

Study of provenance data for the SKA Science Data processor

Julián Garrido, Susana Sánchez
On behalf of the SKA-ProvSDP project



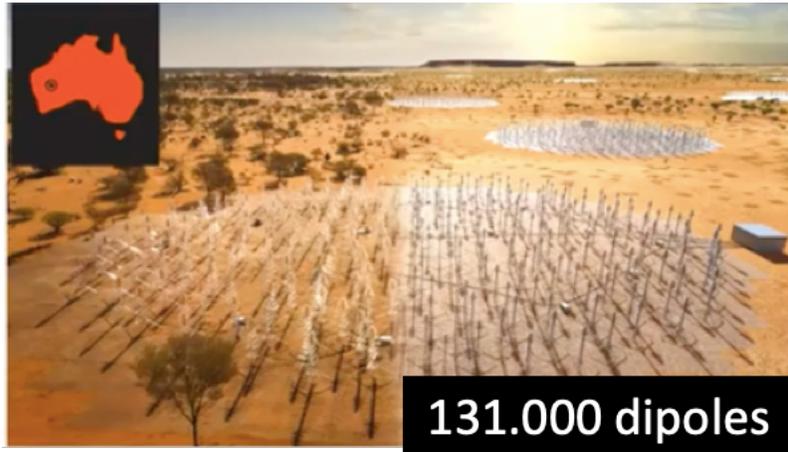
OPEN SCIENCE AND ARCHIVE

SKA-ProvSDP: “Supporting Open Science in the SKA by collecting Provenance from its Science Data processor”

Funded by “Centre for the Development of Industrial Technology (CDTI)”

→ Partners: GMV company, University of Cambridge and IAA-CSIC

→ 5 Technical reports under review by the SKAO

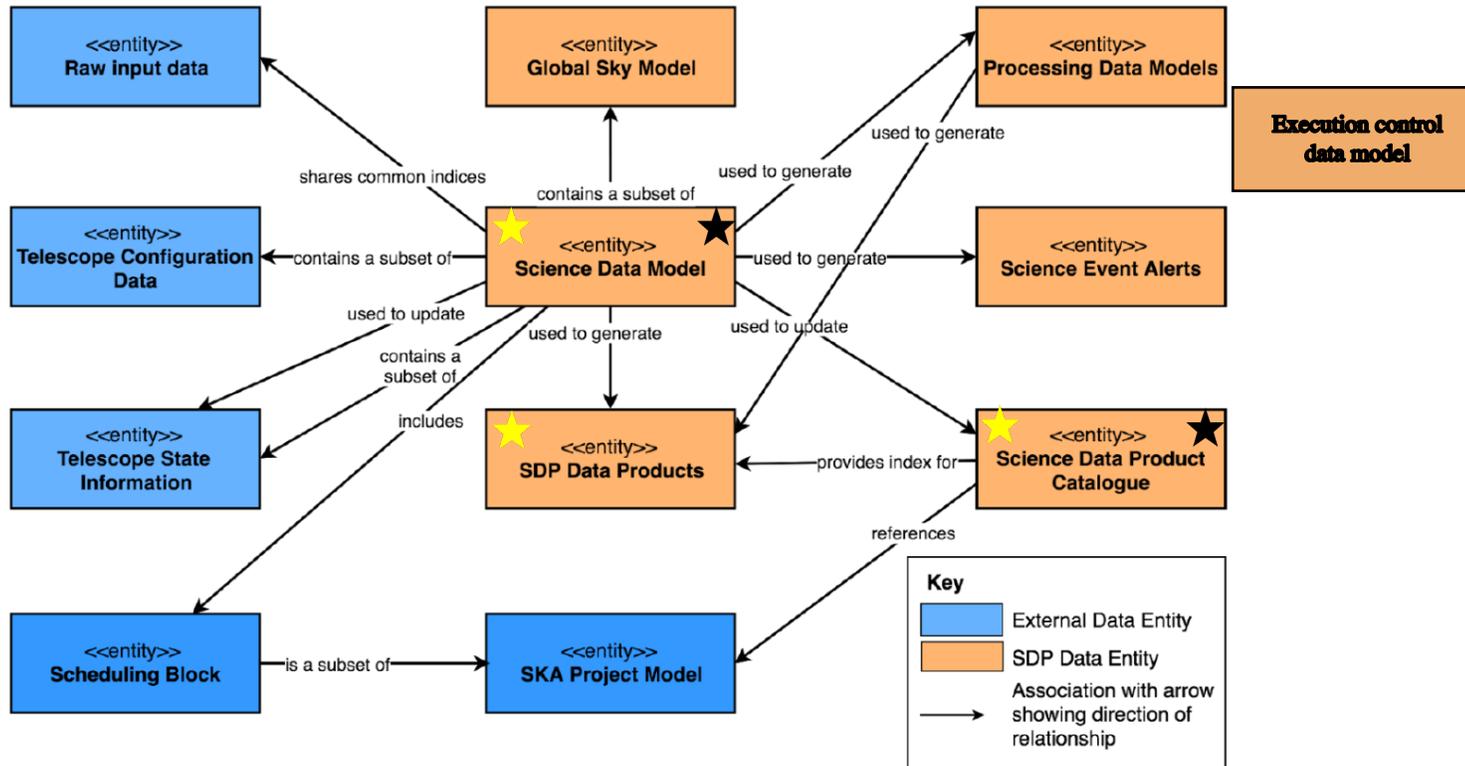


Credits: SKA Organisation



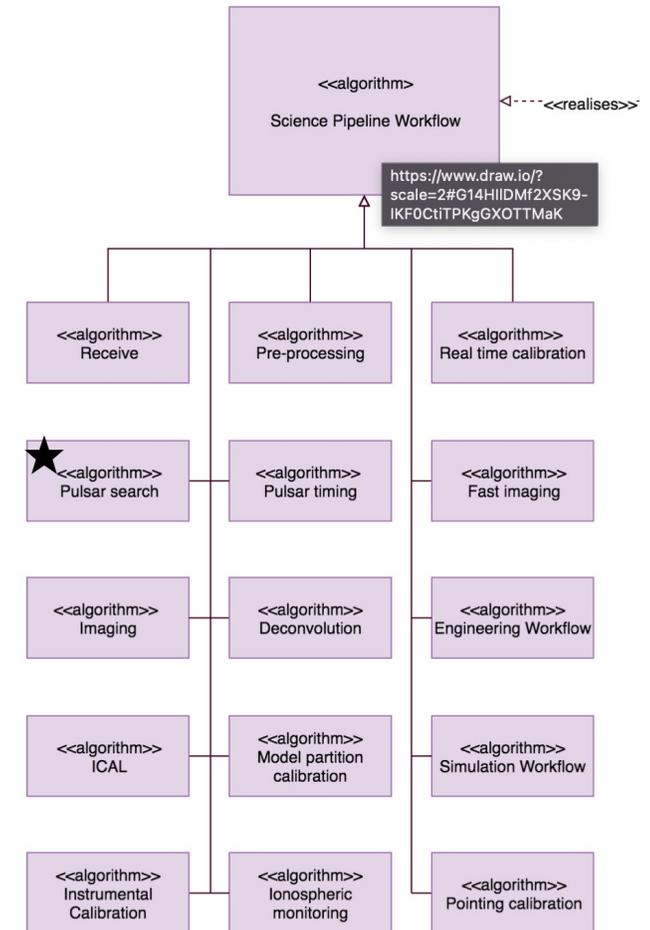
SKA SDP METADATA ANALYSIS

→ SDP Conceptual Data model



Credit: SKA SDP CDR Documents

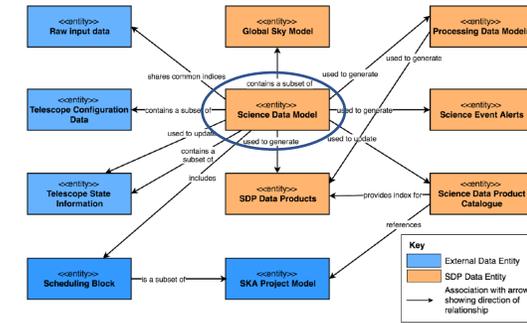
→ SDP Science Pipelines



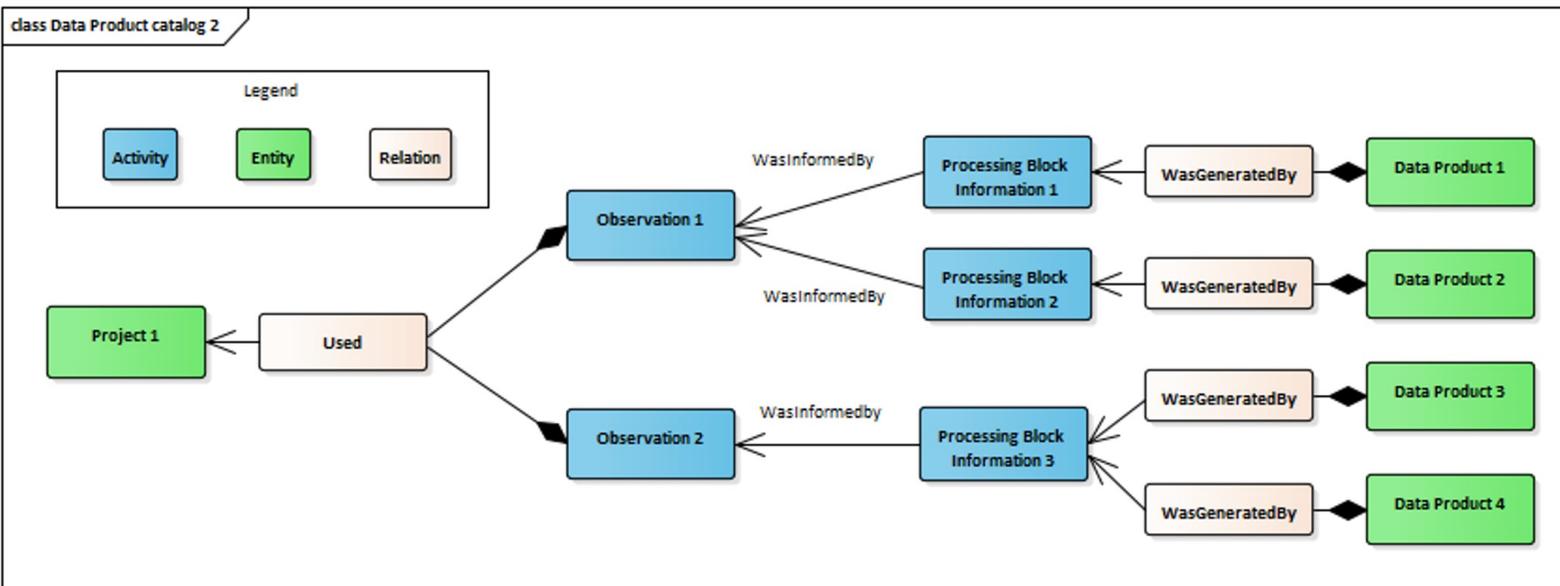
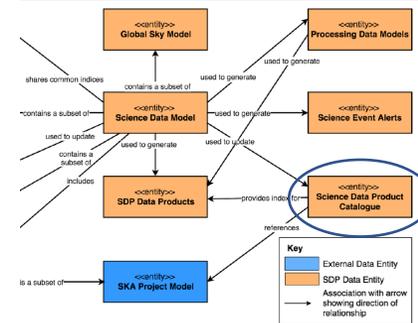
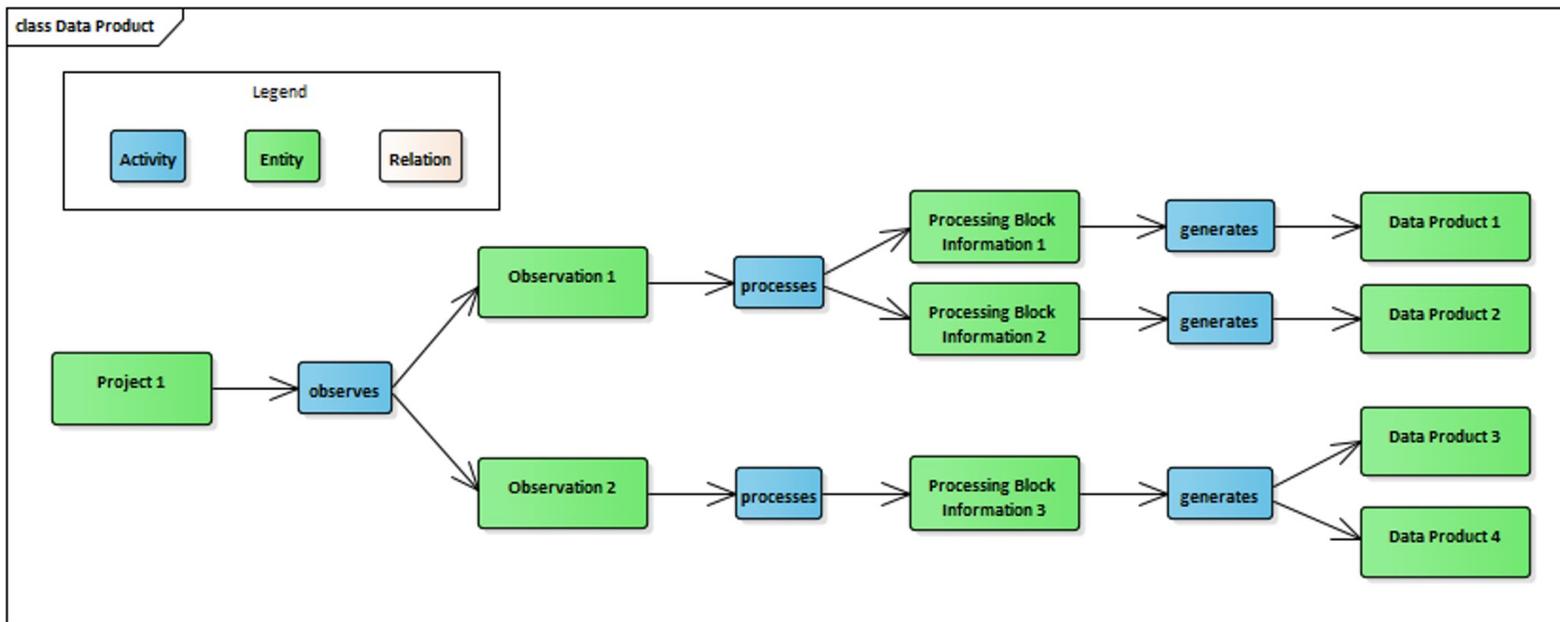
SKA SDP METADATA ANALYSIS

→ Science Data Model

- Local Sky Model (LSM)
 - Subset of the Telescope Configuration Data
 - Subset of the Telescope State Information
 - Gain Tables (Calibration Solutions)
 - SDP Quality Assessment metrics (not yet defined)
 - Processing Block Configuration ---> e.g. it includes reference to the Pipeline/Workflow Script
 - Processing Logs
 - Other Data Product Catalogue metadata
-
- Each instance of the Science Data model is stored as a data product when the processing finishes.
 - At least one instance of the SDM is associated with each Scheduling Block.
 - The SDM does not include the Raw Input Data, Processing Data Models or SDP Data Products, but does include references to these data entities.



SDP DATA PRODUCT CATALOGUE



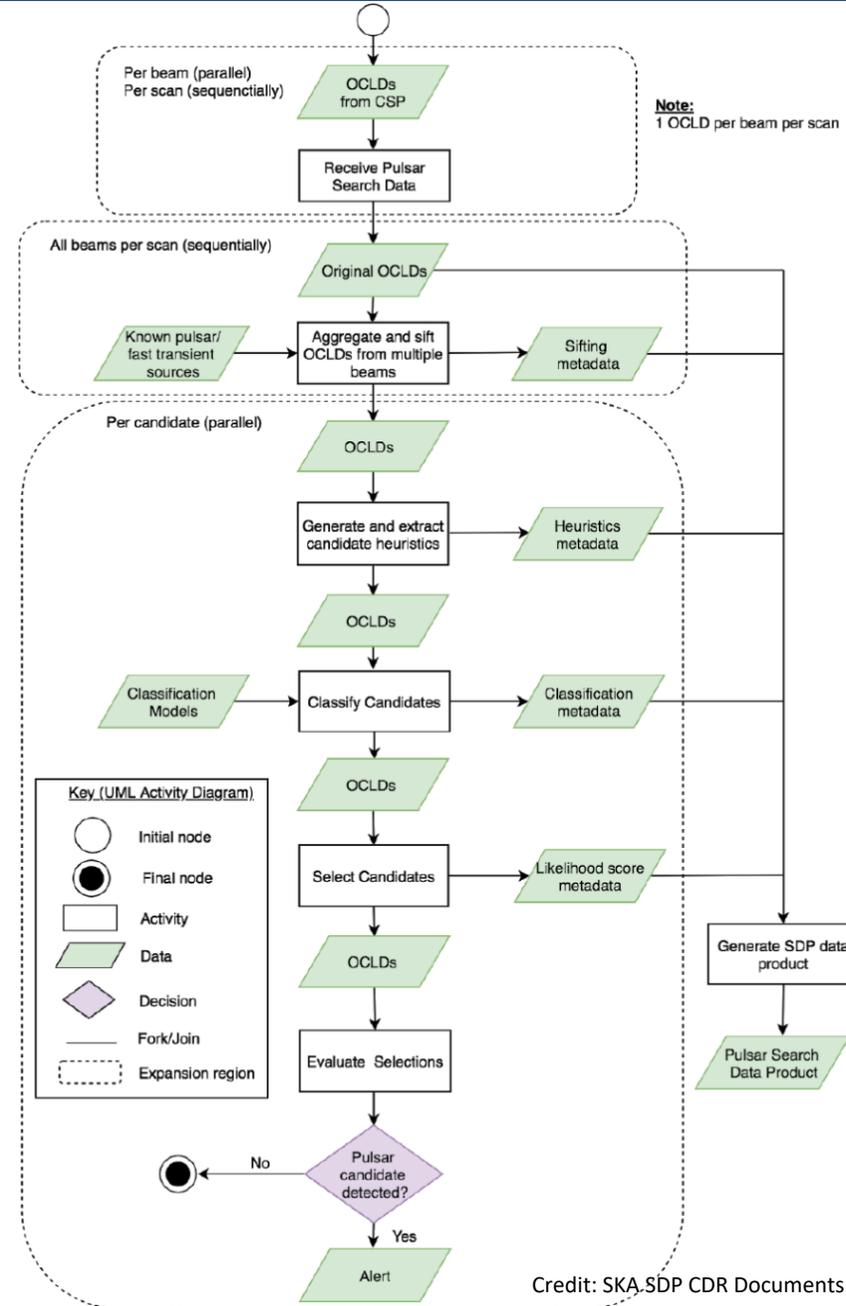
Data Product catalogue:

- Project
- Observation (Access Rights, Obs ID,...)
- Processing block Information (LSM, Telesc. State, target,...)
- Data products



SDP Science Pipelines: Pulsar Search workflow

OCLD: Optimised Candidate List and Data



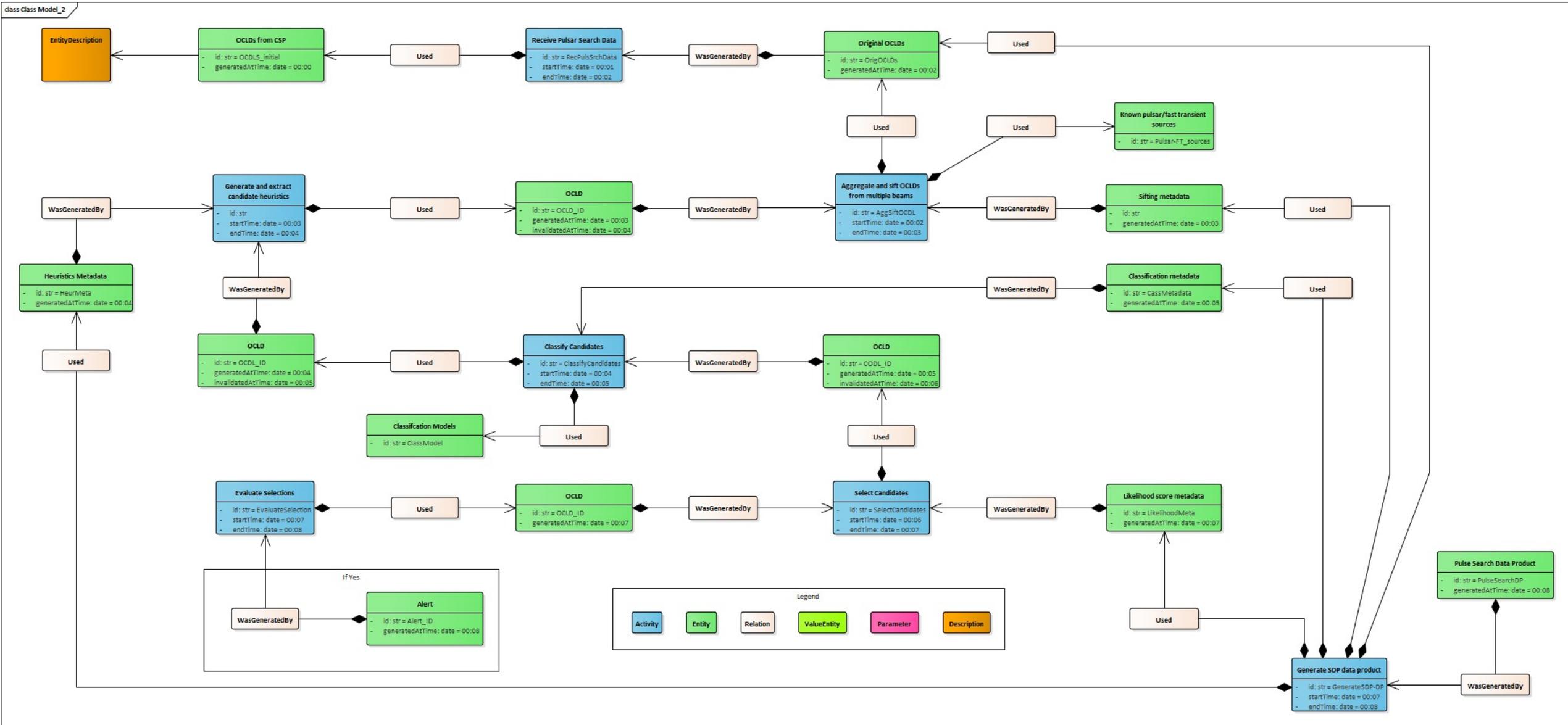
Credit: SKA.SDP CDR Documents



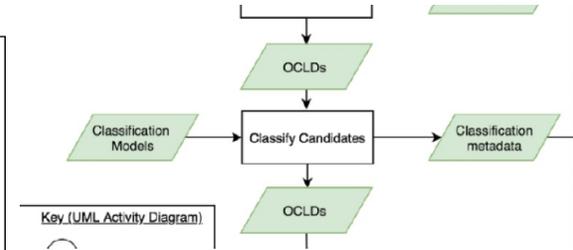
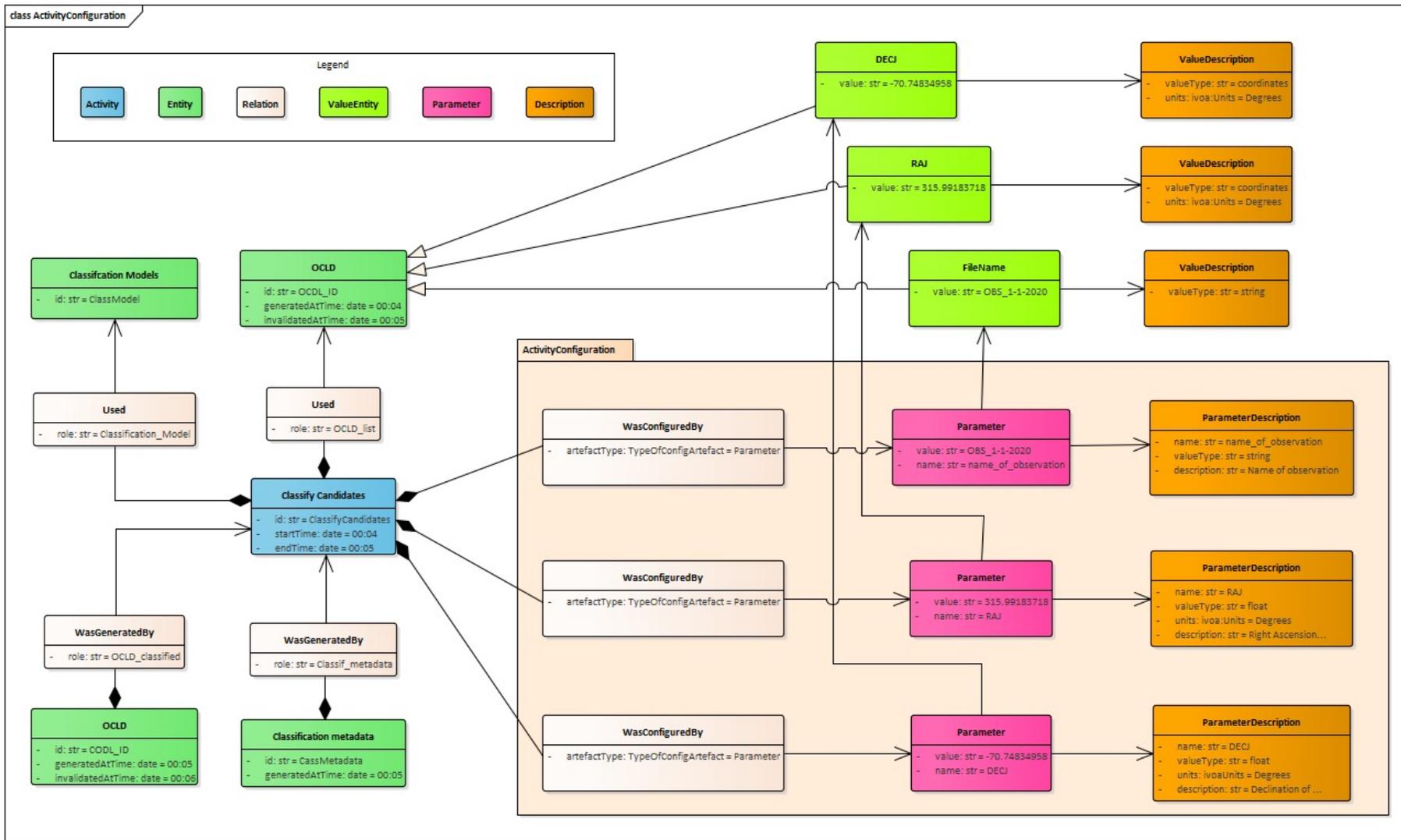
IAA-CSIC



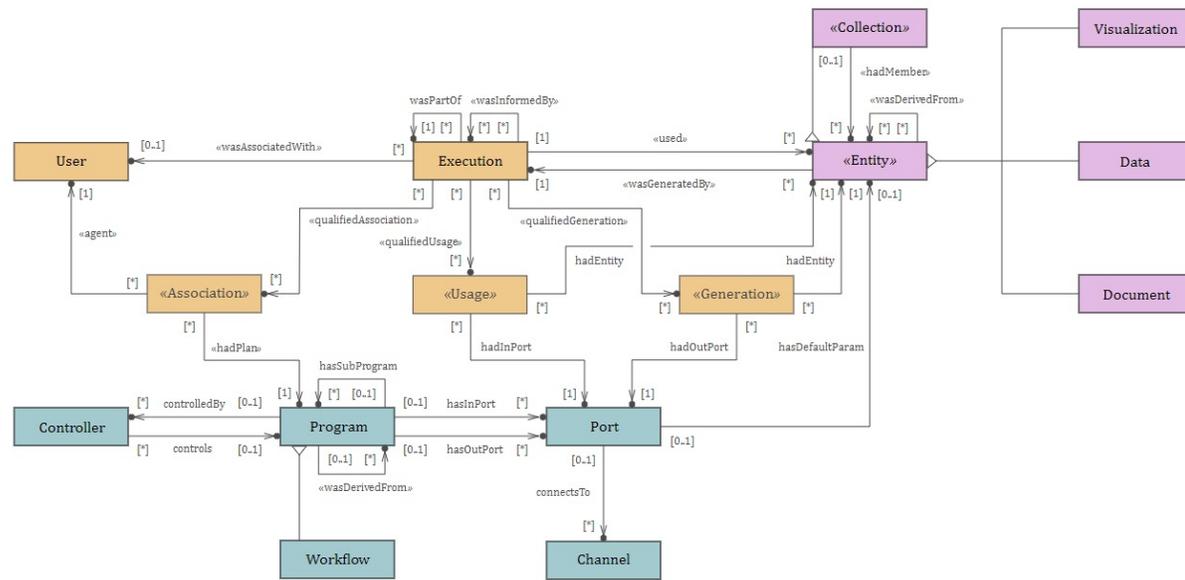
Pulsar Search workflow and IVOA Provenance Data Model



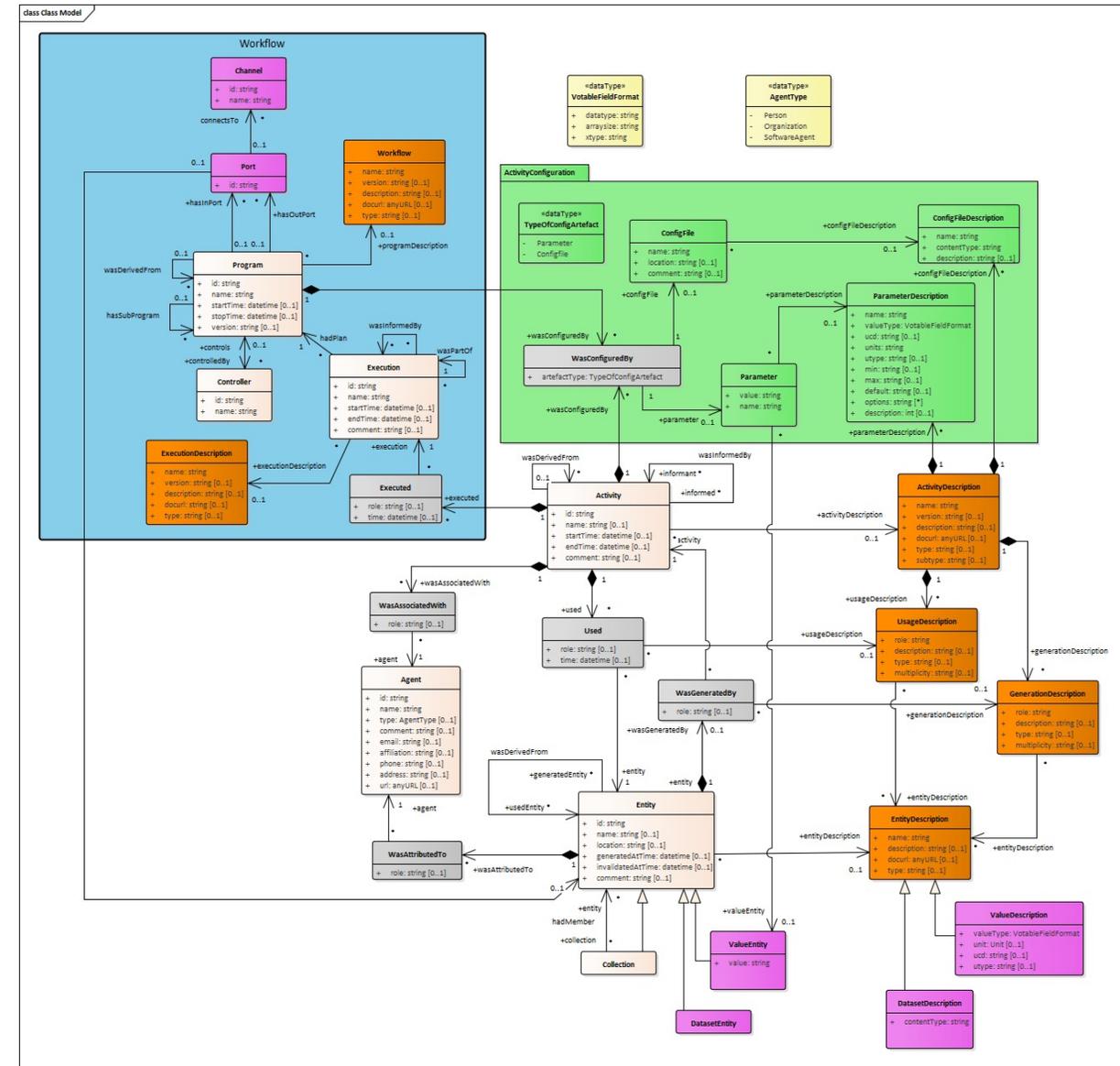
Pulsar Search workflow and IVOA Provenance Data Model



ProvONE and IVOA Prov

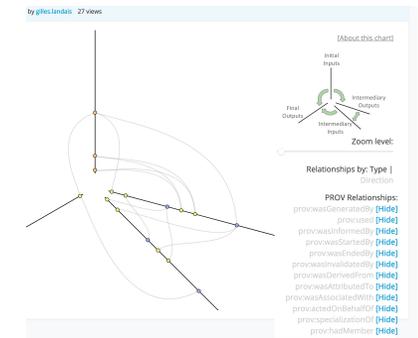
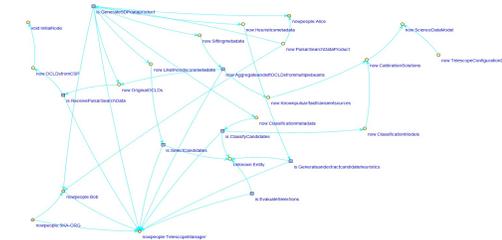
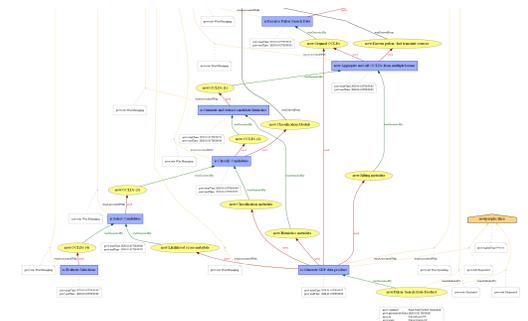


Visualization
Data
Document



PROVENANCE VISUALIZATION

- Interoperability with existing tools by using standard formats: PROV-O, PROV-N, PROV-JSON or PROV-XML
- Open Provenance
- VOProv (based on the Open Provenance library)
 - Static provenance graph
- Prov Viewer
 - Generic
- Active Provenance
 - Powerfull UI but restricted to S-Prov DM
- Prov Store
 - Web service that allows you to store, browse and manage your provenance documents



Thank you!

Julián Garrido: jgarrido@iaa.es

