NRC·CNRC

# Common Archive Observation Model: Inside the Data Centre

**Patrick Dowler**

**Canadian Astronomy Data Centre**

**on behalf of CADC developers and scientists + external contributors from EAO/JCMT and MAST/STScI**

# Common Archive Observation Model

- What is CAOM? Why did we do this?
- CAOM content at CADC
- Services based on CAOM
- Open Source community
- no UML here!
    - try Demo 4 …

# Common Archive Observation Model

- common model to describe all science data at CADC
- design goals:
  - metadata only (no data formats)
  - static collections & data flow from telescope(s)
  - public & proprietary metadata & data

# Common Archive Observation Model

- common model to describe all science data at CADC
- design goals:
  - metadata only (no data formats)
  - static collections & data flow from telescope(s)
  - public & proprietary metadata & data
  - support data discovery services
  - support data access services
  - support for operational modes (e.g. peer-to-peer replication)

# Common Archive Observation Model

- common model to describe all science data at CADC
- design goals:
  - metadata only (no data formats)
  - static collections & data flow from telescope(s)
  - public & proprietary metadata & data
  - support data discovery services
  - support data access services
  - support for operational modes (e.g. peer-to-peer replication)
  - **describe the data we actually have**
  - **evolve to meet new data challenges (currently: CAOM-2.2)**

# Common Archive Observation Model – Current Content

- CAOM describes all science data products at CADC

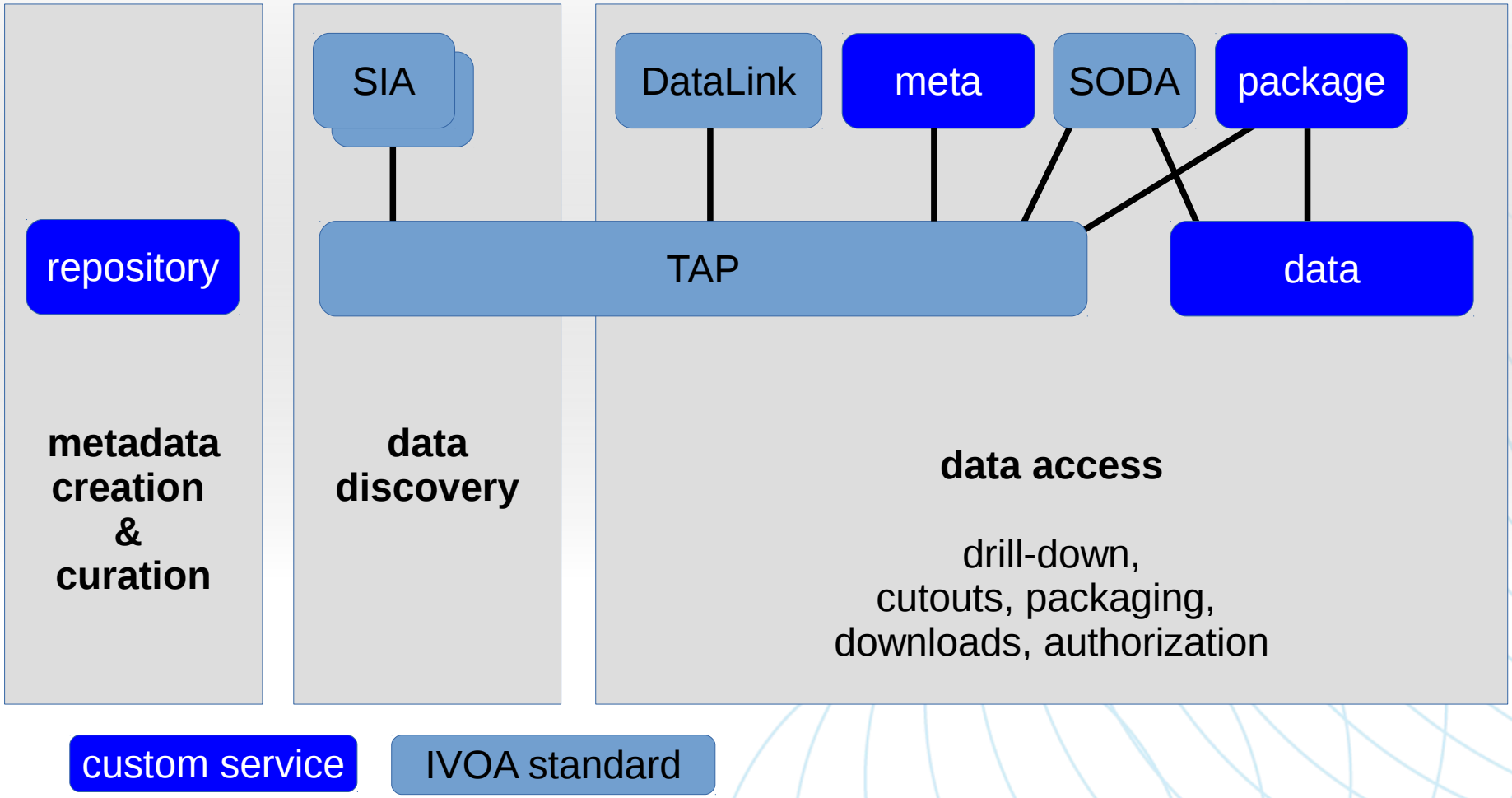
| collection | # products | | collection | # products |
|---|---|---|---|---|
| APASS | 311858 | | GEMINI | 1969640 |
| BLAST | 134 | | HST | 2706843 |
| CFHT | 2663343 | | HSTHLA | 80848 |
| CFHTMEGAPIPE | 7792 | | IRIS | 1720 |
| CFHTTERAPIX | 1048 | | JCMT | 1422092 |
| CFHTWIRWOLF | 1697 | | MACHO | 237116 |
| CGPS | 609 | | MOST | 3262 |
| DAO | 328606 | | OMM | 88162 |
| DAOPLATES | 5017 | | UKIRT | 1288458 |
| FUSE | 5887 | | VGPS | 39 |
| **Collections** | **20** | | **Products** | **→ 12 million** |

# Common Archive Observation Model – Current Content

- CAOM describes a variety of science data product types at CADC

| product type | # products | |
|---|---|---|
| catalog | 159349 | ~160k |
| cube | 128806 | ~130k |
| image | 7064469 | ~7m |
| spectrum | 1747542 | ~1.8m |
| time series | 3262 | ~3k |
| (unclassified) | 2020749 | ~2m |

# Common Archive Observation Model – Services



# Common Archive Observation Model

- 20 collections → CAOM
- pros:
  - much less software to write & maintain
  - software improvements have broad benefits
- same:
  - each collection (telescope / instrument) requires it's own data engineering effort
  - tension: where do I put this telescope metadata **vs.** where do I get this CAOM metadata
- cons:
  - all collections must evolve (more or less) together

# Common Archive Observation Model

- open-source model: http://www.opencadc.org/caom2
- open-source code: https://github.com/opencadc
- discussions, contributions, & collaborations are welcome
- evolution of CAOM is assumed