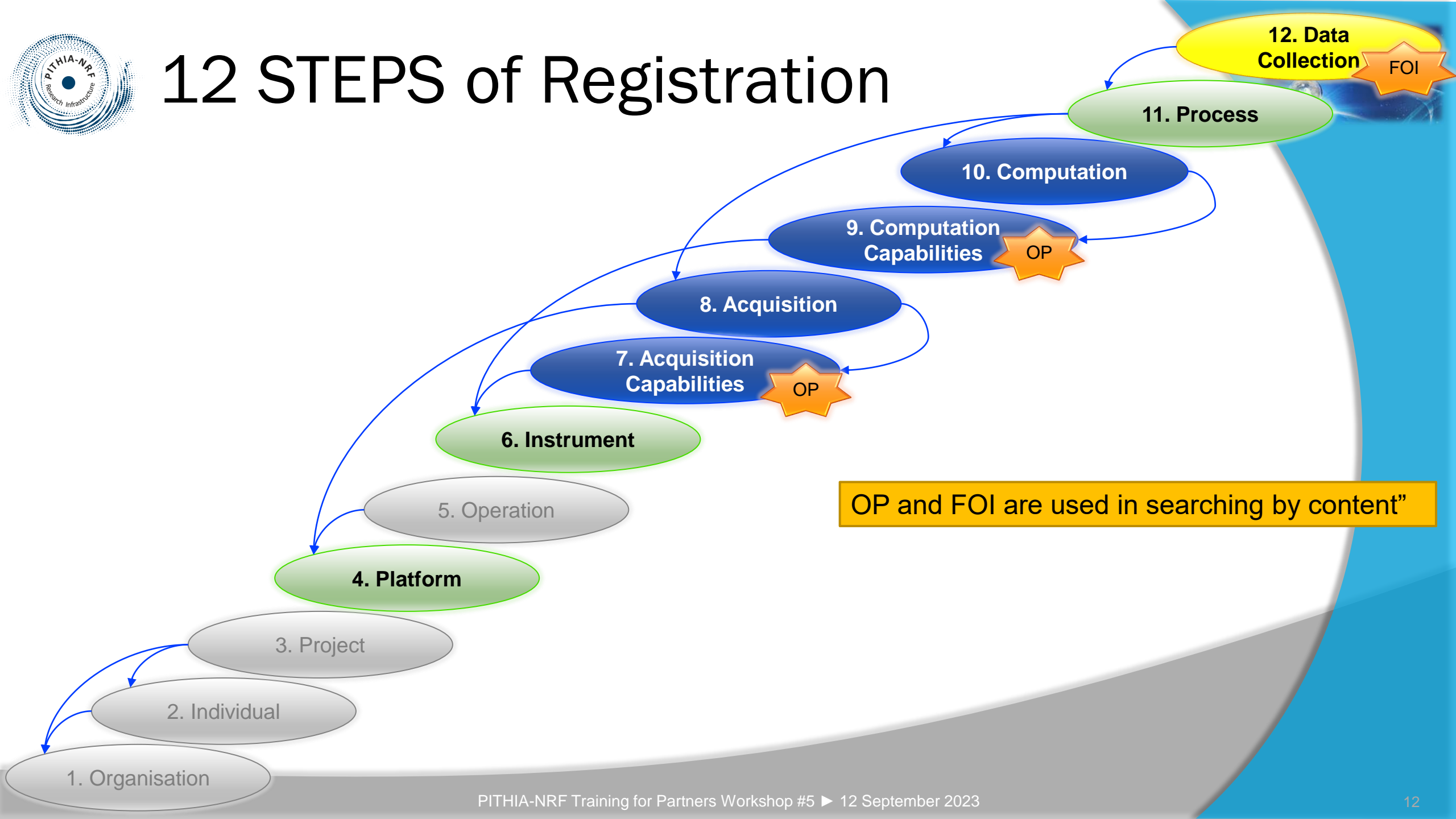
Acquisition

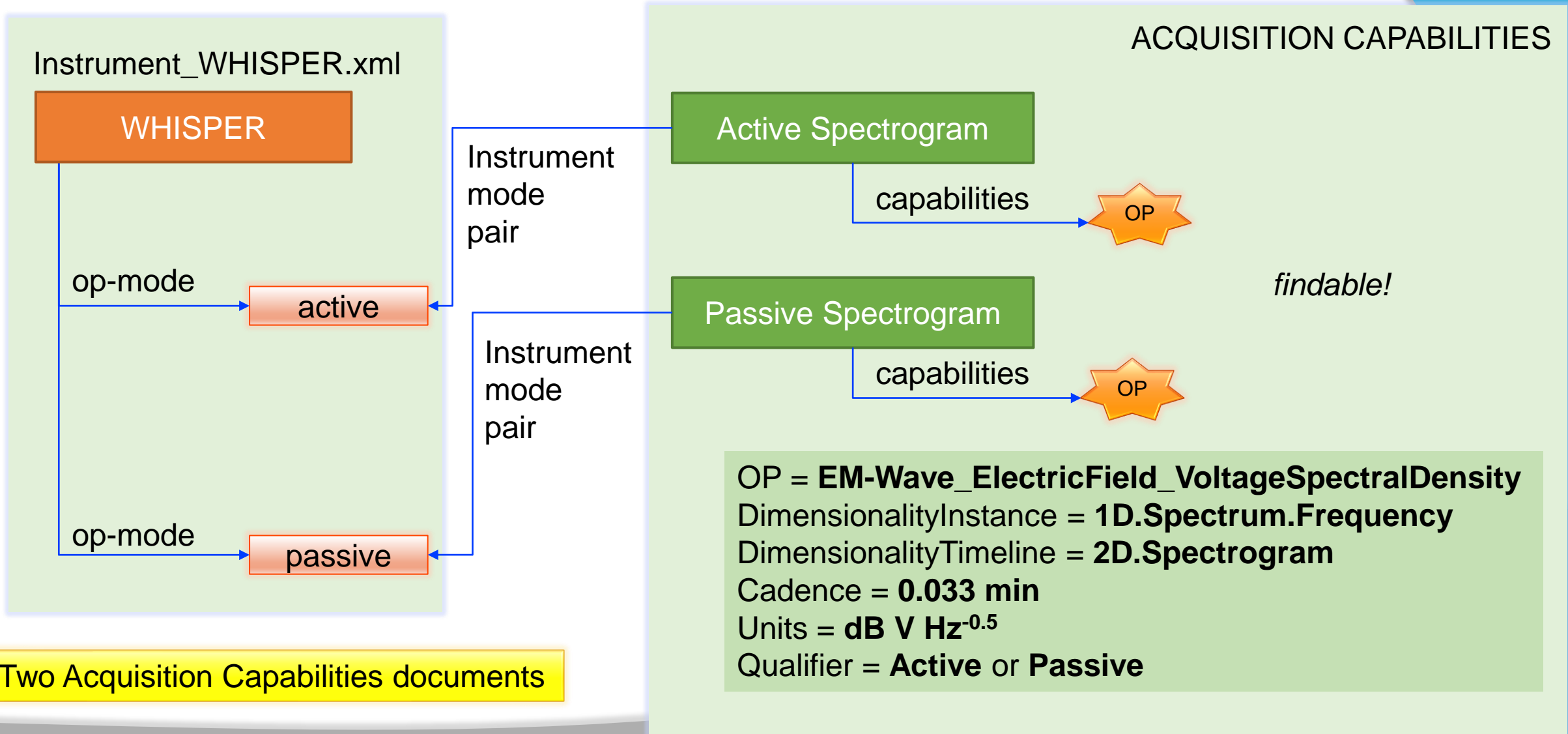- Acquisition links Platforms and Acquisition Capabilities
  - Acquisition Capabilities has Instrument
  - Instrument has different modes of operation
  - Example: ionosonde network
    - ~10 different ionosonde brands and their different capabilities
    - ~120 different ionosonde observatories (platforms)
- Advanced topic: history of different ionosonde brands at the same observatory location

Example: WHISPER on CLUSTER (2)

ACQUISITION CAPABILITIES

Instrument_WHISPER.xml

WHISPER

op-mode → active

op-mode → passive

Instrument mode pair → Active Spectrogram

Instrument mode pair → Passive Spectrogram

Active Spectrogram → capabilities → OP

Passive Spectrogram → capabilities → OP

findable!

OP = **EM-Wave_ElectricField_VoltageSpectralDensity**
DimensionalityInstance = **1D.Spectrum.Frequency**
DimensionalityTimeline = **2D.Spectrogram**
Cadence = **0.033 min**
Units = **dB V Hz$^{-0.5}$**
Qualifier = **Active** or **Passive**

Two Acquisition Capabilities documents

Example: WHISPER on CLUSTER (3)

Acquisition Capabilities

- Active Spectrogram (OP)
- Passive Spectrogram (OP)

Platform

- SALSA
- SAMBA
- RUMBA
- TANGO

ACQUISITION

Acquisition_WHISPER_ActiveSpectrogam.xml

- Active Spectrogram
- link — capability links
- link
- link
- link

Passive Spectrogram

(the same)
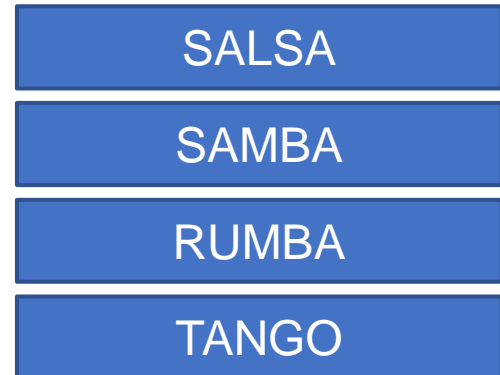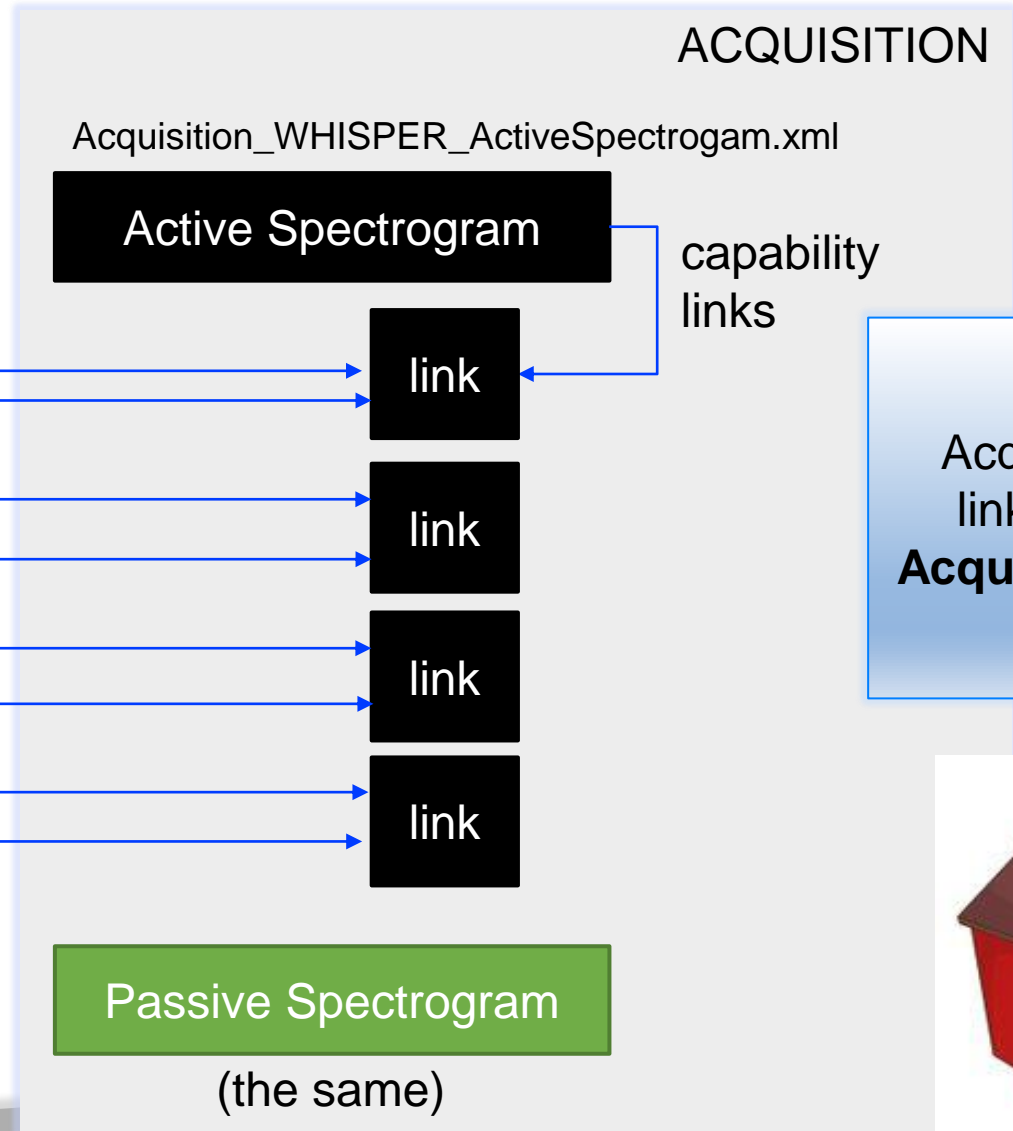
Acquisition document links **Platforms** and **Acquisition Capabilities**

Two Acquisition documents

Example: WHISPER on CLUSTER (4)

COMPUTATION CAPABILITIES

Selection of the optimal spectrogram for processing
(**active**, **passive** or **combined**)

Spectrogram Selection

capabilities → OP

findable!

OP = **EM-Wave_ElectricField_VoltageSpectralDensity**
DimensionalityInstance = **1D.Spectrum.Frequency**
DimensionalityTimeline = **2D.Spectrogram**
Cadence = **0.033 min**
Units = **dB V Hz$^{-0.5}$**
Qualifier = **Active** or **Passive**

Electron Plasma Frequency

capabilities → OP

findable!

OP = **ResonanceFrequency_Electron_Plasma**
DimensionalityInstance = **1D.Points**
DimensionalityTimeline = **2D.TimeSeries.Stacked**
Cadence = **0.033 min**
Units = **kHz**

Dimensionality is 1D.Points because of 4 platforms in Cluster

Two Computation Capability documents

Example: WHISPER on CLUSTER (5)

Computation document links 4 Cluster Platforms
with WHISPER Computation Capabilities
(just like Acquisition does)

Two Computation documents

Example: WHISPER on CLUSTER (6)

PROCESS

- Acquisition: Active Spectrogram
- Acquisition: Passive Spectrogram
- Computation: Spectrogram Selection
- Computation: Electron Plasma Frequency from Spectrogram
- Computation: Plasma Density from fpe

DATA COLLECTION

- Procedure
- Collection Results
  - URL to resource: WHISPER landing page
- Link to "Project_IASB_Cluster"
- Data Level: L2V
- Data Quality Flag
- Metadata Quality Flags
- Access Permission (licence)
- Contact information: POC and organisation

One Process document

One Data Collection document

Example: DTM2020 Thermosphere

COMPUTATION CAPABILITIES

Neutral Density

capabilities

findable!

All Observed Properties:
DimensionalityInstance = **2D.Map**
Units = **gcm^{-1}**

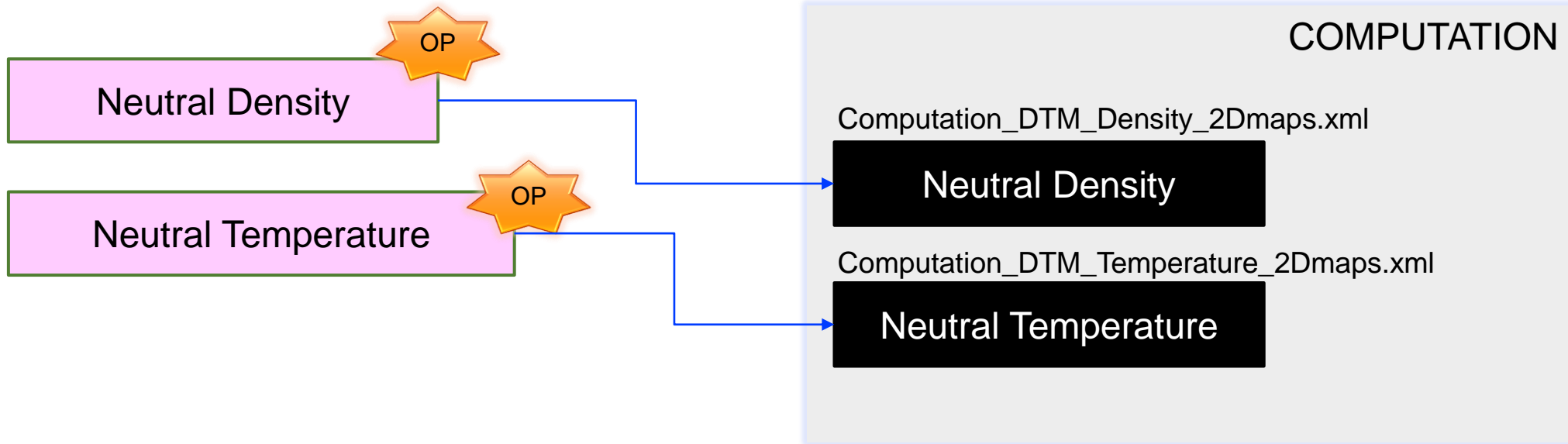OP = **Density of Neutral, O, O2, N2, He**

Neutral Temperature

capabilities

findable!

All Observed Properties:
DimensionalityInstance = **2D.Map**
Units = **degree**

OP = **Temperature of Neutral, O, O2, N2, He**

Two Computation Capability documents

Computation Capabilities:

```
                                    COMPUTATION

                        Computation_DTM_Density_2Dmaps.xml

Neutral Density  (OP) ─────────►    ┌──────────────────┐
                                    │  Neutral Density │
                                    └──────────────────┘

                        Computation_DTM_Temperature_2Dmaps.xml

Neutral Temperature (OP) ──────►    ┌────────────────────────┐
                                    │  Neutral Temperature   │
                                    └────────────────────────┘
```

Two Computation documents

Example: DTM2020 Thermosphere (3)

PROCESS

Computation: Neutral Density 2D Map

Computation: Neutral Temperature 2D Map

DATA COLLECTION

- Procedure
- Link to "Project_CNES_DTM" doc
- Data Level: L4
- Data Quality Flag: DQ4
- Metadata Quality Flags
- Access Permission (licence)
- Contact information: POC and organisation

One Process document

One Data Collection document

Ontology work

- Activity Indicator is new Computation Type
  - Solar
  - Magnetospheric
  - Ionospheric
  - Geomagnetic
- Advanced topic: all indicators have observed property definition(s)
  - Where?

COMPUTATION TYPE

Activity Indicator
- Solar
  - Sunspot Number
    - 2.0
  - $f_{10.7}$
    - observed
    - adjusted
- Magnetospheric
- Ionospheric
  - S4
  - σ_φ
  - ROT
  - ROTI
- Geomagnetic
  - global
    - Kp
    - Hp30
    - Hp60
    - Ap
    - ap
    - ap30
    - ap60
  - local
    - a
    - K

Model
- Theoretical
- Empirical
- Assimilative
- Mathematical
  - Fitting
  - Geometry
  - Tomography
  - Discrete Line Integral
  - Mapping Function
- Semi-Empirical
- Semi-Theoretical

Measurement Processing
- Data Conditioning
- Image Interpretation
  - Automatic
    - Ionogram Autoscaling
    - Plasmagram Autoscaling
    - Spectrogram Autoscaling
  - Manual
- GNSS Signal Processing

Where do we go from here?

◉ Keep cranking!

◉ Train the Trainers

◉ Review board for providers' submissions