HDF5 for NPOESS Data Products

Alan M. Goldberg
The MITRE Corporation
agoldber@mitre.org

This work was performed under NOAA contract 50-SPNA-9-00010.
Opinions expressed are those of the author, and do not represent the United States Government.

Organization: W803
Project: 1400NT01-SE

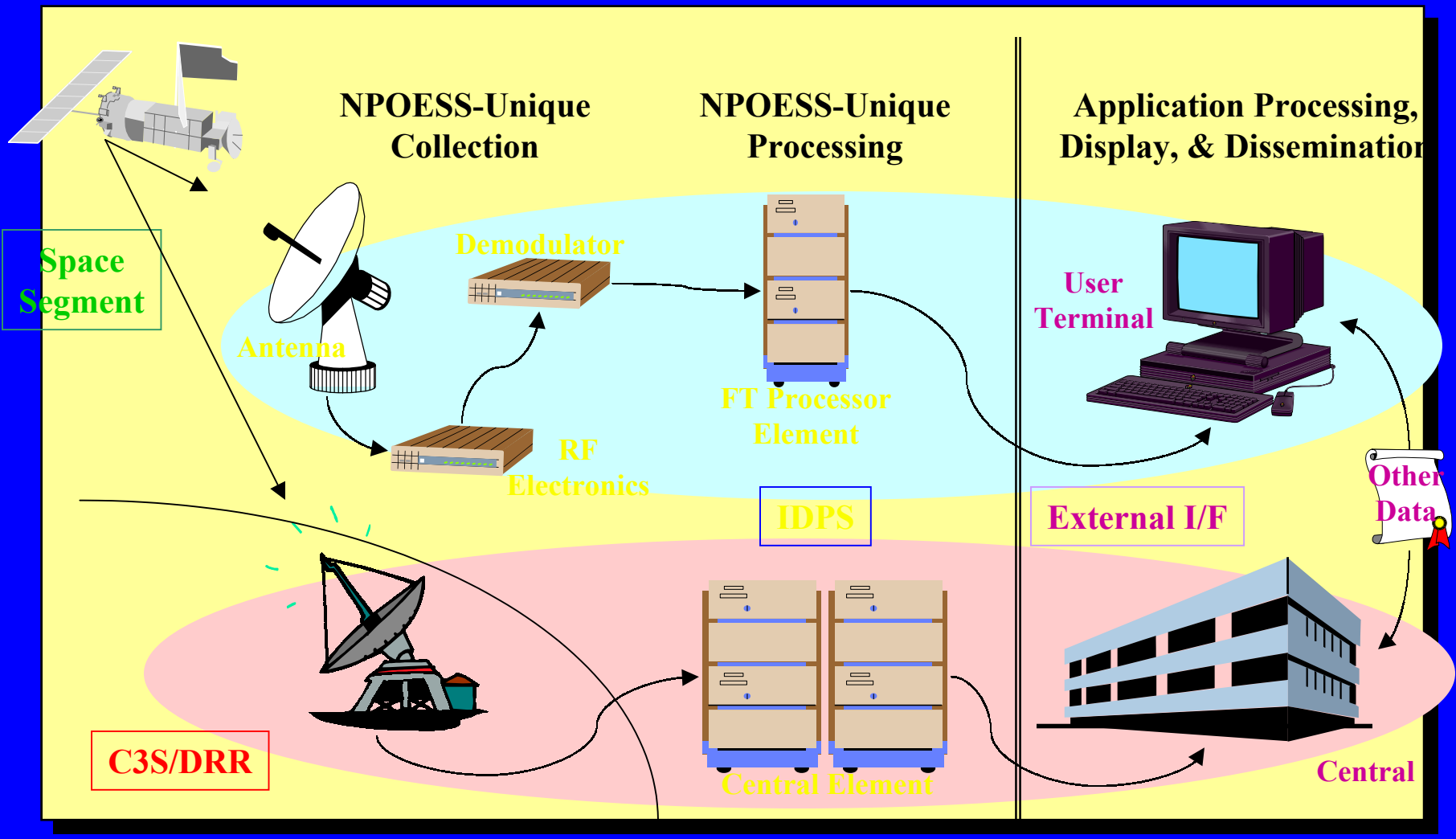
MITRE

The NPOESS Program

- Products
- Requirements
- HDF Implementation (Presentation by Chad Fox / Raytheon)

MITRE

Data Flow Overview
Field Terminals and Centrals

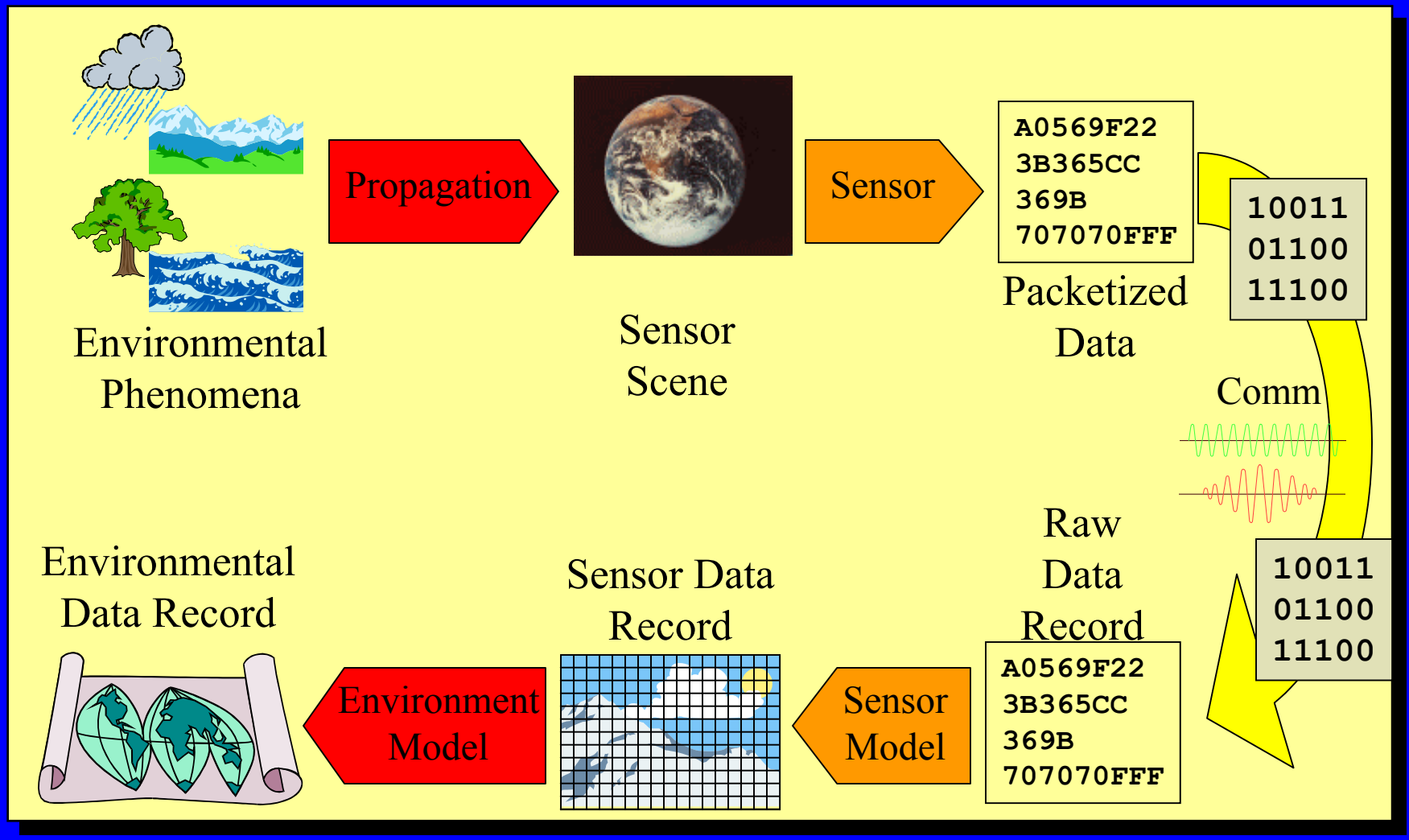


MITRE

Data Delivery to Users

- Centrals *
  - National Environmental Satellite, Data, and Information Service
  - Naval Oceanography Command
  - Fleet Numerical Meteorology & Oceanography Command
  - Air Force Weather Agency
- Field terminals *
  - US government
  - Private
  - Worldwide users
- NESDIS Archives *
- Misc
  - NASA (for NPP), SARSat, DCS, etc.

Major Product Relationships



MITRE

Product Types

- Major products
  - Raw data records *
  - Sensor Data Records *
  - Temperature Data Records *
  - Environmental data records *
- Other
  - Raw data
  - Auxiliary data
  - Ancillary Data
  - Calibration & correction data
  - Telemetry & housekeeping
  - Memory dumps and diagnostics

Delivery Requirements

- Operational *; platform independence; suitable for transmission, efficient conversion, and storage
- Standards based: Interoperability*, consistent with Joint Technical Architecture and National Spatial Data Infrastructure
- Rapid & available*: >95% delivered within 30 min. at Centrals; 15 min. goal
- Self-documenting
- Flexible for over 100 initial products
- Flexible for evolutionary or new sensors and algorithms
- User selectable aggregation (from one granule up to full orbit)
- Delivery: Push or pull; Assured; Secure*
- Efficient: 3 TB per day at each Central
- Consistent interface for delivery to Centrals*, Field terminals*, & long term archives*

* key requirement

MITRE

Schedule

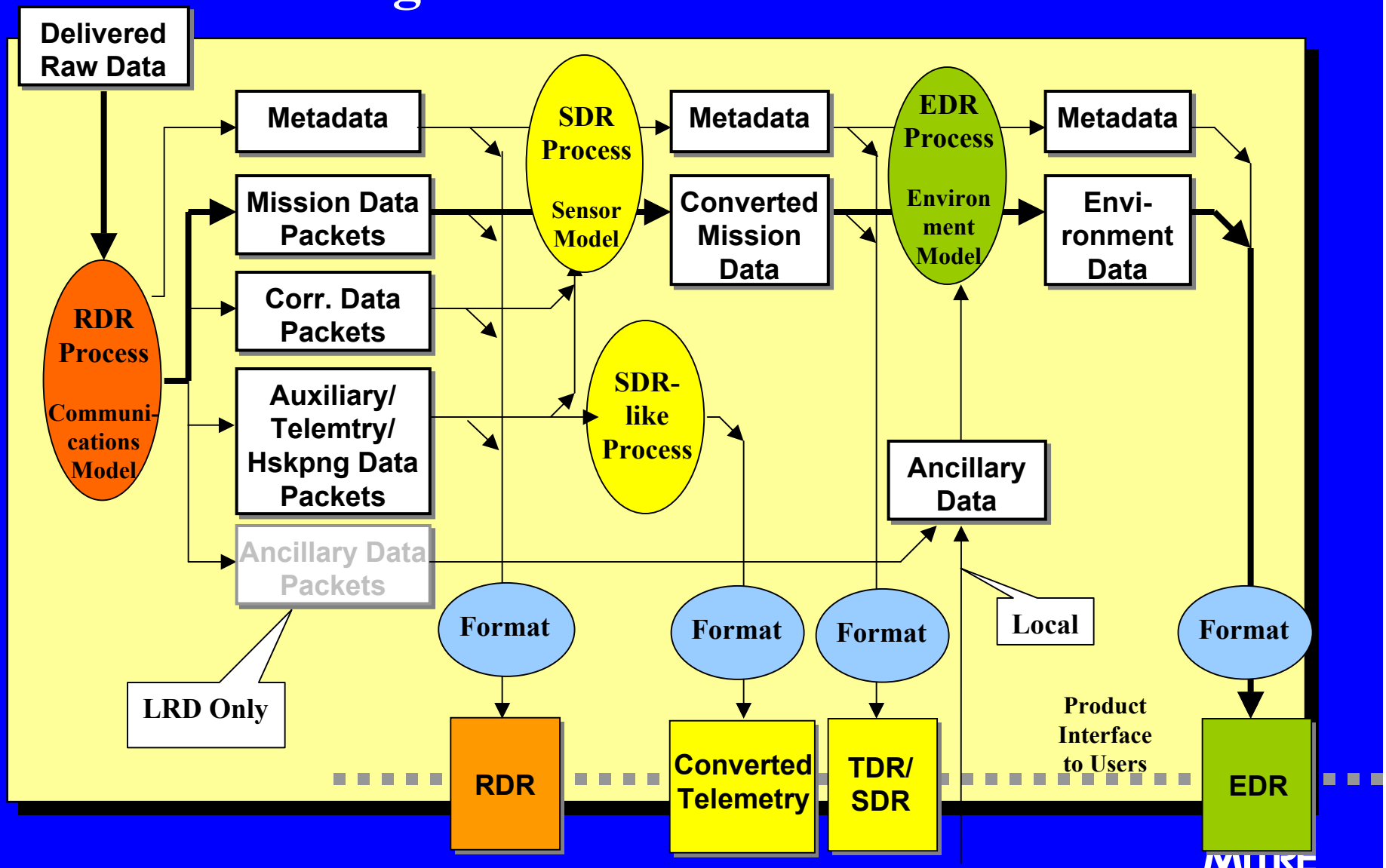
- Interface with Centrals to be published (draft) in spring 2003.
- First deliverable version of IDPS to be ready in spring 2005 for NPP risk-reduction mission in spring 2006
- Hardware specification for software support at field terminals to be published by Oct. 2005
- Operational system to be ready in mid 2008 for first NPOESS launch in mid 2009.

Issues

- How much EOSDIS heritage to retain, if any, *e.g.*
  - whether to use EOS swath construct
- Develop an NPOESS profile to handle particular attributes of NPOESS data, *e.g.*
  - variable length compressed packets in RDRs
  - conical scan geometry
- Assure long-term stability of the standard
- Provide user support
- Suitability for archival use
  - HDF5 is primarily defined by its API, not the format

Backup

Data Processing Flow in IDPS



Raw Data Granule Contents

- CCSDS Application Packets, unprocessed, generally from one sensor over a short time interval
- Mission data as produced by the sensor or payload
- Telemetry
- Calibration data & Correction parameters
- Auxiliary data: spacecraft location & attitude

MITRE

Sensor/Temperature Granule Contents

- Radiometrically corrected data, which is an estimate of the flux at the sensor aperture.
- Mission data by in-track position, cross-track position, channel, detector, etc.
- Georeferenced in spacecraft coordinates, & intersection with the ellipsoid
- Generally, not filtered, resampled, etc.

- TDRs are microwave SDRs without antenna pattern removed

Environmental Data Granule Contents

- SDRs processed with environmental models and ancillary data (by others) to produce estimates of environmental parameters
- Reported as cross-track / in-track values identified in earth coordinates
  - unless resampling is specifically required

Metadata - Scope

- **Identifications:**
  - Name / Level / Type-subtype / Sequential number
- **Description:**
  - Start-end date-time; Start-end orbit
  - Subsatellite coordinates; Corner coordinates
- **History:**
  - Space segment components / DRR components / Source files / Algorithm versions / Processing facility / Processing date-time
- **Contents:**
  - Size / Granules
  - Predecessor quality / Missing data / Out-of-bounds / Processing errors / Quality free text
- **Formats:**
  - Data elements / Structures / Formats

Metadata - Structure

- As goal, all metadata will be compliant with NSDI standards
  - <tag> = “parameter”, where <tag> is defined by FGDC or the NPOESS extension profile.
- Metadata will be hierarchical, as appropriate:
  - file-metadata
    - granule_metadata
      - scan_metadata
        - » point_metadata
- Metadata should map to the file naming convention
- Metadata will be included with associated data files, and accessible as a stand-alone file.

MITRE

Data

- Hierarchical
- Sensor specific format
- “Internal” metadata, such as detailed geolocation, quality, annotation, illumination & view geometry

MITRE

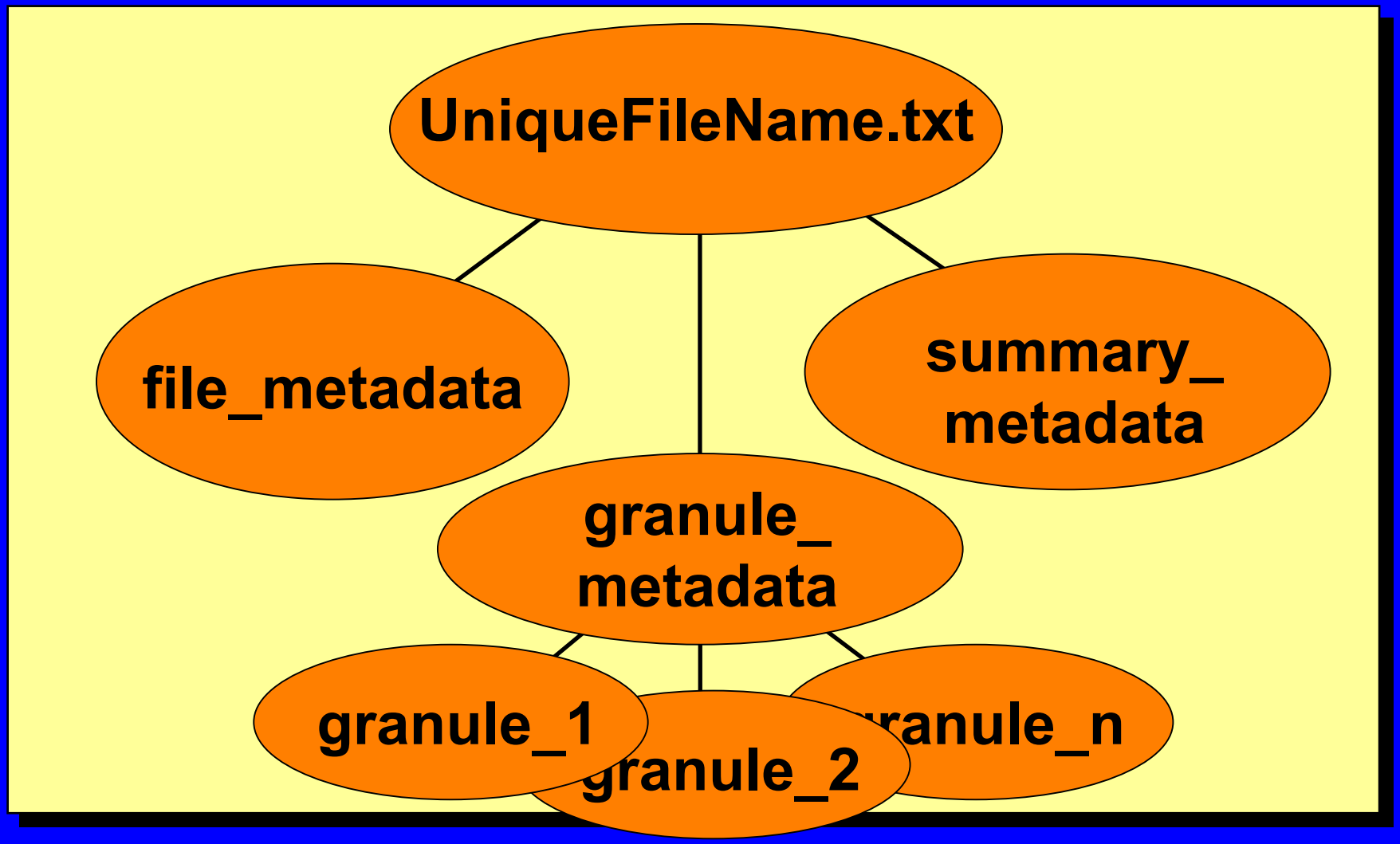
Why HDF5?

- Familiarity -- Environmental scientists already have experience with the standard, most recently from EOS products.
- Maturity -- HDF has shown its "staying power", and has been available long enough to have matured from user experiences. NASA, DOE, and others invested heavily in its development.
- Capability -- HDF was designed to manage large, compound data sets within high performance computing environments. HDF5 incorporates new features that are important for NPOESS.
- Compatibility -- HDF operates on multiple appropriate operating systems and languages: C, C++, Java and Fortran.

Why HDF5? (concluded)

- Availability -- HDF was developed in the public interest at NCSA, and is freely available. HDF also has many free, share and commercial supports tools available.
- Interoperability -- The DoD Joint Technical Architecture is in the process of accepting HDF as a standard for interoperability among DoD systems.
- Efficiency – Low overhead compared with legacy formats: GRIB and BUFR. Efficient indexing and subsetting. Support for data compression.

Notional Structure -- external metadata



Notional Structure -- HDF portion