Systematizing the Record of Earth's Shapes & Colors:

A Framework for Data and Metadata Models

IGARSS03

03.1893

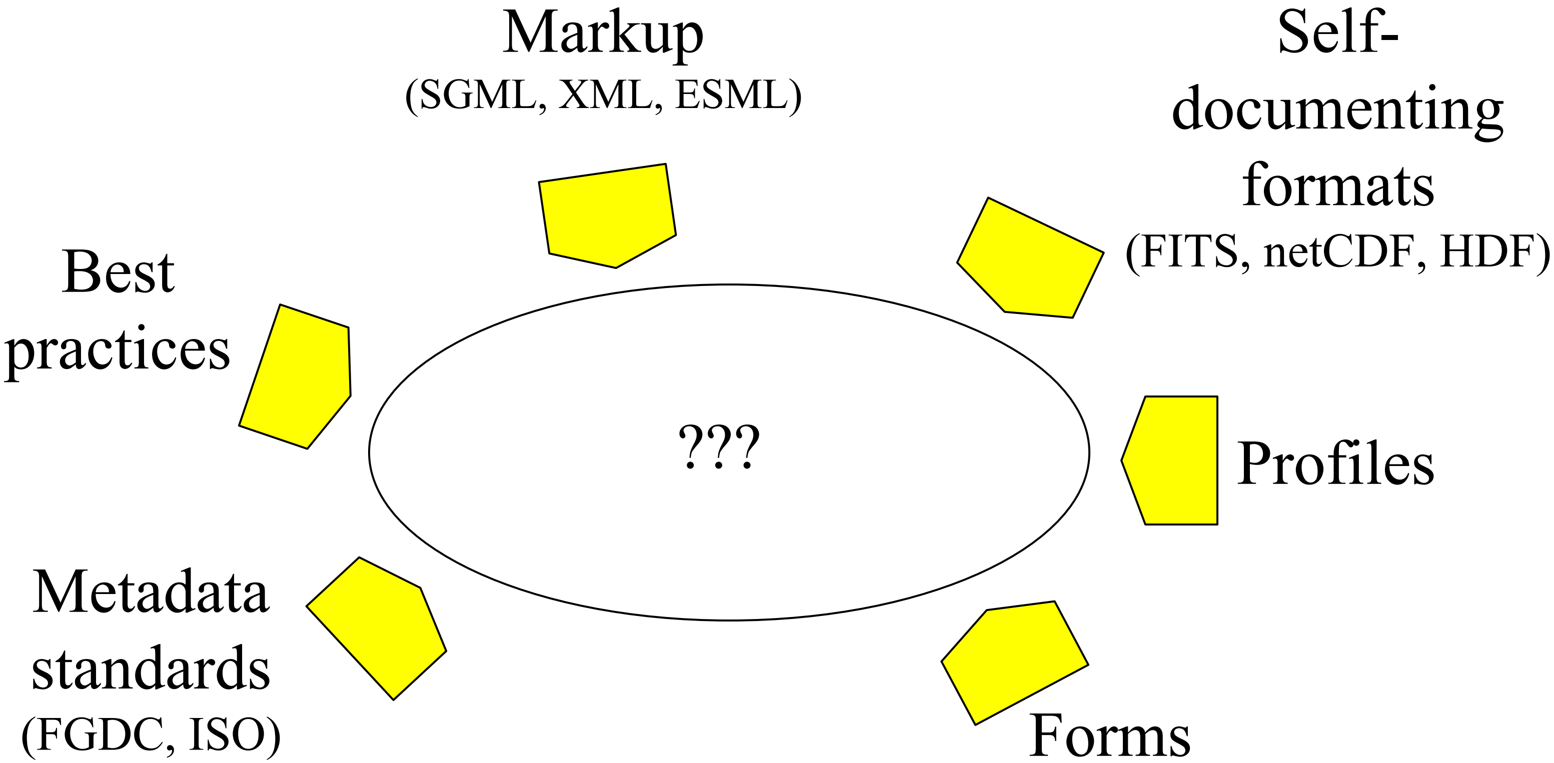
Alan M. Goldberg

MITRE

Goal of the Presentation

- Present an analysis of some underlying problems in designing complex remote sensing data sets
- Contribute to evolution of data and metadata standards
- Provide support to strong domain profile
- Not define a new standard
- Apologies to those who see this as trivial

Motivation


Markup
(SGML, XML, ESML)
Self-documenting formats
(FITS, netCDF, HDF)
Best practices
???
Profiles
Metadata standards
(FGDC, ISO)
Forms

Stress on the System

- **At the same time, sensors and data structures are more complex**
  - **Staring imagers (film, CCD arrays) and point scanners with maybe a filter wheel give way to whiskbroom and pushbroom, conical scan, and FT sensors**
    - non-linear
    - different geometry in-track and cross-track
    - bands with different geometries must be fused
    - multiple independent detector responses
    - hyperspectral
- **As temporal, spatial, and radiometric resolution have increased, so has the need for precision documentation**

Residual Problems

- Attempt to force-fit individual independent attributes to be the array indices
  - Multiple independent attributes may be associated
  - An attribute may be f() multiple indices
  - An attribute may be multi-dimensional
- Ad hoc solutions
  - Warning: someone else's general solution may appear *ad hoc* to me, and vice versa
- Lack of consistency
  - Follow all the rules, but the data is still a mystery, complex to process, or imprecisely characterized; sometimes due to the force-fit

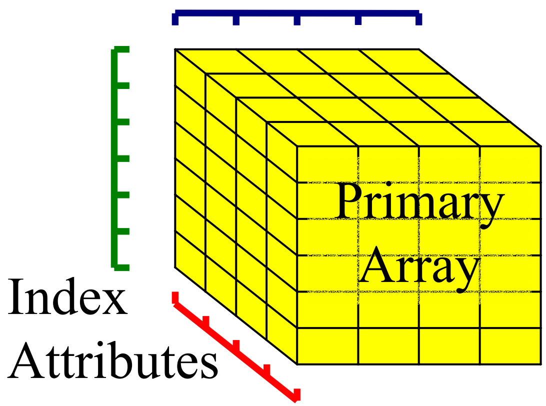
Framework Approach

- Dependent variables ("entities") are stored as arrays
  - As close to native format as practical
- Independent variables ("attributes") are associated with the array dimensions
  - One "primary" independent variable associated with each dimension
  - Zero, one, or more "secondary" independent variables associated with each dimension
  - Zero, one, or more "secondary" independent variables associated with combinations of dimensions

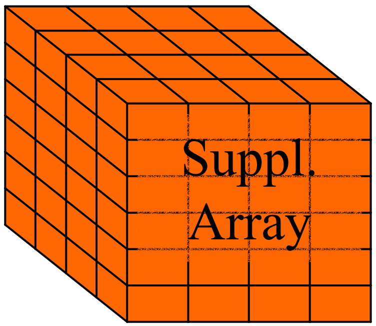
Framework Approach (cont.)

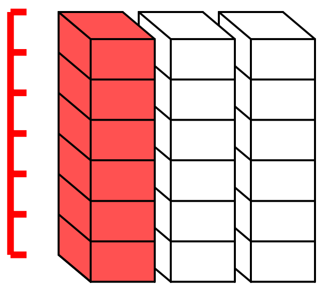
- Entity and attribute metadata
  - Semantic, syntactic, descriptive metadata
  - Hierarchically associated with values, dimensions, or combinations of dimensions
  - Scalings, polynomials, or LUTs may be applied

Building Up the Parts of the Framework

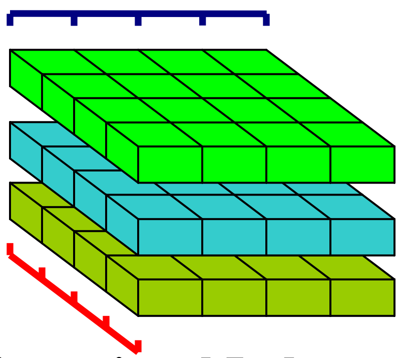


n-Dimensional Dependant Variable (Entity) Arrays



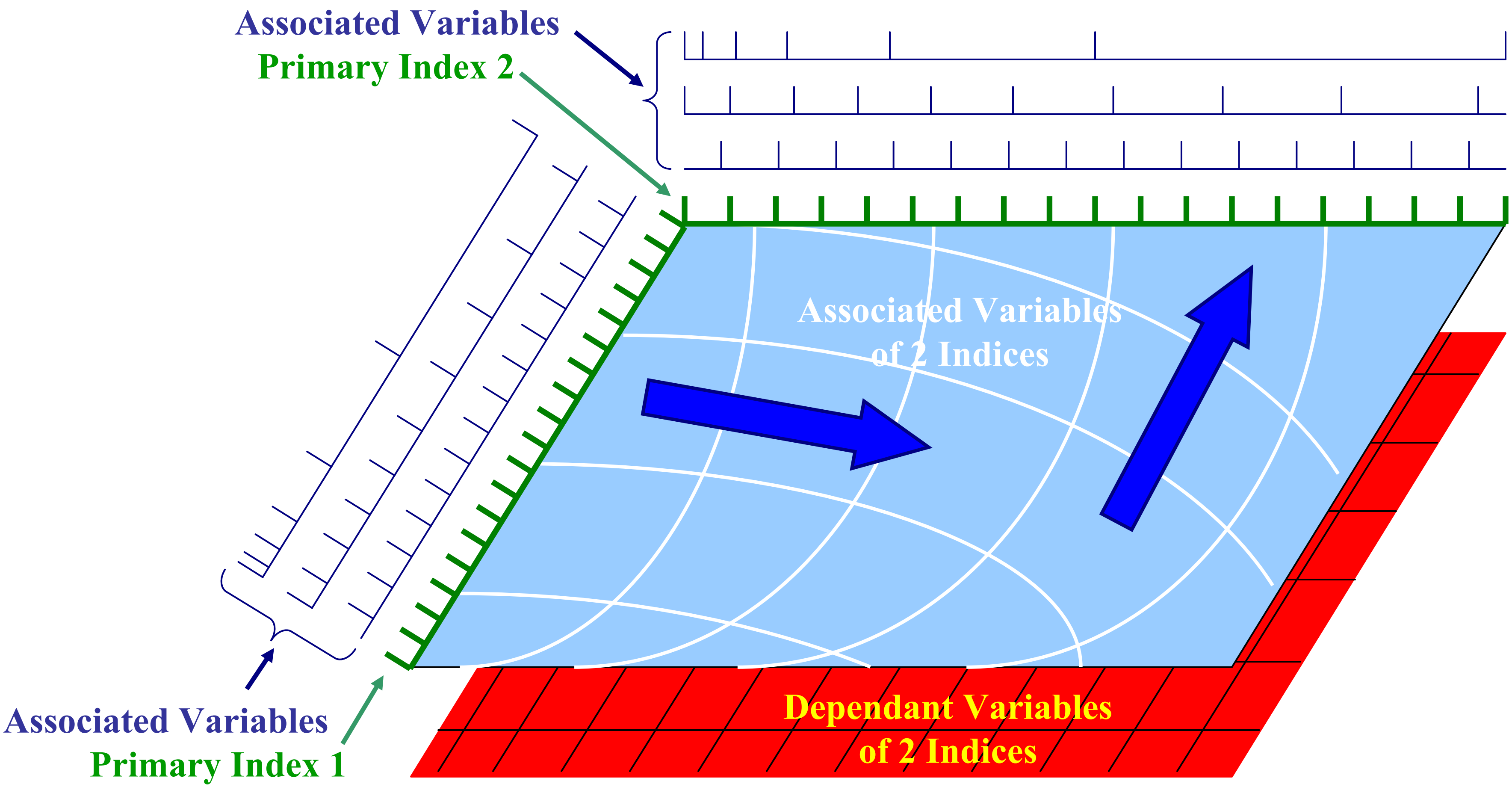


Primary and Supplementary (Attribute) Variable Arrays

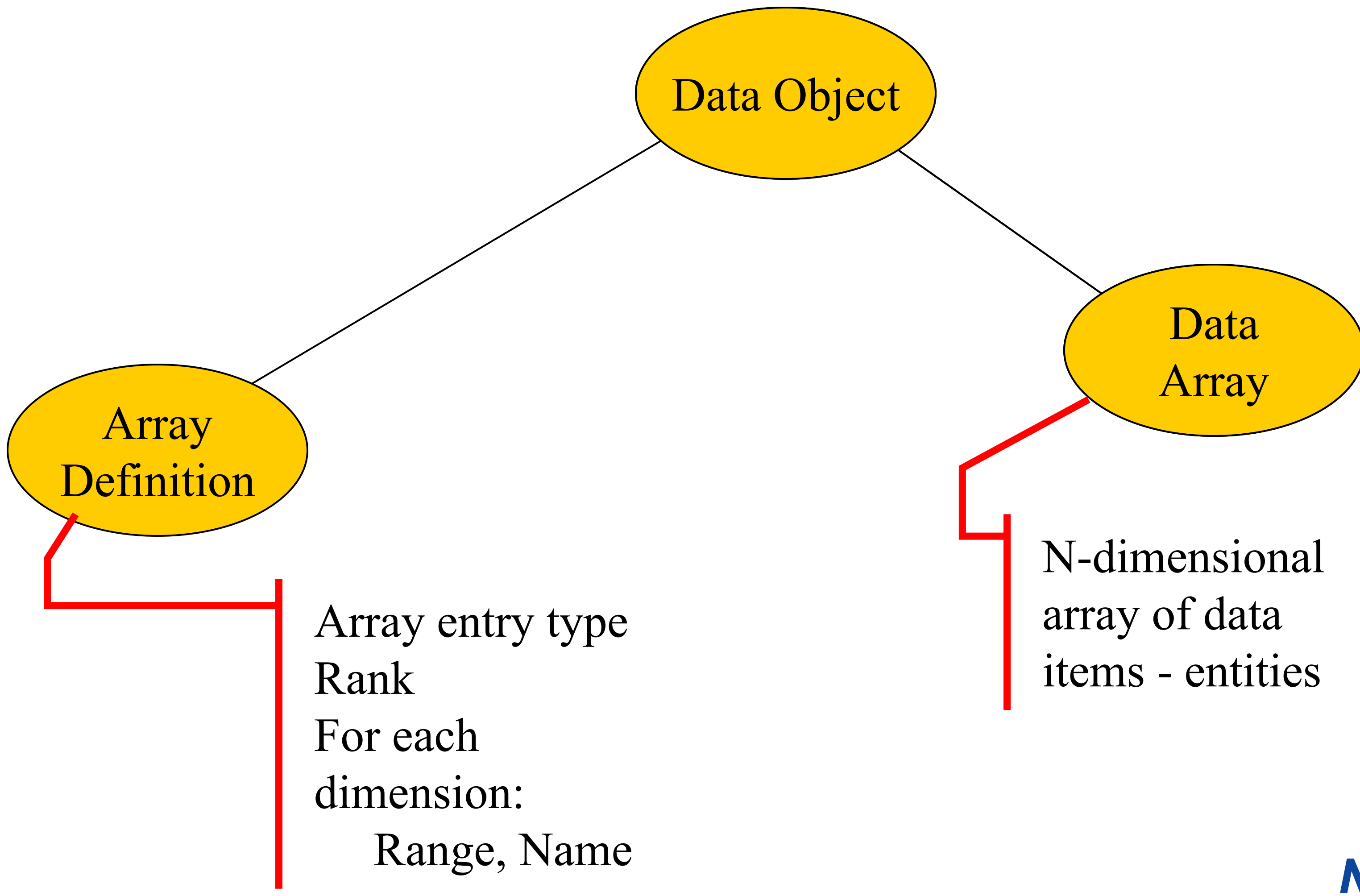


2-Dimensional Independent (Attribute) Variable Arrays

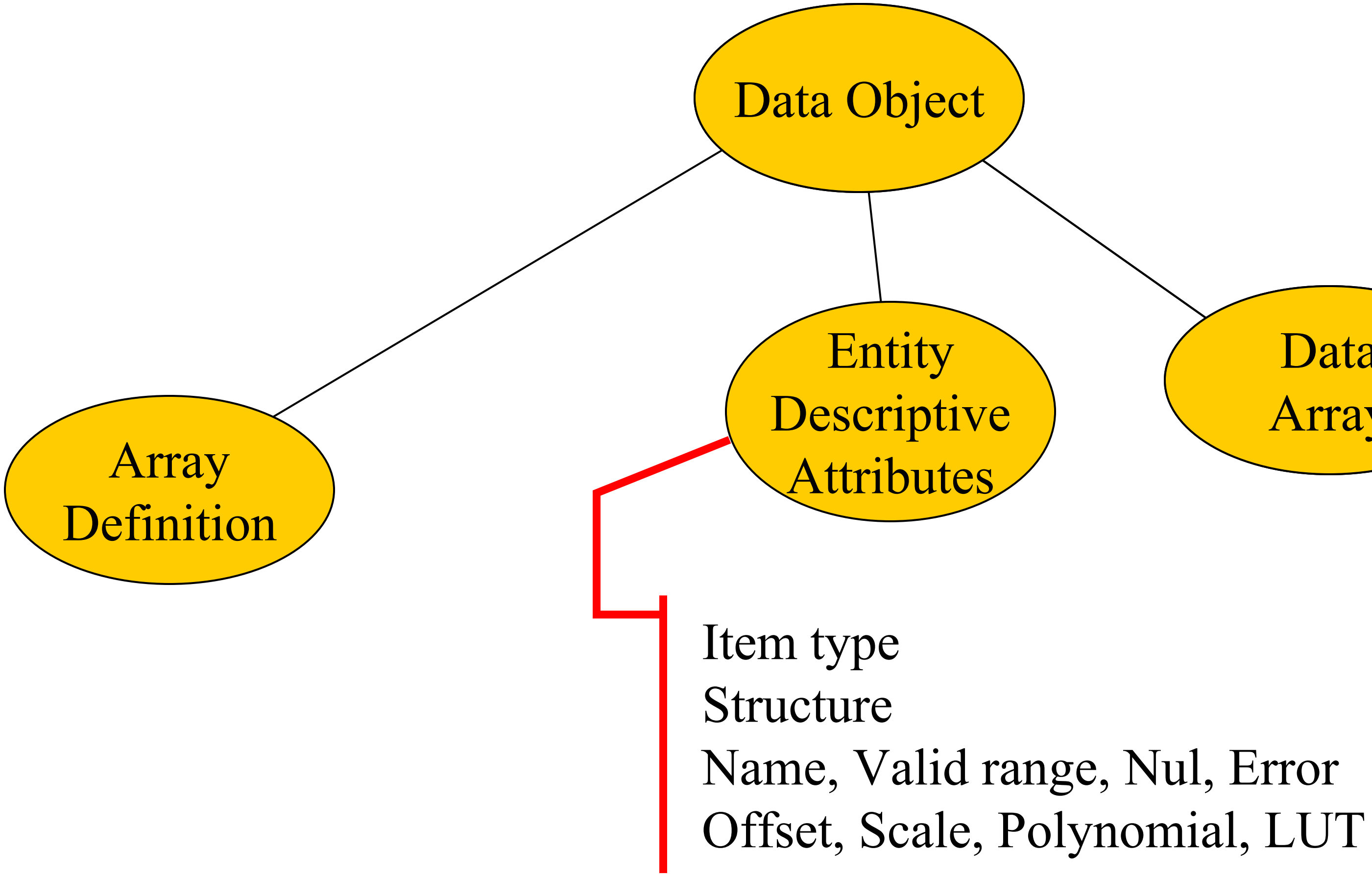
Primary and Associated Independent Variables



Basic Data-Metadata Association

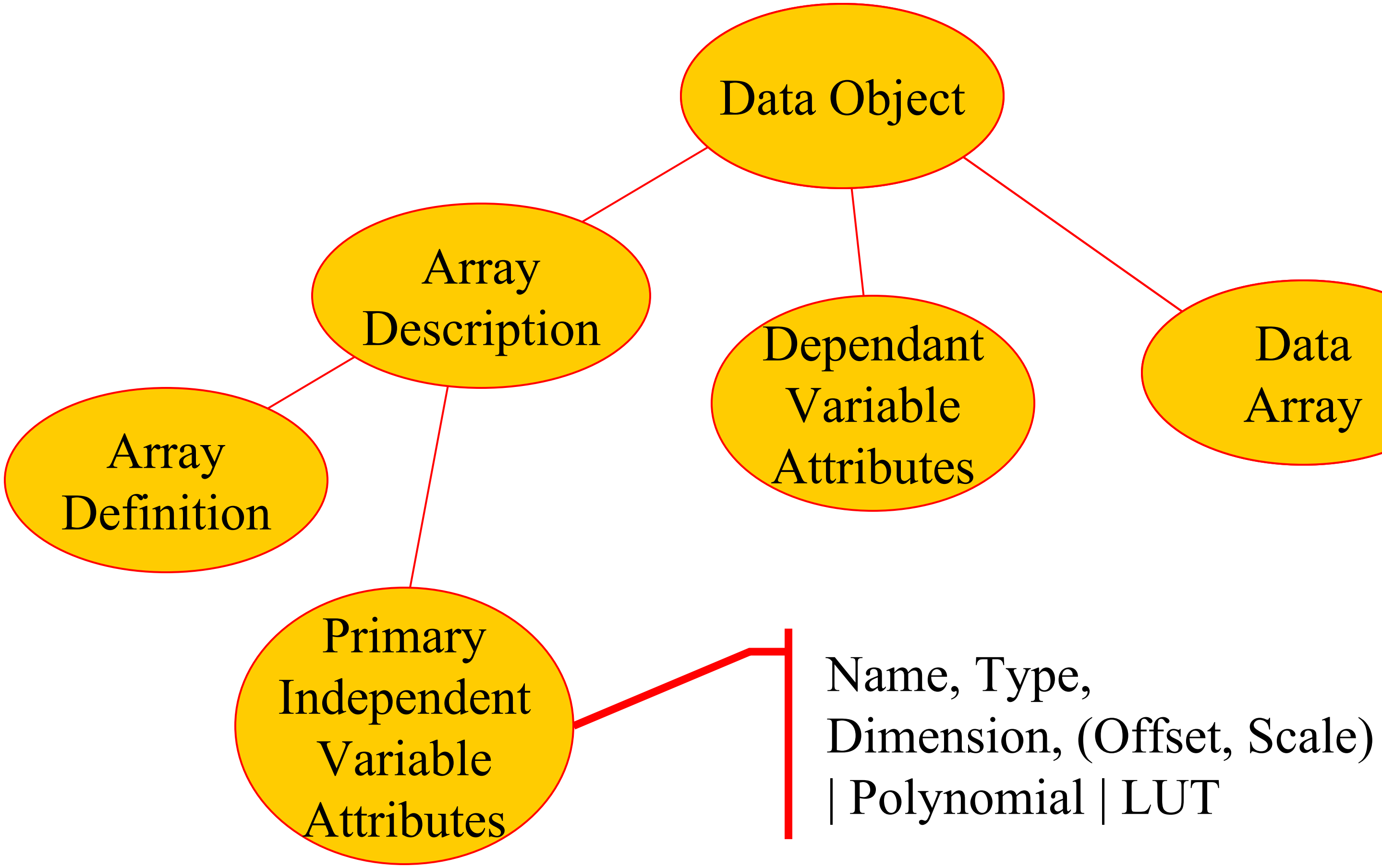


Complete Dependant Variable Description Through Metadata

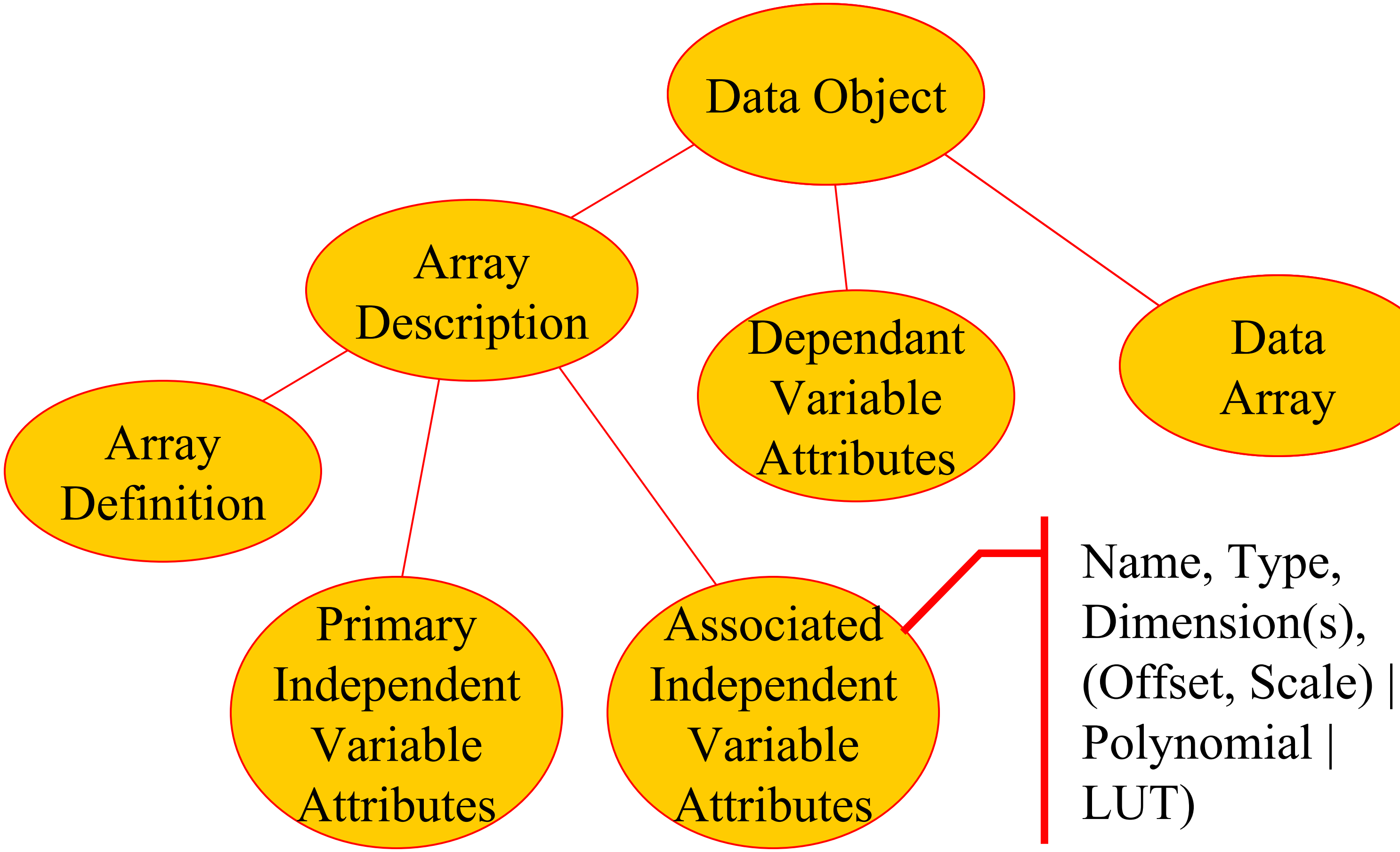


MITRE

Primary Independent Variable Description Through Metadata



Complete Independent Variable Description Through Metadata



Steps

- Identify the primary attributes of the measurand(s)
- Identify the associated attributes, and functional relationship to primary attributes
- Add annotation
  - File
  - Entities
  - Attributes
- Some dependent variables (*e.g.*, telemetry) can be organized by the same attributes

Identify the Primary Independent Variables of the Measurement System -- Attribute Indices

- **Key characteristics**
  - Fundamental parameterization of the discrete measurement system
  - One-dimensional
  - Integer or enumerated
  - Unambiguous
  - Algorithmic convenience
- **Examples**
  - Time index
  - In-track index
  - Cross track index
  - Band index
  - Channel/polarization/ energy/ time-of-flight index
  - Detector index/indices
    - row, column

Identify the Associated Independent Variables -- Attribute Variables

- **Key characteristics**
  - Items which are thought to be known, based on experiment design, and external facts
  - Functions of one or more primary independent variables
  - Meaningful types
- **Association mechanisms**
  - Complete table
  - Interpolated table
  - Scaling
  - Function
- **Examples**
  - Primary Indices converted to physical units
  - Time
  - Natural illumination
  - Viewing geometry
  - Scene orientation
  - Motion
  - Sensor associated parameters (*e.g.*, spectral calibration)

MITRE

Associate Fundamental & Ancillary Dependent Variables (Data) -- Entities

- **Key characteristics**
  - **Items measured and recorded by the sensor system**
  - **Mission data usually associated with all Primary Variables**
  - **Other data associated with one or more primary variables**
  - **Only one of each measurand at each combination of primary variable indices**

- **Examples**
  - **Mission Data**
    - raw engineering units
    - calibrated units
    - QC evaluation
  - **Ancillary Platform Data**
    - position, orientation, operational mode, environmental parameters
  - **Ancillary Sensor Data**
    - scan, operational mode, engineering parameters

MITRE

Data Items - Entities

- Simple types
  - integer, float, boolean, text
- *n*-dimensional array
  - *e.g.*, spectral radiance; vector magnetometer measurement; rotation matrix
- Structure
  - *e.g.*, measurement and QC flags; hierarchical data objects
- Clear decision must be made between associated entities and metadata

Determine Secondary Independent Variables -- Derived Attributes

- **Key characteristics**
  - **Functions of the primary variables, associated variables, and "trusted" measurements**
  - **Derived from what is known and thought to be known**
  - **Single or multi-dimensional**
- **Examples**
  - **Location, orientation, rates of change**
    - alternative frames and coordinate systems
  - **Sensor IFOV direction**
  - **Scene geometry**
    - geolocation, height, orientation
  - **Scene illumination; glint**
  - **Coarse range**
  - **Expected scene type**
    - land, ocean, space

Determine Derived Dependent Variables -- More Entities

- **Calibrated measurements**
- **Illumination-corrected & geometry-corrected scene properties**
- **Quality control parameters**
- **New independent variables may be needed**
  - ***e.g.*, inferred range or height**

MITRE

A Simple Example: Color CCD Camera

- Primary independent indices
  - frame number F[1...128]
  - column C[1...480]
  - row R[1...640]
  - band B[1...4] = B[r,g,b,m]
- Associated
  - vertical offset angle V = f(R)
  - horizontal offset angle H = f(C)
- Measurements
  - intensity counts I(F,R,C,B)
  - GPS camera location O(F)
  - time t(F)
- Derived associated
  - solar elevation S(X(F), t(F))

MITRE

A Better Example: Whiskbroom Sensor

- Linear array for each band, scanned cross-track while platform moves in-track
- Primary independent indices
  - scan line F[1...]
  - cross-track position C[1…6000]
  - detector R[1…32]
  - band B[1…20]
- Associated
  - in-track offset angle V = f(R)
  - cross-track offset angle H = f(C)
- Measurements
  - intensity counts I(F,R,C,B)
  - sensor location O(t)
  - sensor orientation Θ(t)
  - time t(F,C)
- Derived associated
  - calibrated brightness B(I)
  - geolocation X(F, C, R, O, Θ, t)
  - solar direction S(X, t)
  - satellite direction ...
  - orbit number
  - mission elapsed time

MITRE

Observations

- **Limitations**
  - **Everything changes when you resample the data**
  - **Irregular structures**
- **Opportunities**
  - **Assertion: some explicit derived data layers can be eliminated by providing the functional relationship**
    - trading storage against processing
  - **Hierarchical implementation is possible**
    - dependant variable might be vector, tensor, …
  - **Opportunity for content-specific compression**

Next Steps

- Demonstrate a specific implementation
- Work with standards bodies to facilitate generalization
  - Metadata and markup
- Evaluate the role of functions replacing tables
- Applying these framework concepts to NPOESS

Conclusions

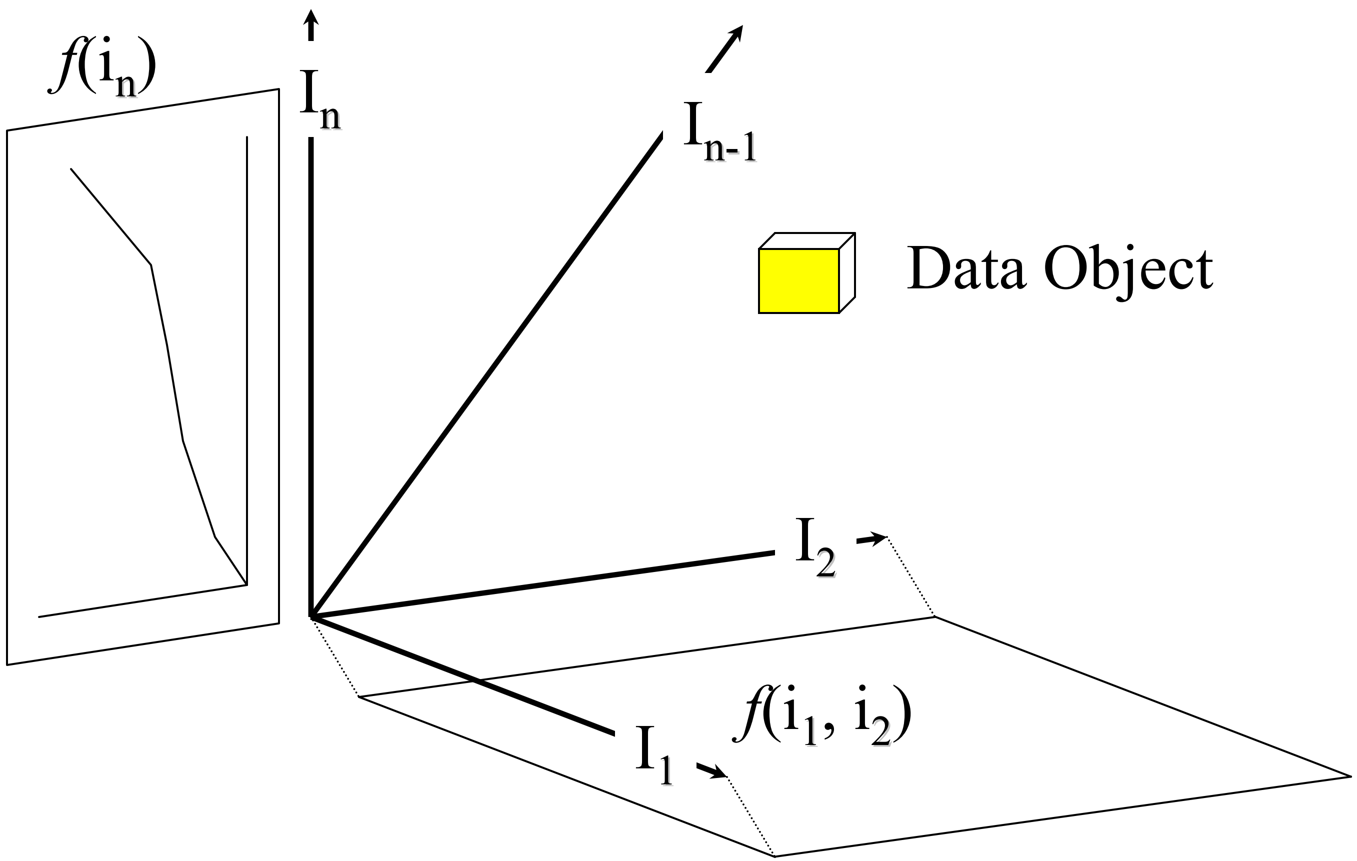
- Most or all of the pieces of the puzzle exist
- A consistent approach can be defined for planning a data system
- Requires insight into the application and engineering characteristics of the end-to-end system
- Requires considering data system impact
- Heritage systems are not always the best guide
  - How do we migrate?
  - How will the future migrate?

Backup

Scaling the Variables

- Data objects are fundamentally identified by indices
- Generally, a physically meaningful variable may be related to the index by
  - scaling
  - polynomial
  - look-up-table (LUT)
- There can be more than one variable associated with an index
  - e.g., time and scan angle; height and pressure
- A vector or matrix variable may be associated with an index
- A variable may be associates with more than one index

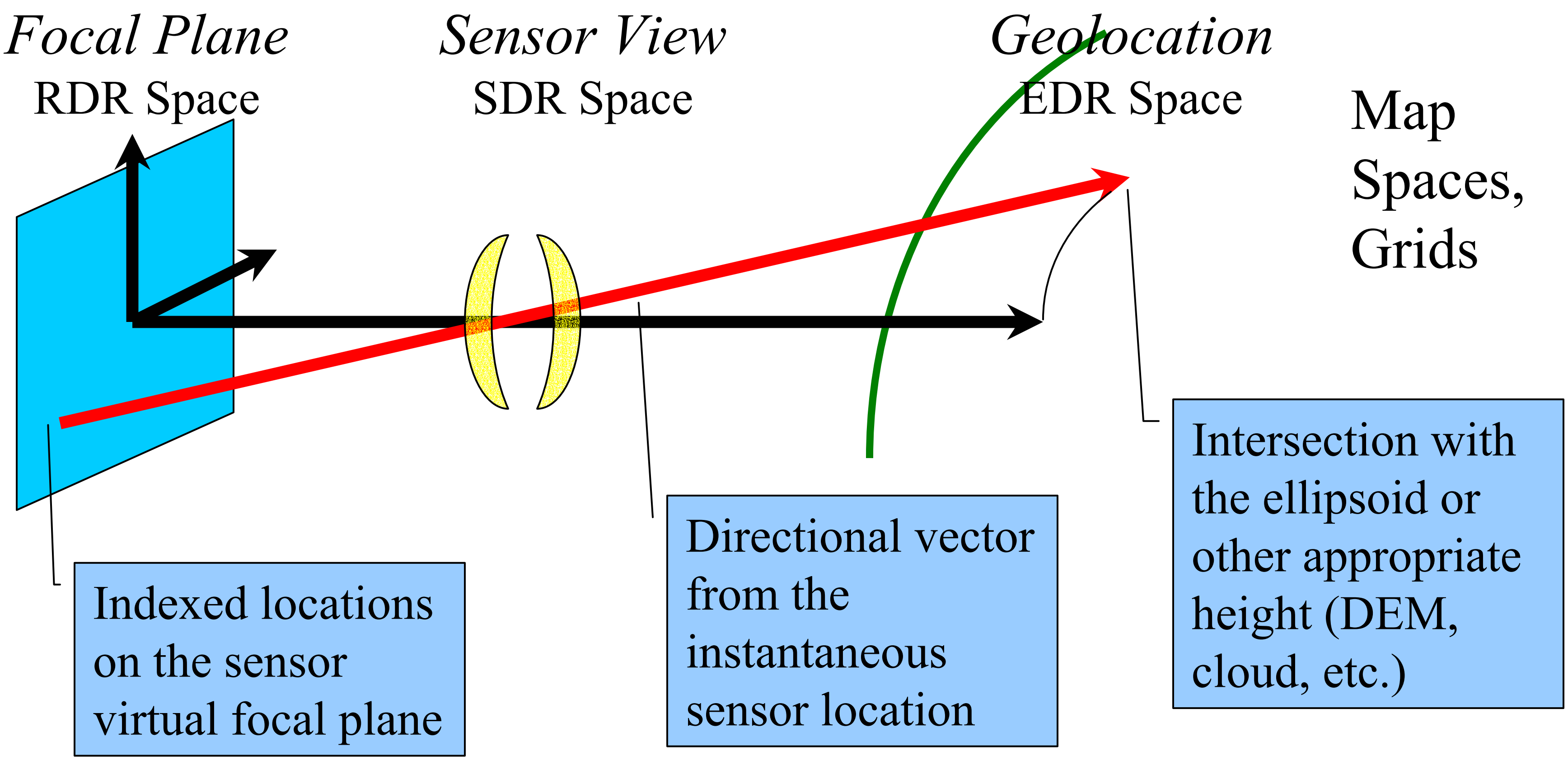
Data Array Dimensions



Data objects may by located by any number of indices.
Independent variables may be functions of one or more indices.

Coordinate System Transformations in the Framework

Measurements Appear in 3 Coordinate Spaces in IDPS



Defined Coordinate Frames

- Detector focal plane frame
  - rotated by scanner to

⇐ RDR data

- Sensor frame
  - rotated by alignment to

⇐ SDR data (primary)

- Spacecraft frame
  - rotated by orbit & jitter to
- Earth centered inertial (ECI) frame
  - rotated by Earth rotation to
- Earth centered fixed (ECF) frame

SDR data (secondary) &
⇐ EDR data

Auxiliary Data for Whisk-Broom Imager Geolocation

$$[X_{view}] = [M_{earthrot}][M_{\text{❹}}][M_{align}][M_{\text{❸}}][X_{\text{❶❷}}]$$