Science and Technology Facilities Council

uk | SRC

SKAO Regional Centre United Kingdom

Interoperability in Astronomical Science Metadata Model Mapping

Michael Johnson & Erin Brassfield Bourke

June 2025

Overview

e-MERLIN - Multi-Dish Radio telescope with no public facing archive

CAOM - Metadata archive model designed for single-dish optical astronomy observations

Square peg → Round hole

Does the interoperability of using a standardised model compensate for the complexity added by applying it outside of what it was designed for?

e-MERLIN

Multi-Element Radio Linked Interferometer Network (MERLIN)

Science includes:

- Evolution of the Universe
- Physics of extreme conditions
- Stellar, Galaxy and Planetary Evolution

2009-present, e-MERLIN

1990-2009, MERLIN

1980-1990, MTLRI

Image credit:
https://www.e-merlin.ac.uk/

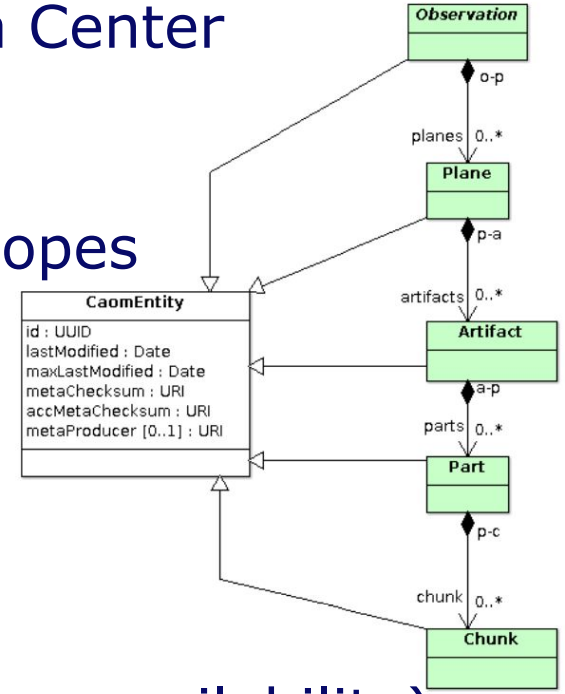
uk | SRC

Common Archive Observation Model (CAOM)

Designed by Canadian Astronomical Data Center

Focus on Findability and Accessibility

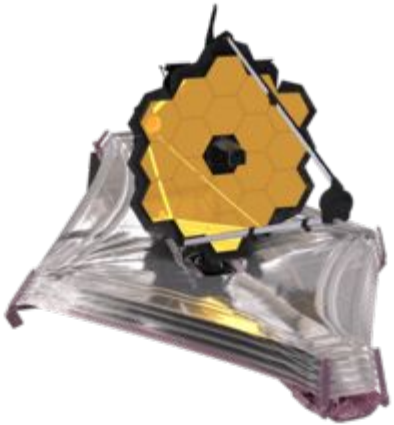
Applied to all of the CADC's optical telescopes



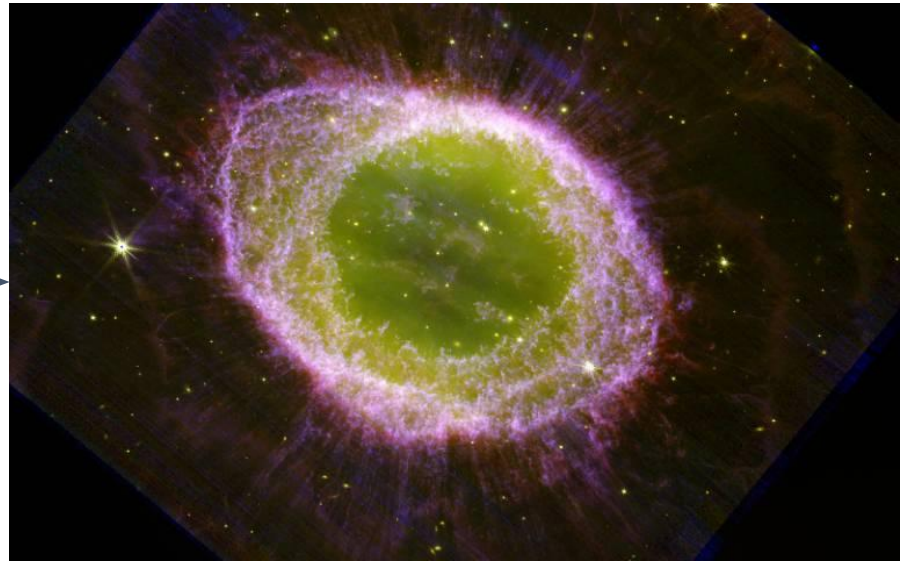
Contains metadata on:
- telescope, target, proposal
- time, energy, provenance
- data product metadata (size, hash value, availability)

uk | SRC

Optical Astronomical Data



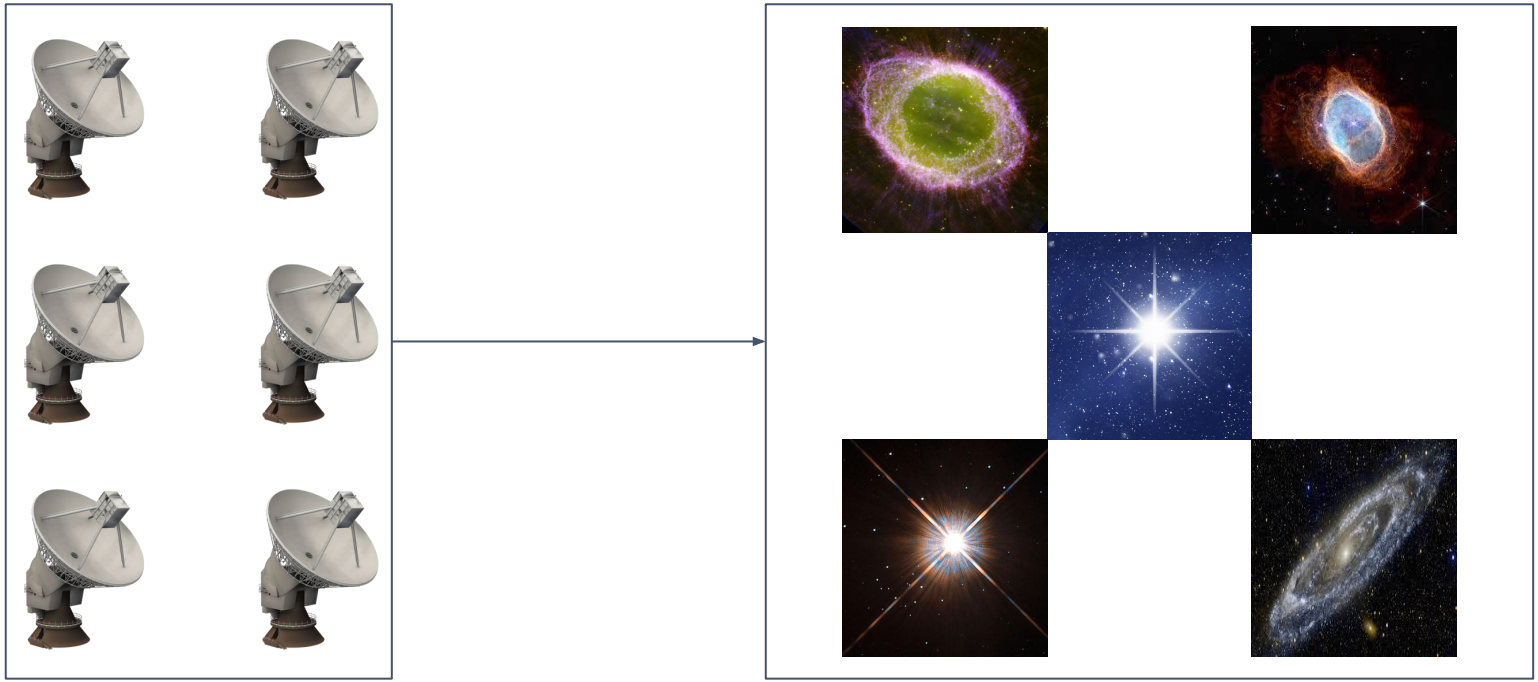
James Webb Space Telescope. Image credit: https://www.jwst.nasa.gov/content/webbLaunch/whereIsWebb.html



Ring Nebula. Image credit: https://www.ucl.ac.uk/news/2023/aug/james-webb-space-telescope-captures-stunning-images-ring-nebula

uk | SRC

Radio Astronomical Data

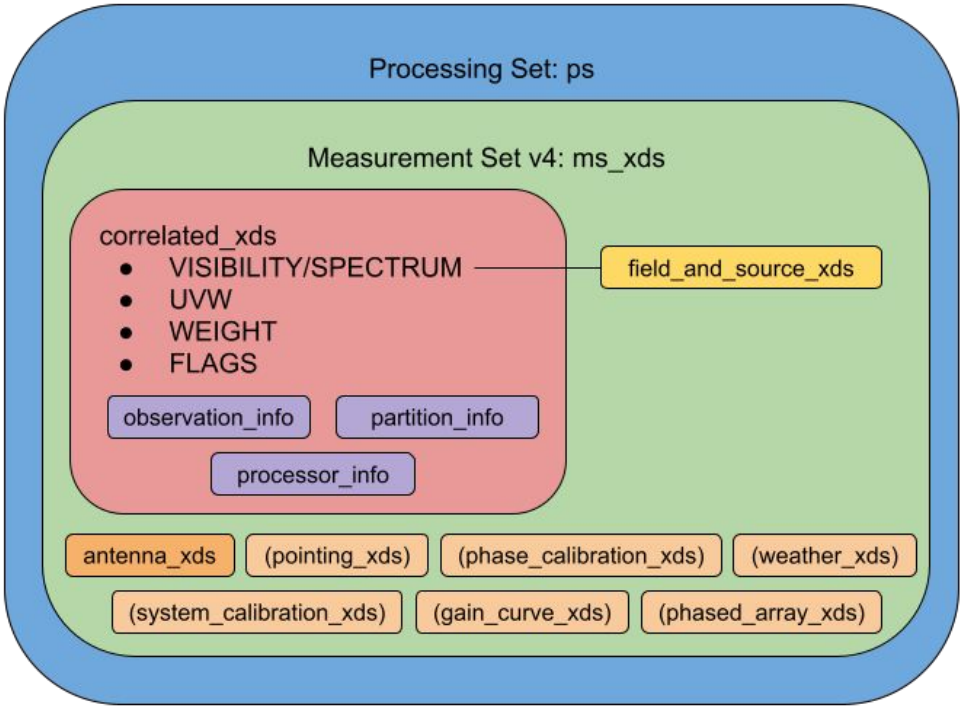


Representation of Radio Data

Measurement Sets

- Hierarchical tabular data
- Coupled data and metadata
- Similar to HDF5

Data format is specific to multi-antenna radio astronomy ONLY



uk | SRC

Incompatibilities

- The multiplicity for telescope and targets in CAOM is 1
- Radio metadata ≠optical metadata
- No way to model some radio metadata
- Extraneous optical specific information in model

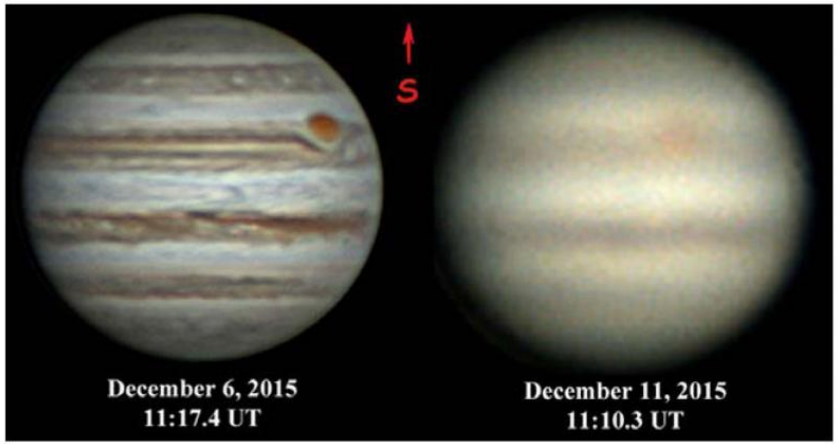
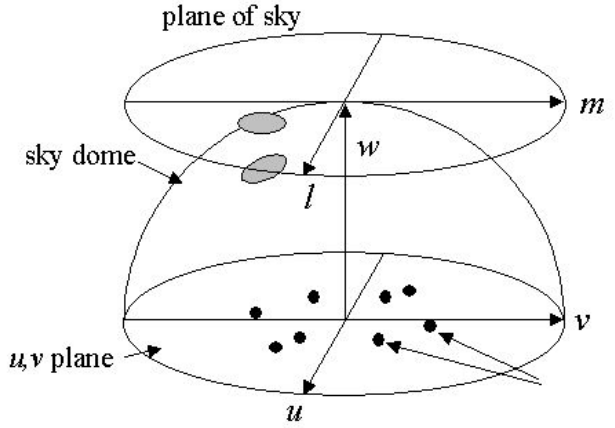


Figure 2. Jupiter, captured with the same equipment in good (left) and bad (right) seeing conditions. *Cliff Ashcraft, New Jersey, USA*



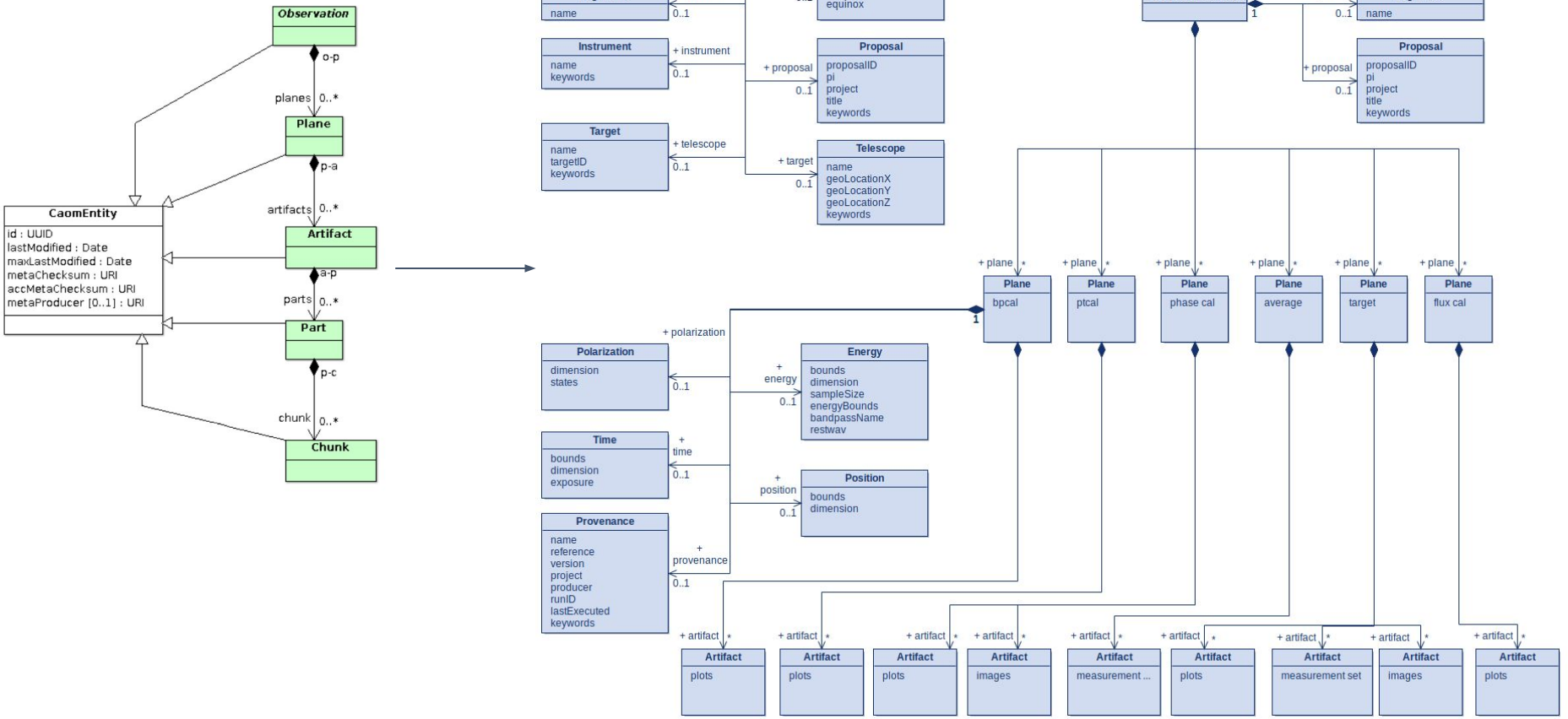
uk | SRC

Solutions

Adapting the model itself, new radio-specific additions

Bending the model

- Extra observations per telescope/target
- Adding information as keywords
- Leaving many fields unused


Observation
o-p
planes
0..*
Plane
p-a
artifacts
0..*
Artifact
a-p
parts
0..*
Part
p-c
chunk
0..*
Chunk
CaomEntity
id : UUID
lastModified : Date
maxLastModified : Date
metaChecksum : URI
accMetaChecksum : URI
metaProducer [0..1] : URI
SimpleObservation
+Target X
SimpleObservation
+Band Pass Calibrator
SimpleObservation
+Pointing Calibrator
SimpleObservation
+Flux Calibrator
SimpleObservation
+Phase Calibrator X
SimpleObservation
Lovell
SimpleObservation
Mark II
SimpleObservation
Pickmere
SimpleObservation
Darnhall
SimpleObservation
Knockin
SimpleObservation
Defford
SimpleObservation
Cambridge
0..1
1
TargetPosition
coordinates
coordsys
equinox
+ targetPosition
Algorithm
name
+ algorithm
Instrument
name
keywords
+ instrument
Proposal
proposalID
pi
project
title
keywords
+ proposal
Target
name
targetID
keywords
+ telescope
Telescope
name
geoLocationX
geoLocationY
geoLocationZ
keywords
+ target
+ derivedObservation
+ members
1..*
DerivedObservation
+ algorithm
Algorithm
name
+ proposal
Proposal
proposalID
pi
project
title
keywords
+ plane
Plane
bpcal
Plane
ptcal
Plane
phase cal
Plane
average
Plane
target
Plane
flux cal
+ polarization
Polarization
dimension
states
+ energy
Energy
bounds
dimension
sampleSize
energyBounds
bandpassName
restwav
+ time
Time
bounds
dimension
exposure
+ position
Position
bounds
dimension
+ provenance
Provenance
name
reference
version
project
producer
runID
lastExecuted
keywords
+ artifact
Artifact
plots
Artifact
plots
Artifact
plots
Artifact
images
Artifact
measurement ...
Artifact
plots
Artifact
measurement set
Artifact
images
Artifact
plots

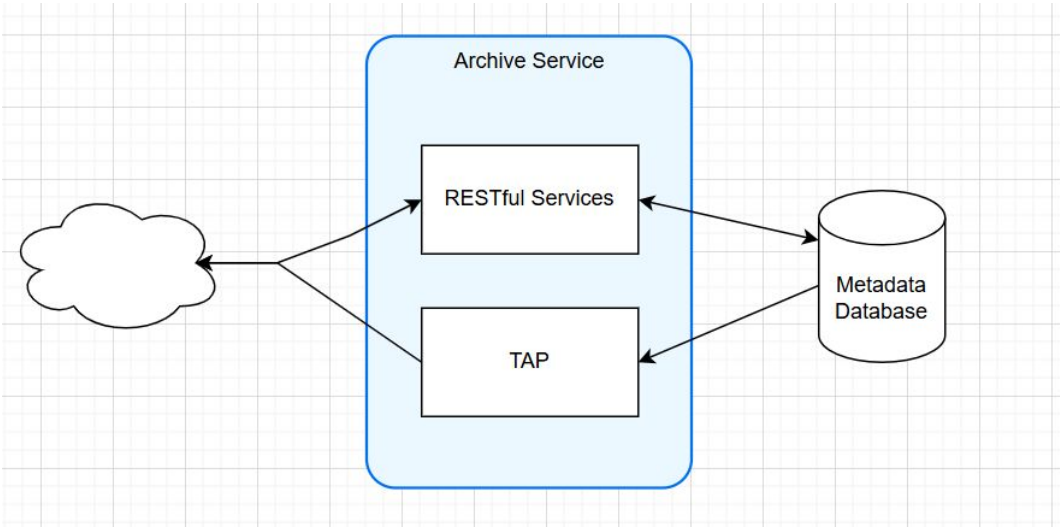
Implementation

An advantage of using and existing model is infrastructure

Query service

Database schema

User interface



Testing the Implementation - Use Cases

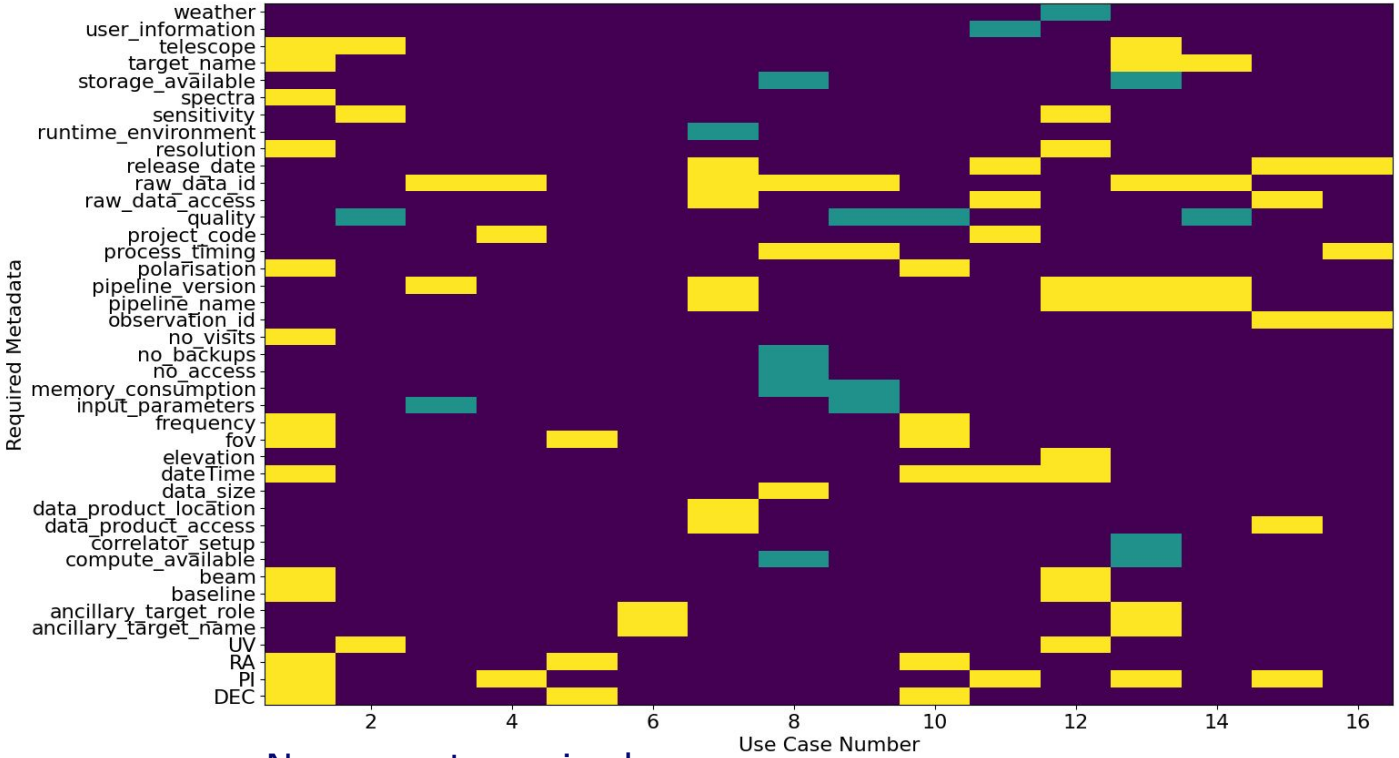
16 use cases were formulated to verify the application of the model

The primary focus being findability for the archive

For example:

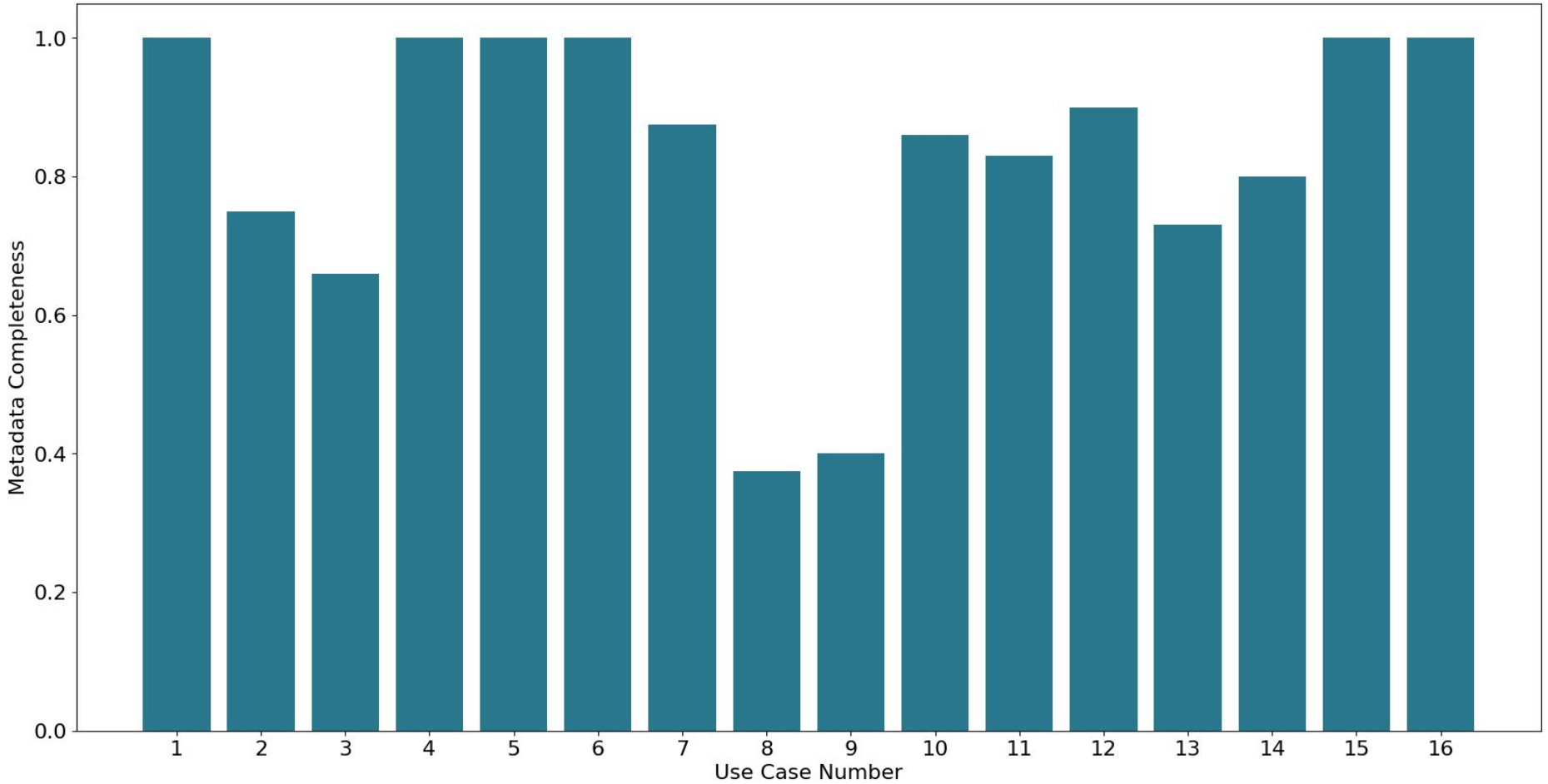
As an astronomer, I want to be able to use geometric spatial searches for my target data

Use Cases - Information loss

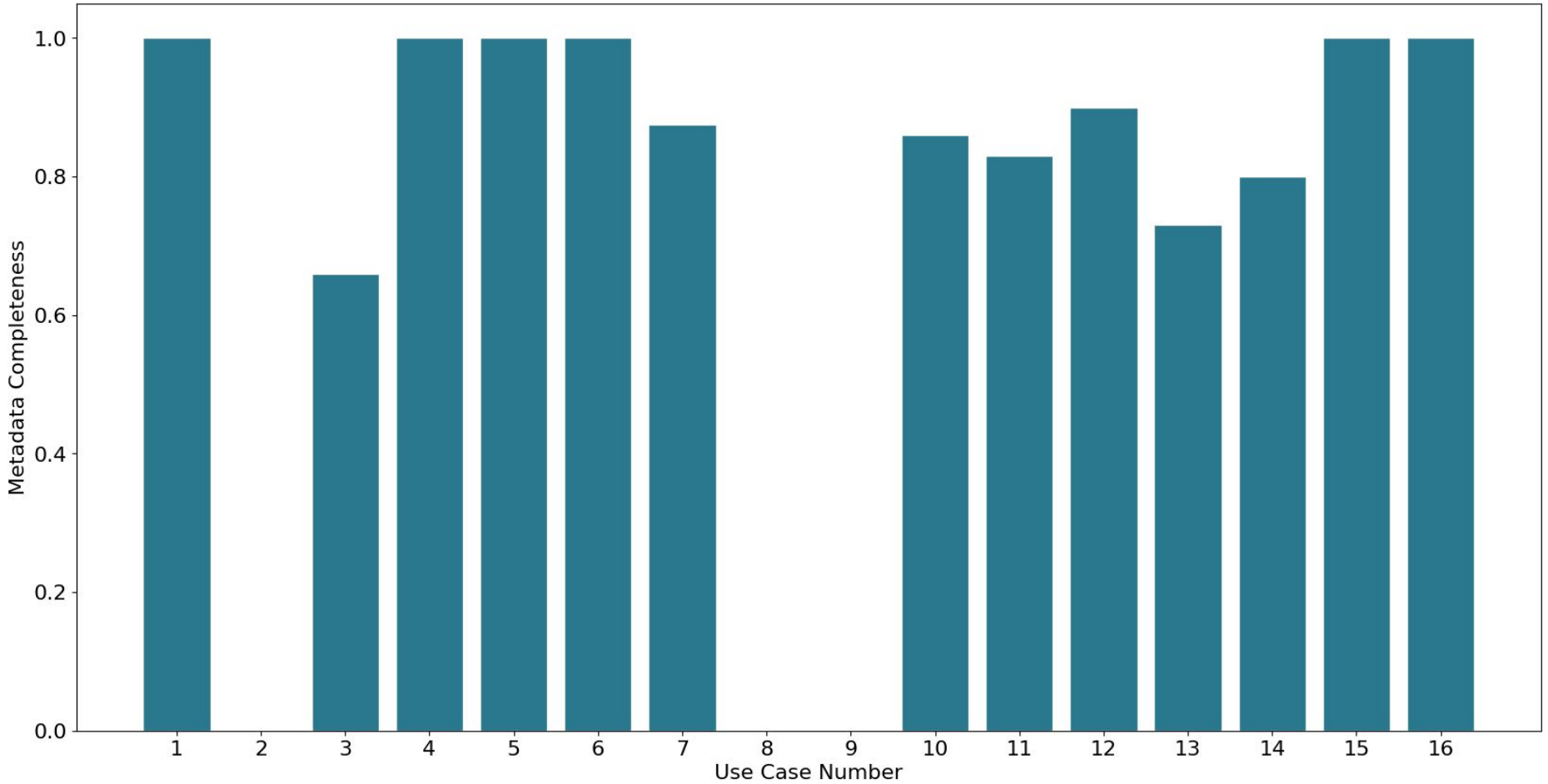


Navy - not required
Yellow - required and present in CAOM
Turquoise - required and not present in CAOM

uk | SRC


Metadata Completeness
1.0
0.8
0.6
0.4
0.2
0.0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
Use Case Number

uk | SRC


Metadata Completeness
1.0
0.8
0.6
0.4
0.2
0.0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
Use Case Number

uk | SRC

Does the interoperability of using a standardised model compensate for the complexity added by applying it outside of what it was designed for?

Pros

Interoperability

Existing infrastructure

User familiarity

Community enrichment

Cons

Query complexity

Information loss

Superfluous DB fields

DM knowledge barrier

Metadata modelling (use and creation) takeaways

- Drive the data model decision with Use Cases!
- Simplify first, iterate with added fields afterwards.
- When mapping gets complex, go back to use cases.
- Correct model architecture is more important than field inclusion.

Conclusions

What aspects does your application prioritize?

Completeness, query and storage efficiency -> build your own data model

Accessibility, and interoperability -> consider using existing data models