VIRTUAL ASTRONOMICAL OBSERVATORY

Image Data Model – Fall 2013

Doug Tody, NRAO, USVAO
on behalf of the full ImageDM author group

Image Data Model

- Status
  - Working Draft
    - Updated working draft released to author group 16 Aug
    - Complete except for access model
    - Includes a full data model spreadsheet with all metadata
  - Prototyping
    - VOA prototype (within SIAV2) has been available for several weeks
    - Can see DM instances in VOTable by querying against service
    - A number of real image/cube datasets included
      - ALMA, JVLA, legacy VLA, JWST (partial), JCMT, Keck, Ned, Califa, etc.
  - Future Plans
    - Access model is major item to be done next – need prototyping with real data and real client software

ImageDM and FITS

- Why Not Just use FITS?
  - Not just because VO does its own thing
  - Separate DM abstraction from serialization
    - Retain key FITS models, e.g. Image, WCS
    - Use VO-compatible representations, e.g. Utypes
  - Inherit / Extend from common VO models
    - ObsCore, Characterization, etc.
    - Consistent with other models, e.g, SpectralDM
  - Explicit support for issues of cube data
    - Sparse data, time cubes, better support for polarization
    - Access model for advanced access to large images/cubes
  - Support multiple serializations
    - FITS, VOTable, CASA, Starlink NDF, JPEG-2000, etc.
    - Dynamic FORMAT transformation

ImageDM Use Cases

- Scope of Use Cases
  - Not science use cases
    - Rather, derived from science use cases
    - Used to drive DM architecture
  - Based mainly upon
    - Heidelberg cube session and VAO cube whitepaper
    - Cube test data, cube prototyping
- Feedback
  - Need to tie back in to science use cases better

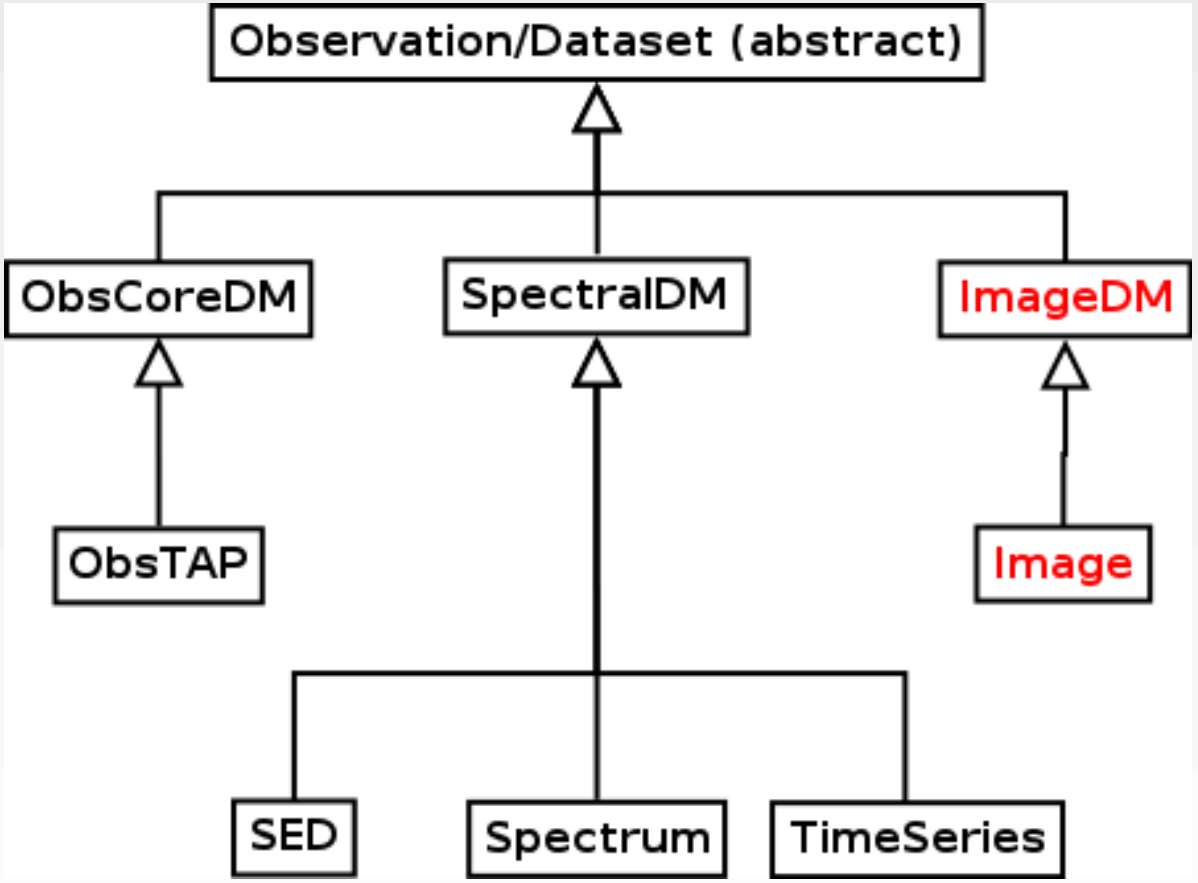
ImageDM Use Cases

- DM Architecture Use Cases
  - 1. Simple image
    - Overall Dataset metadata plus a single Data element (array)
  - 2. Single sparse image
    - Like simple image (single Data element), but not fully sampled
  - 3a. Multiple-subarray as a single Image dataset
    - Multiple Data elements share common dataset metadata
  - 3b. Aggregation of images to model more complex data
    - Aggregation of separate, distinct Image instances
    - Any metadata may differ in each image.

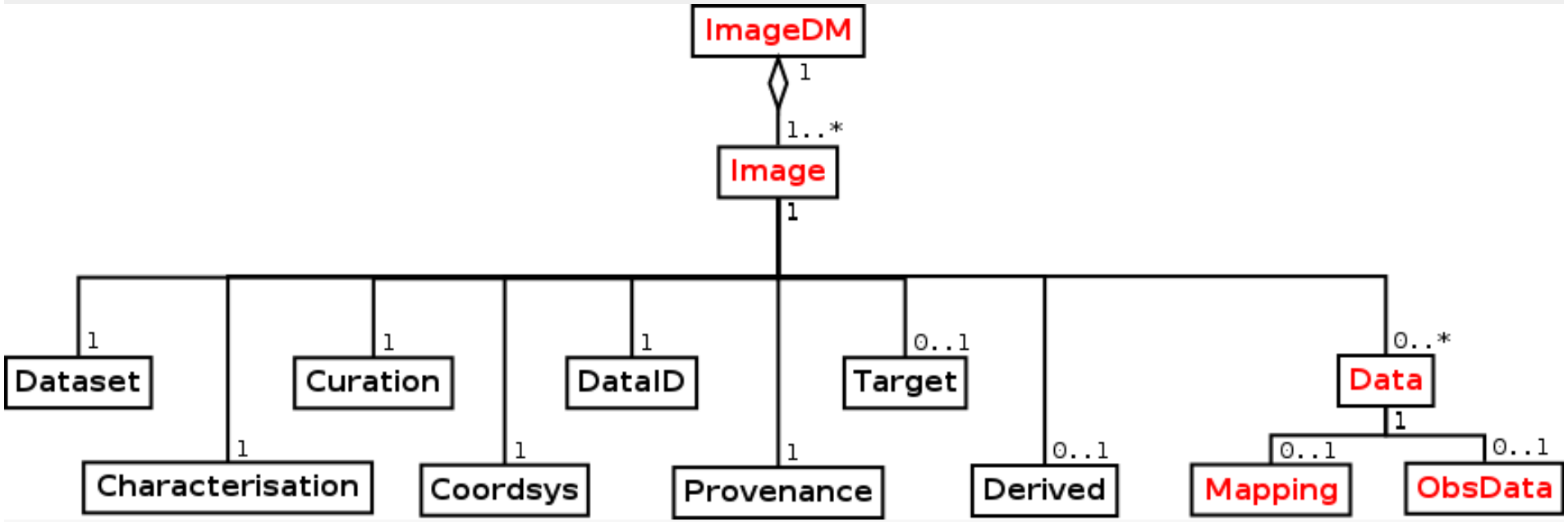
ImageDM Use Cases

- DM Architecture Use Cases
  - 4. Large cube
    - Supported mainly by #1(simple image), #3 (multi-subarray)
    - Data abstraction is key, to hide how data is physically stored
    - Large cubes can no longer be stored as a single file
  - 5. Wide-Field survey
    - May or may not expose individual survey fields
    - May require only virtual data access if fields are hidden
  - (6) Pass-through of other hypercube data, e.g., event data.

ImageDM Architecture



ImageDM – Data Model Elements



Data Model Classes

- Major Elements
  - Main model
    - Generic dataset metadata describing overall dataset
    - Mostly common with other models
  - Data Element model
    - Describes each image sub-array
    - Most Image-specific part of data model

Dataset.Image

- Concept
  - Summarizes most important image-specific metadata
  - Note Data element metadata is not included

UTYPE	*Description*	*Req*	*Default*
Dataset.DataModel.Name	Data model name and version	MAN	Image-2.0
Dataset.DataModel.Prefix	Data model prefix	MAN	im
Dataset.DataModel.URL	Reference URL for the data model	OPT	
Dataset.Type	Type of VO dataset	MAN	image
Dataset.Subtype	Type of data product (archive-specific)	OPT	
Dataset.CalibLevel	Calibration level	MAN	
Dataset.Length	Total number of voxels in image dataset	MAN	
Dataset.Image.Nsubarrays	Number of image subarrays	MAN	
Dataset.Image.Naxes	Number of physical image axes	MAN	
Dataset.Image.Naxis	Length of each image axis	MAN	
Dataset.Image.Pixtype	Pixel datatype	MAN	
Dataset.Image.WCSAxes	Enumeration of the WCS axes types	MAN	
Dataset.Image.DataRef	Reference URL for Data element metadata	OPT	

Data Element

- Usage
  - One Data element per sub-array
  - Geometry, WCS may differ for each sub-array
  - Supports pass-through of non-image hypercube data

UTYPE	*Description*	*REQ*	*Default*
Data.ID	Unique identifier for the Data element	OPT	
Data.Naxes	Number of image axes	MAN	
Data.Naxis	Length of each axis in pixels	MAN	
Data.Pixtype	Pixel / voxel datatype	MAN	
Data.Encoding	Encoding used for the array data	OPT	"FITS"
Data.Length	Array length in voxels (actual count if sparse)	MAN	
Data.Size	Data array element size in bytes	MAN	
Data.Values	Array data	OPT	
Data.Mapping.*	World coordinate system	OPT	
Data.ObsData.*	Reference to original observational data	OPT	

Data Access Model

- Concept
  - Define standard model for accessing *n*-D images
  - Dataflow from input image to output data product
  - Define standard access operations

- Model Elements
  - Filter -> WCS transform -> Pixel transform -> Function
  - All elements are optional

- Status and Issues
  - Current "cutout" prototyping addresses Filter, Pixel transform
    - To be covered in DAL/DM session on Friday
  - Filter can be expressed either using STC-S or per-axis parameters
  - Need to support range lists for spectral and time axis
  - Functions are a challenge, but needed to access large cubes

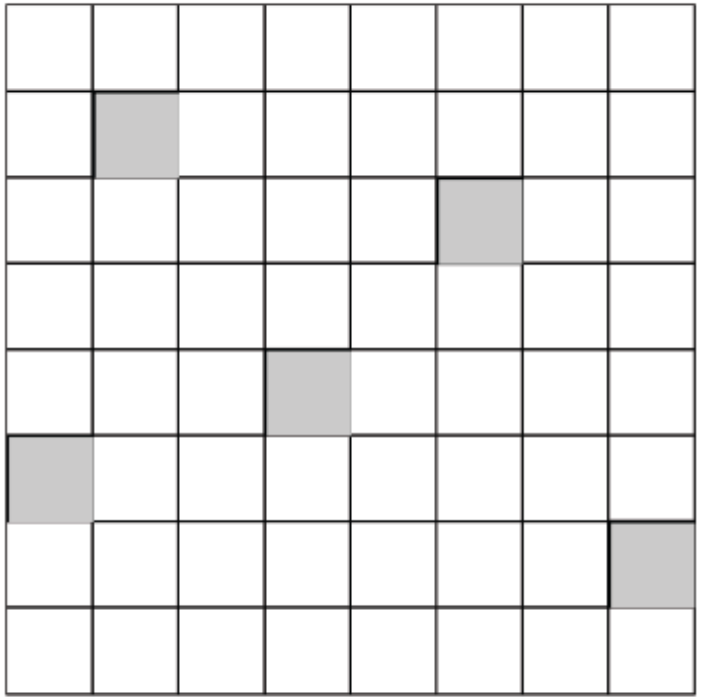
Image DM Spreadsheet

	A	B	C	D	E	F	G
1	**Field ID**	Type	**UTYPE**	**UCD**	**Description**	**FITS**	**CSV**
2	#		***Legend**: (RED) New fields (BLUE) Changed fields*				
3	**## ACCESS Model**						
4		**Query**	**Query**		**Query Metadata**		
5	Score	Double	Query.Score		Degree of match to query parameters		
6	Token	String	Query.Token		Continuation token for large queries		
7							
8		**Association**	**Association**		**Association Metadata**		
9	AssocType	String	Association.Type		Type of association		
10	AssocID	String	Association.ID	meta.id	Association identifier		
11	AssocKey	String	Association.Key		Key used to distinguish association elements		
12							
13		**Access**	**Access**		**Access Metadata**		
14	AcRef	URL	Access.Reference	meta.ref.url	URL used to access dataset		
15	Format	String	Access.Format		Content or MIME type of dataset		8;@ID
16	EstSize	Long	Access.Size		Estimated dataset size		7;@ID
17							
18	# **Field ID**	Type	**UTYPE**	**UCD**	**Description**	**FITS**	**CSV**
19	**## CORE Model**						
20		**Dataset**	**Dataset**		**General Dataset Metadata**		
21	DataModel	String	Dataset.DataModel.Name		Data model name and version	VOCLASS	
22	DataModelPrefix	String	Dataset.DataModel.Prefix		Data model prefix		
23	DataModelURL	String	Dataset.DataModel.URL	meta.ref.url	Reference URL for data model		
24	DatasetType	String	Dataset.Type		Dataset type		
25	DatasetSubtype	String	Dataset.Subtype		Dataset subtype (external type)		
26	DatasetCalibLevel	Long	Dataset.CalibLevel		Calibration level		
27	DataLength	Long	Dataset.Length	meta.number	Number of pixels		
28	Deleted	String	Dataset.Deleted		Set if dataset is deleted		
29		**ImageDataset**	**Dataset.Image**		Image-specific Dataset metadata		
30	Nsubarrays	Long	Dataset.Image.Nsubarrays	meta.number	Number of image subarrays		
31	Naxes	Long	Dataset.Image.Naxes		Number of image axes		2;@ID
32	Naxis	Long[][]	Dataset.Image.Naxis		Length of each axis of each subarray		3;@ID
33	Pixtype	String	Dataset.Image.Pixtype		Pixel datatype		
34	WCSAxes	String[]	Dataset.Image.WCSAxes		WCS axis coordinate types		4;@ID
35	DataRef	String	Dataset.Image.DataRef	meta.ref.url	Access reference URL for Data element metadata		
36							
37		**DataID**	**DataID**		**Dataset Identification Metadata**		
38	Title	String	DataID.Title	meta.title;meta.dataset	Dataset Title	TITLE	5;@ID
39	Creator	String	DataID.Creator	meta.curation	Dataset creator	AUTHOR	
40	Collection	String	DataID.Collection		Data collection to which dataset belongs	COLLECT	6;@ID
41	DatasetID	URI	DataID.DatasetID	meta.id;meta.dataset	IVOA Dataset ID	DS_IDENT	
42	CreatorDID	URI	DataID.CreatorDID	meta.id	Creator's ID for the dataset	CR_IDENT	
43	CreatorDate	String	DataID.Date	time.epoch;meta.dataset	Data processing/creation date	DATE	
44	CreatorVersion	String	DataID.Version	meta.version;meta.dataset	Version of dataset	VERSION	
45	CreationType	String	DataID.CreationType		Dataset creation type	CRETYPE	
46	CreatorLogo	URL	DataID.Logo	meta.ref.url	URL for creator logo	VOLOGO	
47	Contributor	String	DataID.Contributor		Contributor	CONTRIB	

UType Inventory

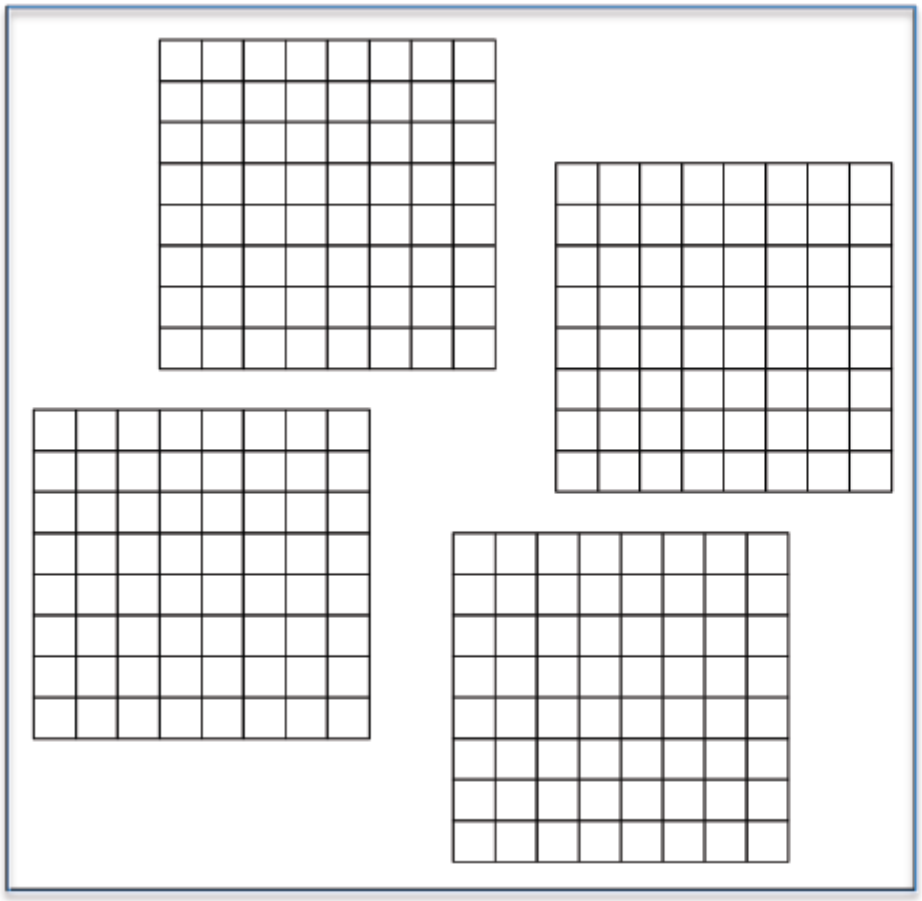
- Scope
  - ObsCore/ObsTAP, Char 1&2, SpectralDM, STC
- Highlights
  - Dataset.Type (DataProductType in ObsTAP)
  - Dataset.Subtype, Dataset.CalibLevel
  - Dataset.Image
  - ObsID (DataID.ObsID?)
  - CoordSys.FluxFrame
  - Characterisation
    - Coverage*LoLimit2Vec, HiLimit2Vec
    - Char.PolAxis (UCD, enumeration of states)
  - Data Element

Sparse Image – Selected Pixels/Voxels



Example of a **sparse image** (image or image cube which is sparse on the two coupled spatial axes). Data was obtained only for the points shown as gray in the figure. Rather than store the entire array, only data for the five sampled regions is stored. The coordinates of each sampled region are stored in a table included in the WCS for the image/ cube. In this example the sparse cube would be represented in 5/64 of the space that would be required to store the fully sampled cube.

Sparse Image - Segmented



Example of a **sparse image** (image or image cube which is sparse on the two coupled spatial axes), that is composed of several sub-arrays. The outer box defines the area of the super-array, or overall Image dataset. The four sub-arrays are individual
smaller images for which data was obtained. This example illustrates the use of multiple sub-arrays to cover a larger spatial region, however the same technique may be used for other axes such as the spectral, time, and polarization axes of a general cube.

Visibility Data

- Characterization
  - Pointing, FOV
  - UV distance plot
    - min/max UV distances, number of antennas, duration of exposure
  - Dirty beam plot
    - FWHM axes, max sidelobe expressed as % of peak
  - Freq sub-bands observed
    - support for velocity units (convention, ref frame, rest freq)
  - Resolution
    - size of synthesized beam (major, minor axes and angle)
  - Flux density, Jy/beam
  - Sensitivity, rms noise
  - Properties of possible generated images/spectra as ranges