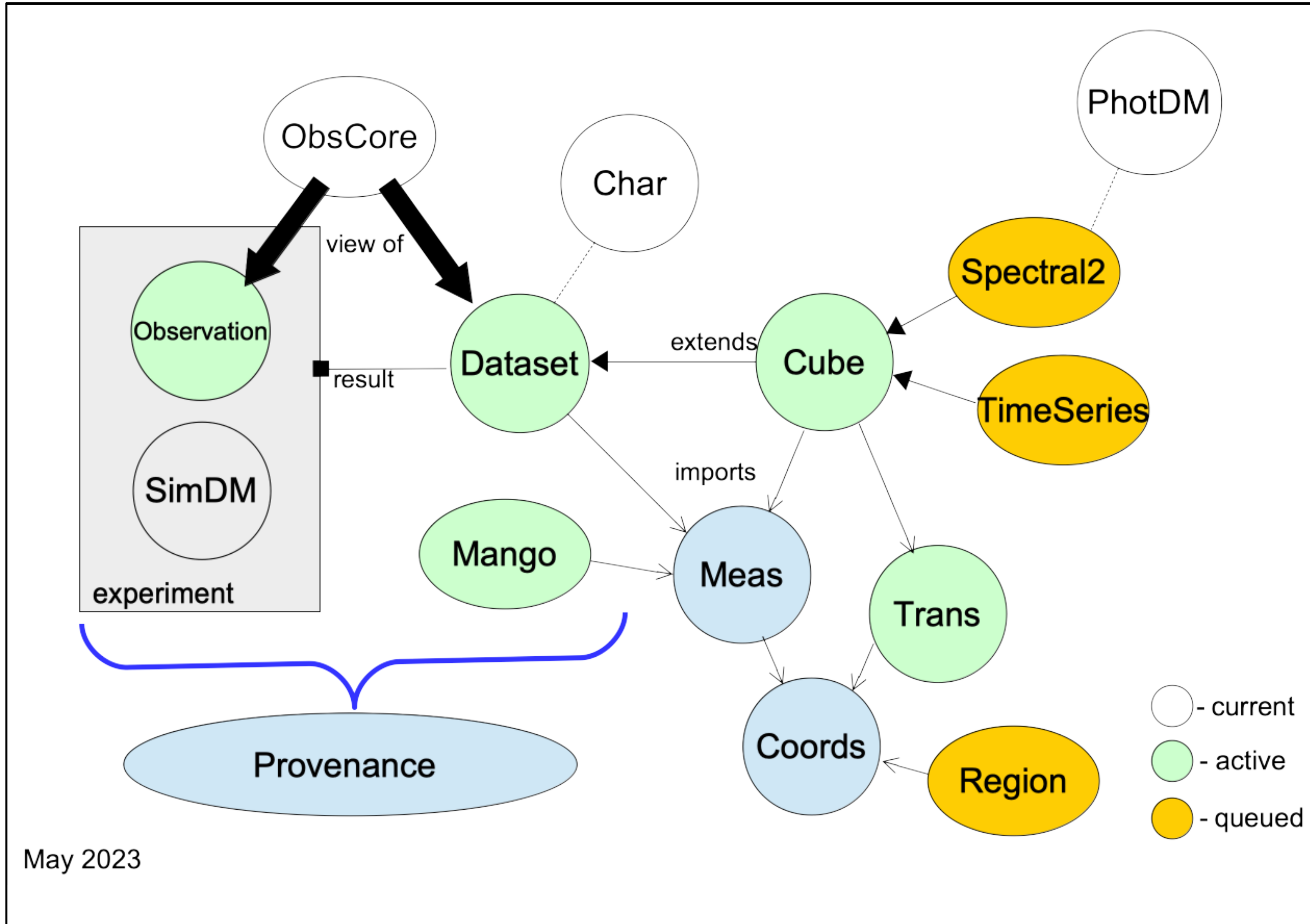


Dataset/Cube model review

High level review of models to assess compatibility with High Energy Domain data products.

Mark Cresitello-Dittmar: 2023-09-14

IVOA Data Model Family



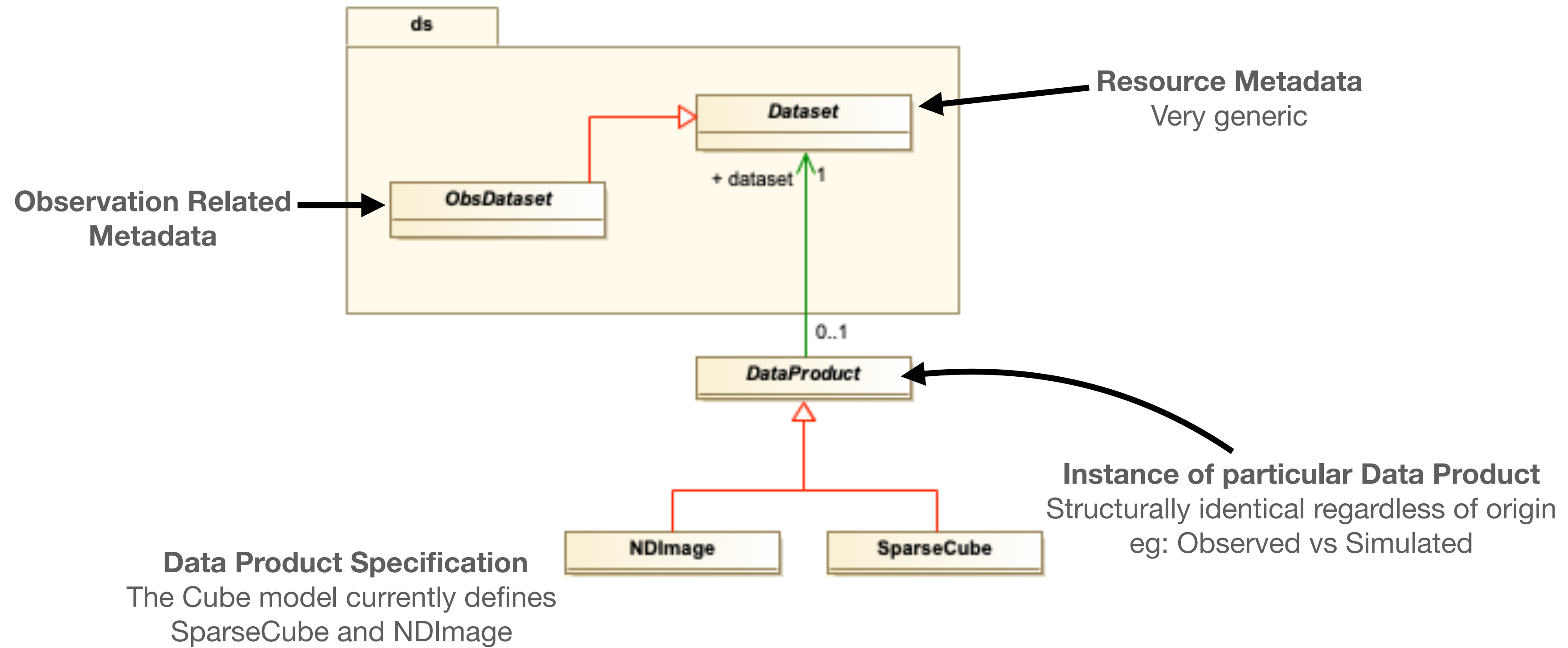
Glossary

- **Dataset:** “This metadata identifies the dataset, and provides information regarding the ownership, rights and associations with other datasets. The primary purpose of this metadata is to facilitate the registry and discovery of datasets within the IVOA community.”
- **Also has this somewhat contradictory statement:** “Since serialization format choices may effect the number of files or components which comprise a dataset, we define an IVOA Dataset as "a file or files which are considered to be a single deliverable". Examples of viable datasets include:
 - An individual data product, such as a Spectrum, or Image.
 - A 'tar' file or directory of processed observational data files.”

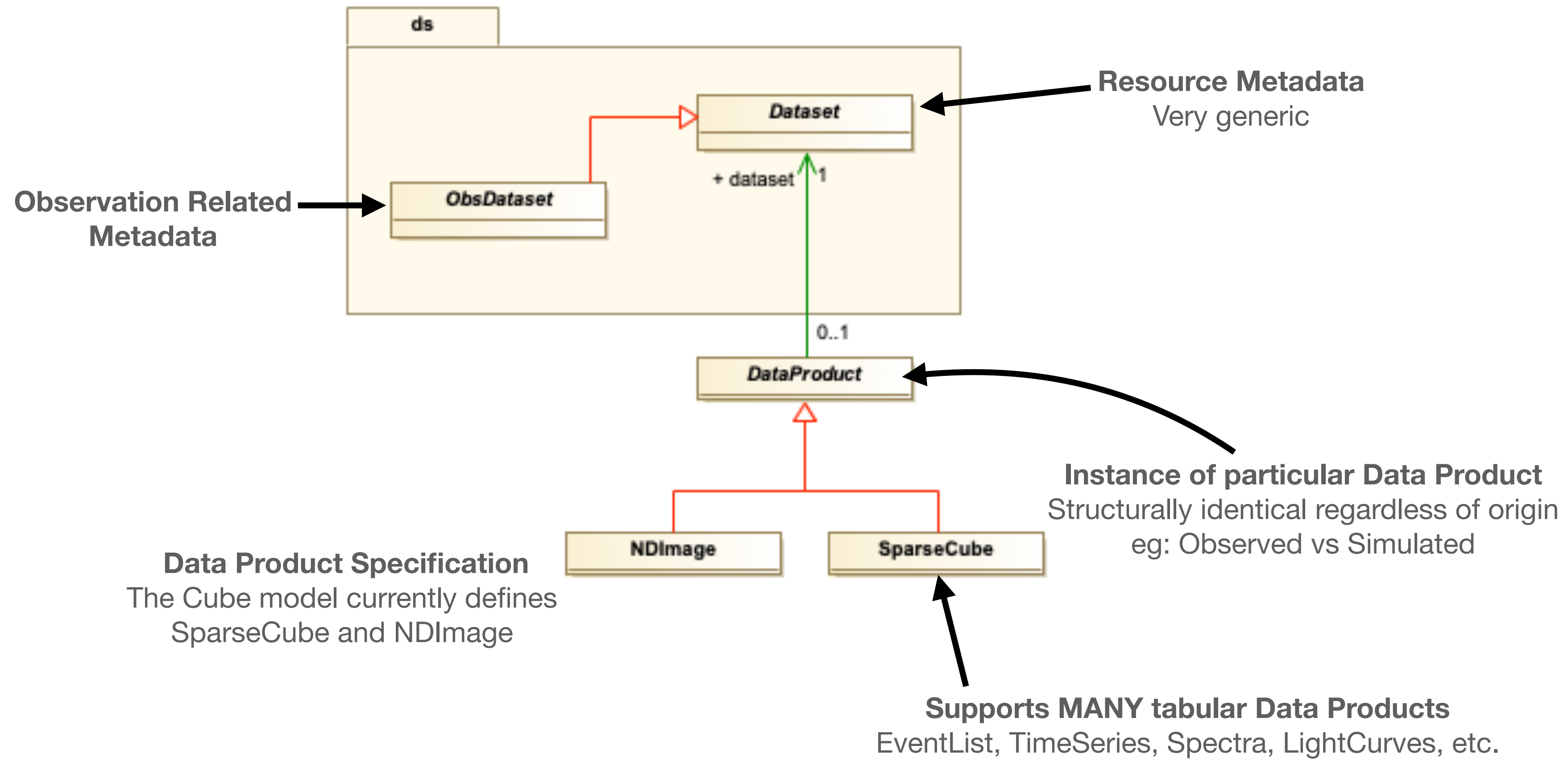
Glossary

- **ObsDataset:** Extends Dataset with additional high-level metadata for Datasets derived from an Observational process. This includes the observation identifier, instrument configuration, facility, etc.
- **DataProduct:** Currently defined in the Cube model; Data Products represent a single instance of a particular type of data. It contains only the metadata required to define the data product itself, all other metadata is left to the Dataset. The Data product definition is constant, regardless of what entity or process is generating it.

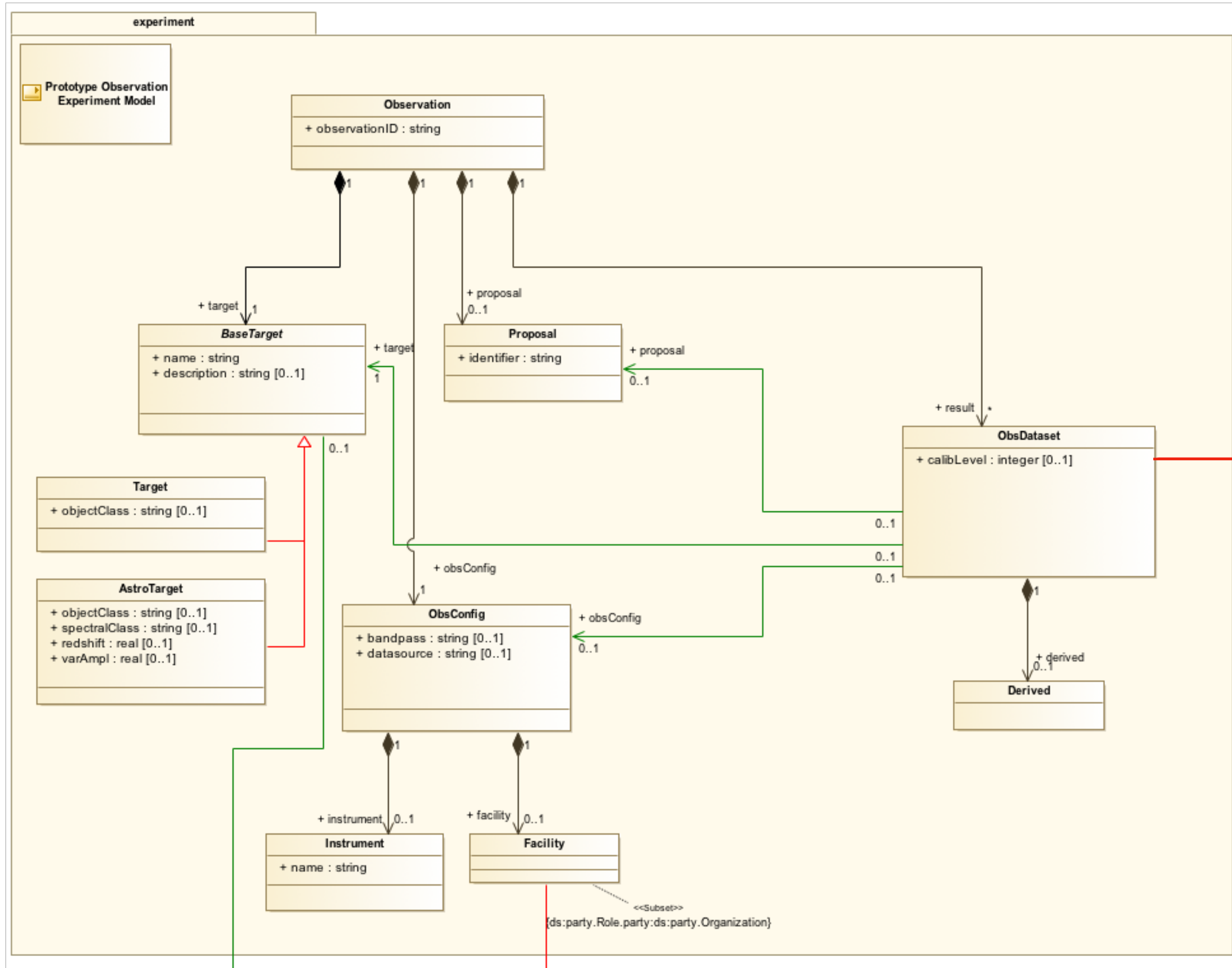
Dataset - Data Product Relation



Dataset - Data Product Relation

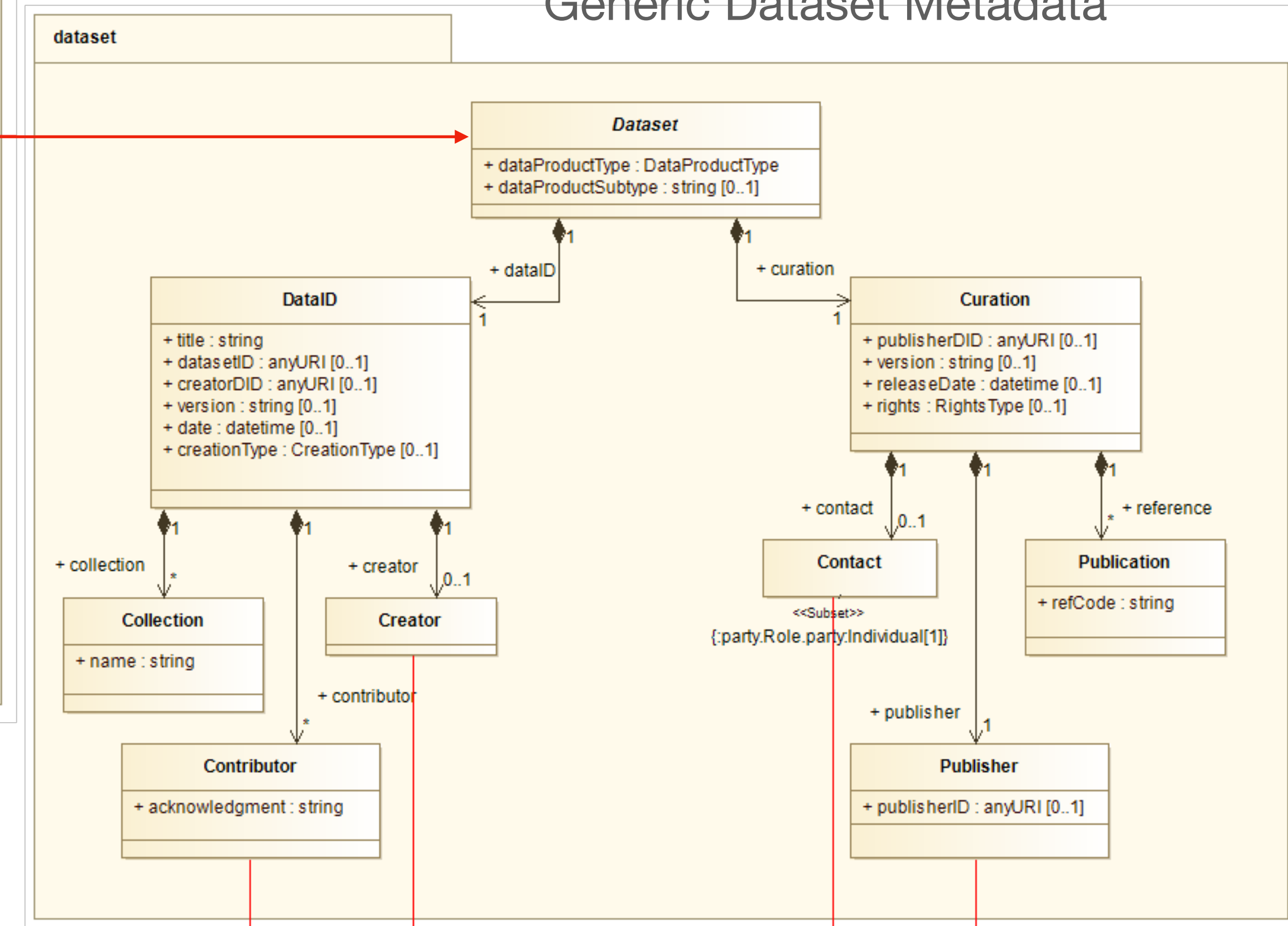


Dataset Model Overview



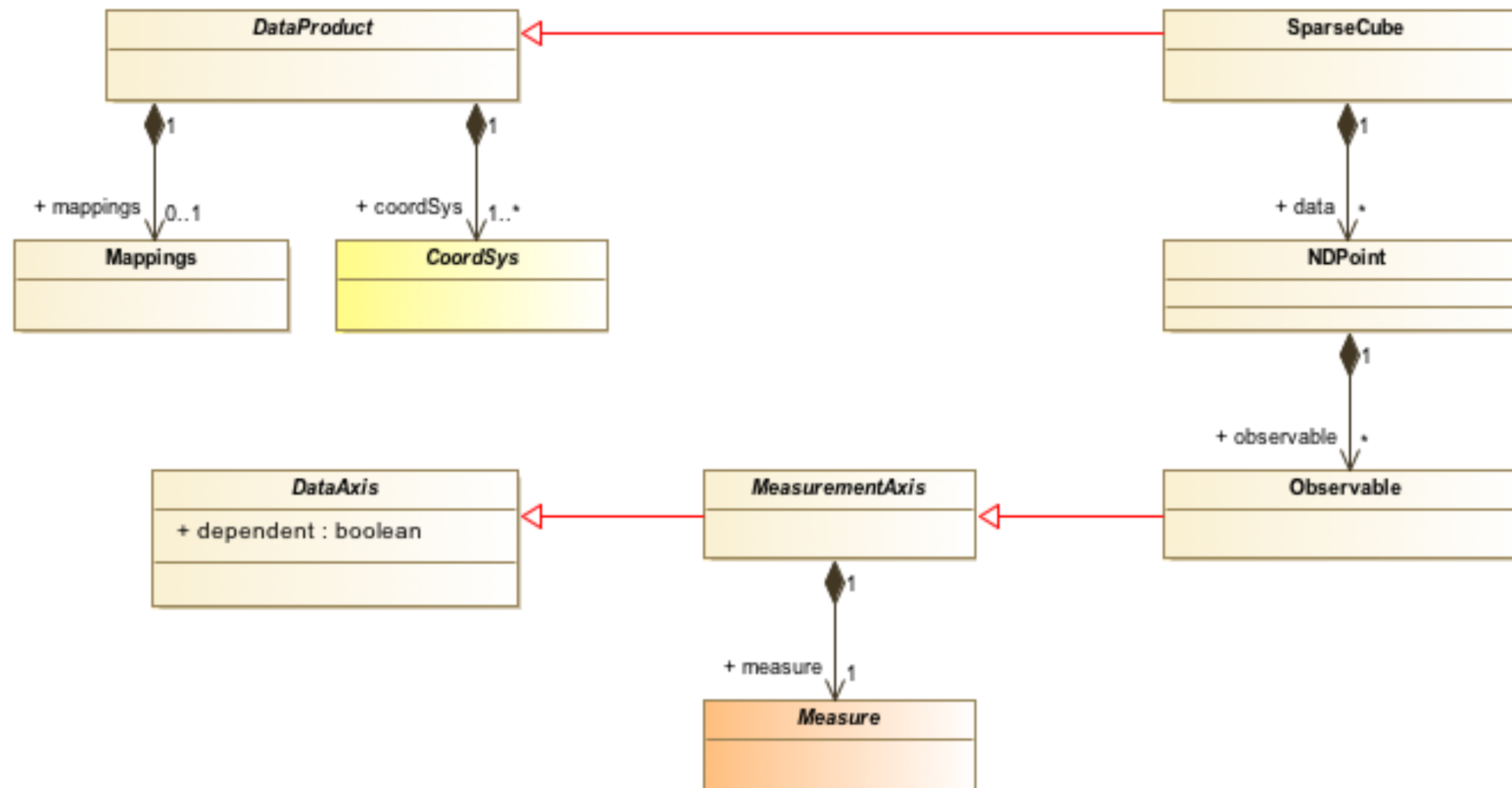
Metadata describing the Observation

Generic Dataset Metadata



Cube Model Overview

SparseCube



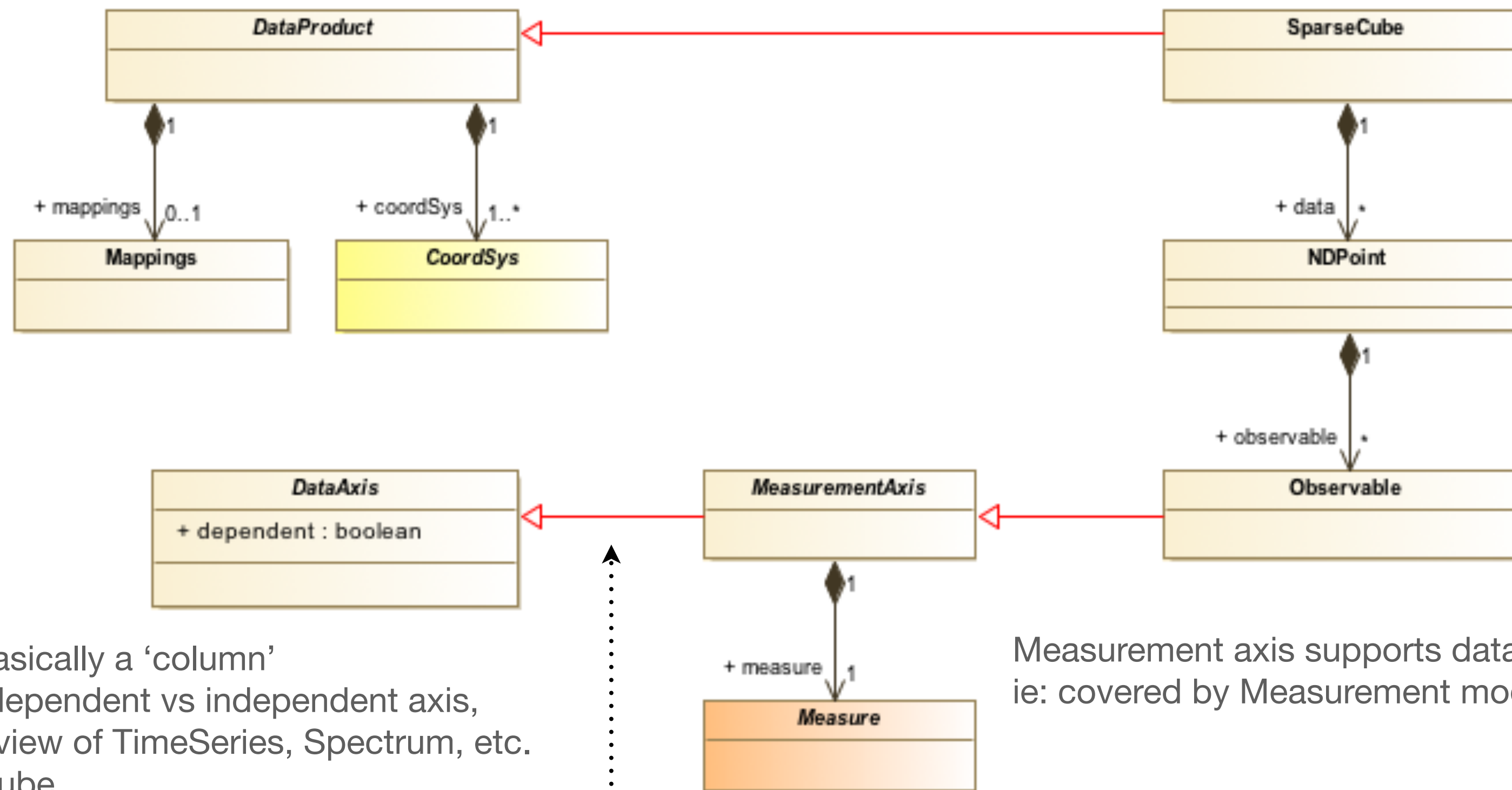
Sparse Cube Data Product

has a collection of "Points"

each Point is a collection of 'Observables'

Cube Model Overview

SparseCube



Base Object for basically a 'column'

* flag identifies dependent vs independent axis, facilitating the view of TimeSeries, Spectrum, etc. as slices of a cube

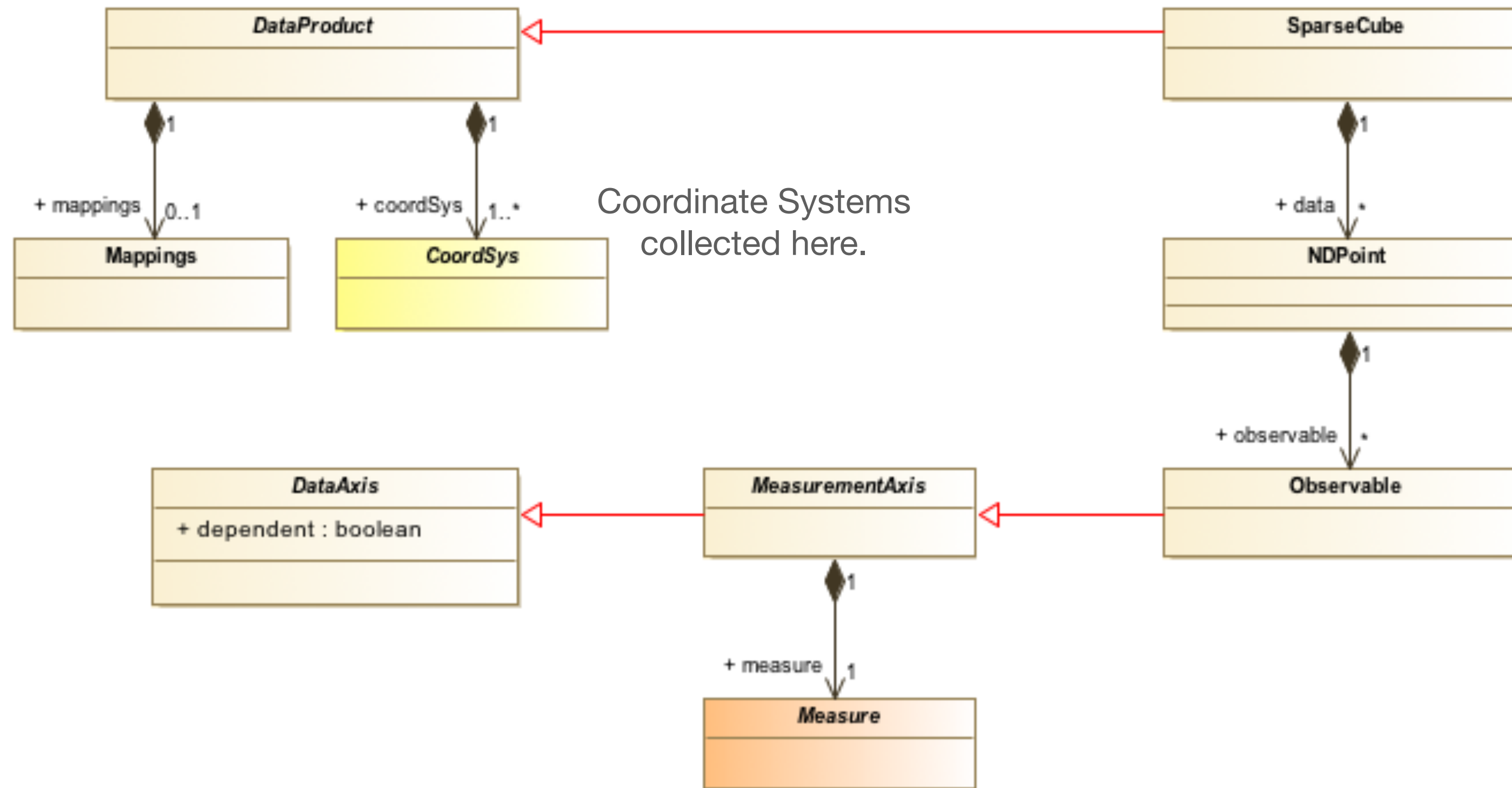
Measurement axis supports data which are physical quantities. ie: covered by Measurement model.

The separation allows for the definition of other types of DataAxis, such as quality flags and complex products such as TimeSeries of Spectra.

Cube Model Overview

SparseCube

Supports Virtual Data
(eg: WCS)



Event List as SparseCube

- Chandra Event List descriptions: https://cxc.cfa.harvard.edu/ciao/data_products_guide/events.html
 - **EVENTS**: primary data hdu
 - **GTI#** - Good Time Intervals: secondary hdu-s
 - One per CCD
 - Associated data making up part of the 'data subspace'
 - **REGION**: secondary hdu, or a simple region string.
 - Associated data making up part of the 'data subspace'
- Gratings - Adds several positions and metadata (GRATINGS=HETG) which may need to be modeled for ObsCore extension
- Data subspace
 - Combination of any filters applied to the event list data.
 - May include columns which no longer exist in the file.

Event List as SparseCube

- Chandra Event List column summary:
 - Temporal: time
 - Position: (chipx,chipy), (tdetx,tdety), (detx,dety), (x,y)
 - ccd_id, node_id = positional? (ccd_id,chipx,chipy) => (tdetx,tdety)
 - Energy: phas, pha, pha_ro, pi, energy
 - Other: fltgrade, grade, status <== how to model these?
 - Virtual Columns:
 - CPC: (cpcx,cpcy) = T(chipx,chipy)
 - MSC: (phi, theta) = T(detx,dety)
 - EQPOS: (ra,dec) = T(x,y)

Event List as SparseCube

- Chandra Event List mapping to SparseCube
 - Events => cube:SparseCube
 - Row/Event => cube:NDPoint
 - Column => cube:Observables (dependent=False)
 - time => meas:Time
 - (chipx,chipy), (tdetx,tdety), (detx,dety), (x,y) => meas:Position, with custom coordinate spaces
 - pha, pha_ro, pi, energy => meas:GenericMeasure
 - phas: a 3x3 array => ??
 - fltgrade, grade, status ==> ?? not currently supported well
 - Can use meas:GenericMeasure with an appropriate UCD, but this is conceptually incorrect.
 - Virtual Data ==> cube:VirtualMeasure (needs review)
 - CPC: (cpcx,cpcy) => trans:Shift + trans:Scale
 - MSC: (phi, theta) => trans:SkyProjection("TAN-P")
 - EQPOS: (ra,dec) = trans:SkyProjRotate("TAN")

Questions: Areas where HEIG can help to improve the models

- Dataset vs DSMetadata vs DataProduct: clarify definitions and relation
- Dataset with multiple DataProducts: do we have examples?
 - Despite the definition, DataProductType is in Dataset, not DataProduct.
- DataAxis: need to refine non-Measurement types from example products.
- The concept of 'data-subspace'? How are these described/handled by others?
 - Intervals and Regions are sub-components of STC which have not been worked in the new model family. (FOV does touch on Regions)
- Virtual data: The model can accommodate virtual data; Chandra's (ra,dec) are expressed as functions only. How common is this?
 - version 1 could be non-virtual only, requires providers to convert on output.

- Dataset currently == Metadata Only. I would prefer a container with both (unsure of relation), but not sure we can impose defining all IVOA data products as a DataProduct.
 - I think we intuitively expect a “Dataset” to include both the Header and the Data.. ie: the Dataset and the DataProduct objects.
 - There is a difference between ‘multiple files’ and ‘multiple products’. We can serialize an Event list as 1 event per file; this does not effect the model, but is significant for the annotation and software interface.

Discussion points

- Other structural requirements?
- Other associated dependencies?
- Other data types?

