Multi-angle
Imaging
Spectro-
Radiometer

Use of HDF and HDF-EOS in MISR Summary Data Products and Ancillary Datasets

Kathleen Crean

Jet Propulsion Laboratory,
California Institute of Technology

HDF-EOS Workshop IV

September 20, 2000

MISR product design: Summary ("QA") files

- Background
- File requirements
- File design choices
- Issues encountered

MISR ancillary dataset design: SMART dataset

- Background
- File requirements
- File design choices
- Issues encountered

MISR product design: Summary ("QA") files

Background

- Full-swath data files are unwieldy

  => for initial analysis, need smaller, more manageable files

- Want smaller files for quality assessment activities

- Want smaller files to aid in selecting datasets to order

- MISR solution:

  => ***create summary file for each product file***

Requirements:

- Consistent structure across all MISR files
- Flexibility to handle differences between MISR products
- Need a scheme which is flexible for future updates
- Need a structure which is end user-friendly

Design choices:

- What file type to use? text vs. HDF

  Complexity of reported quantities => HDF

- Which HDF data structures to use? SDS vs. vdata

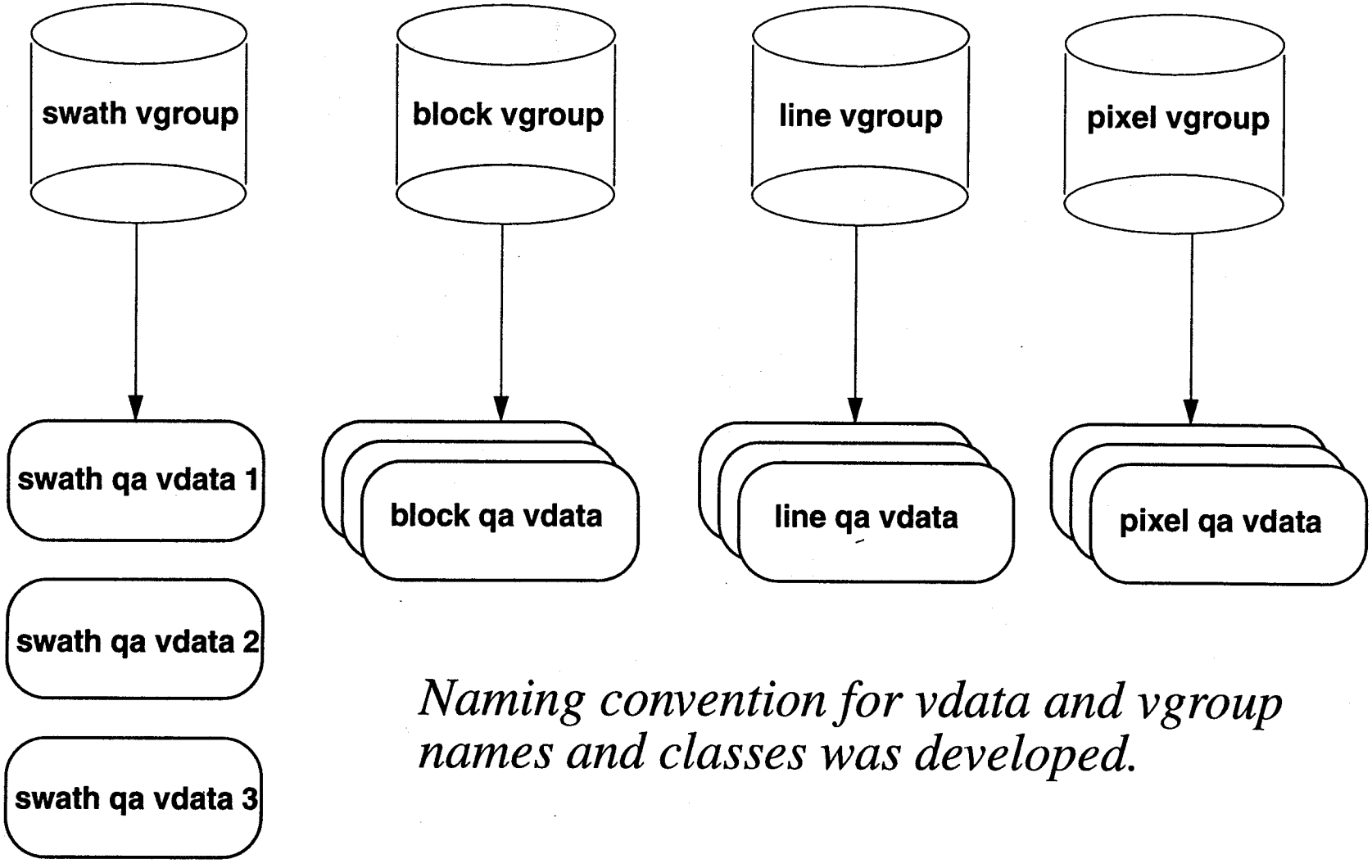
  Tabular nature of reported quantities (per-pixel, per-line, per-block) => vdata

  Large number of reported quantities => vdata

  Logical grouping of reported quantities => vgroup

- How to design MISR QA data structures?

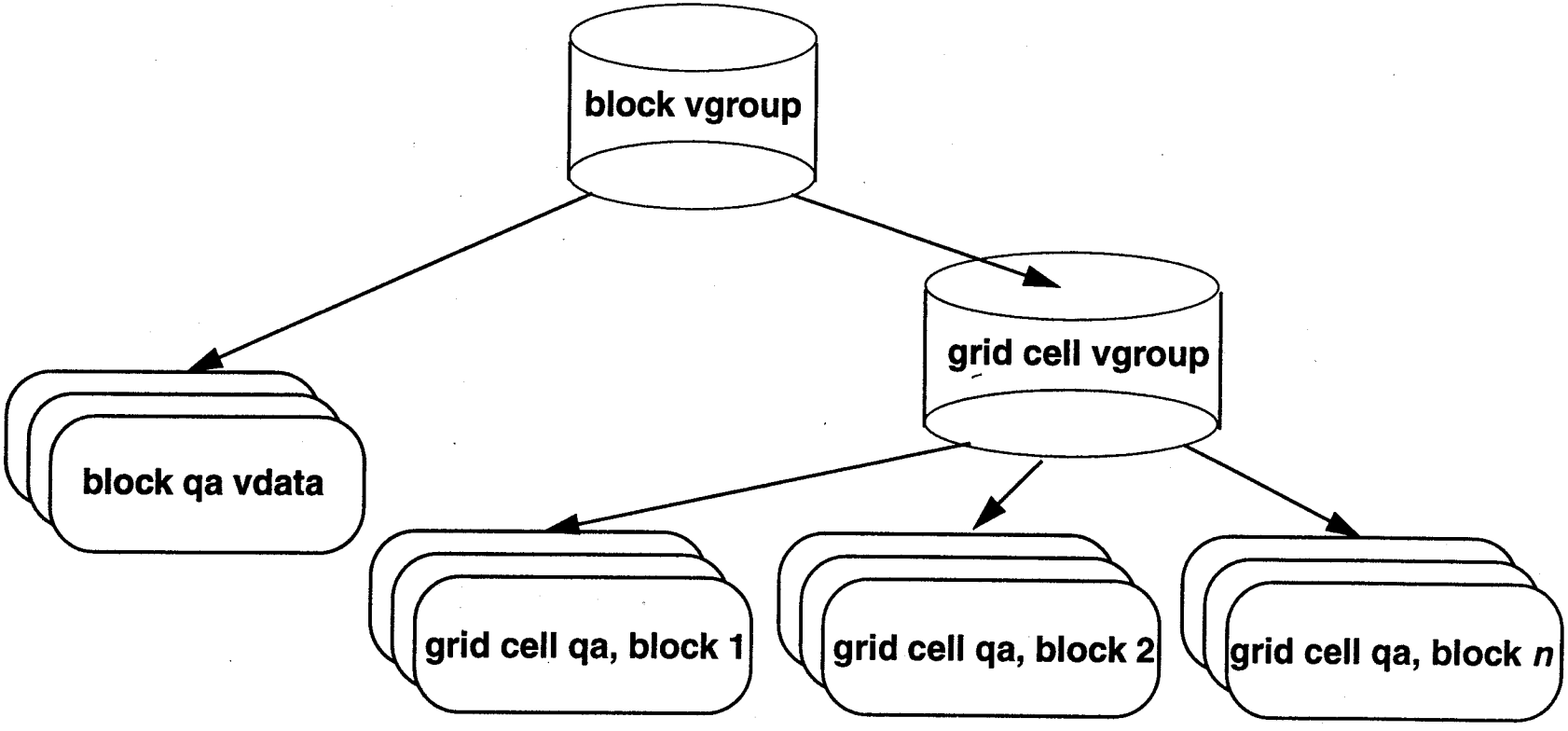
Product Design: Summary (QA) Files



Naming convention for vdata and vgroup names and classes was developed.

Custom QA data structures are also possible

example: "Grid Cell" data structure



Issues encountered:

- Interface with QA_DMS database imposed design decisions

  - Physical index of vdata record was assumed by MISR software; database needed explicit index

    Make physical index of record explicit => add vdata field 'BlockNumber'

  - Don't want database to pick up empty records that are automatically generated by HDF

Specify if a record is valid => add vdata field 'ValidRecord'

Issues encountered:

- Level 2 TOA/Cloud histogram implementation

  - Could not put all block-level histograms into one vdata, due to HDF-imposed limitation of 256 fields per vdata

    Redesign histograms to avoid this limitation => split histograms into 3 separate vdatas

MISR product design: SMART Dataset

Background

- SMART = *S*imulated
  *M*ISR
  *A*ncillary
  *R*adiative
  *T*ransfer

- Contains "look-up tables" of radiative transfer calculations
- Required input to create MISR Level 2 aerosol data products
- Large, complex dataset (9 parameters, each with different dimensions)

Requirements:

- Data must be stored within file size constraints
- Data must be accessed within timing constraints
- Data must be accessed within memory constraints
- Data must be readily understandable to a user

Design choices:

- What file type to use? binary vs. HDF

  Complexity of stored parameters => HDF

  Consistency with other MISR files => HDF

- Which HDF data structures to use?

  Array nature of stored parameters => SDS

Design choices:

- How to organize information into files?

File size constraints: too much data for certain stored parameters

=> store in scaled form to reduce file size
=> use file as a "dimension" for largest stored parameters

End user readability: separate into logical groupings
=> write each stored parameter to its own file

Design choices:

- How to design SDS structure of stored parameters?

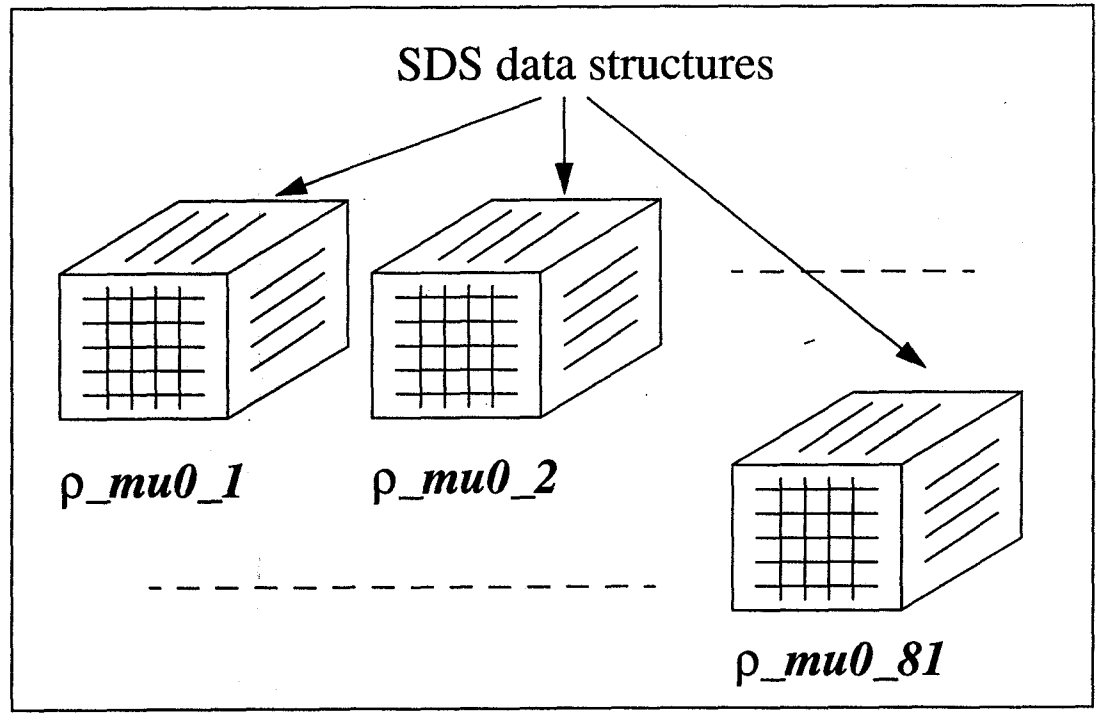
  Timing constraints: must access data to minimize I/O and data transformation operations

  Memory constraints: for certain stored parameters, can't read entire SDS into memory at one time

  => store "slowest" dimension as a separate SDS; all other dimensions are array dimensions within the SDS

File Structure for SMART Dataset File: "Rho_surf_ws1"

(equivalent reflectance at the surface of a dark water body, at given wind speed, wind speed #1)



Issues encountered:

- No problems encountered due to HDF
- Timing/memory issues drove the HDF file design
- Use of HDF features such as dimension scales and dimension names were useful in documenting file structure

References

- Scott Lewicki, Brian Chafin, Kathleen Crean, Scott Gluck, Kyle Miller and Susan Paradise, "**Data Products Specifications**", JPL Internal Document D-13963, Rev. C, December 14, 1999.
- Alec Shaner and Kyle Miller, "**QA File Structure for MISR PGEs**", MISR SDS Design File Memo #257-B, August 4, 1998.
- Brian Chafin, "**SMART Dataset HDF File Structure (Rev. A)**", MISR SDS Design File Memo #309-F, July 19, 1999.
- David Nelson, personal email communication regarding Level 2 TOA/Cloud QA file/QA DMS interface issues, August 14, 2000.