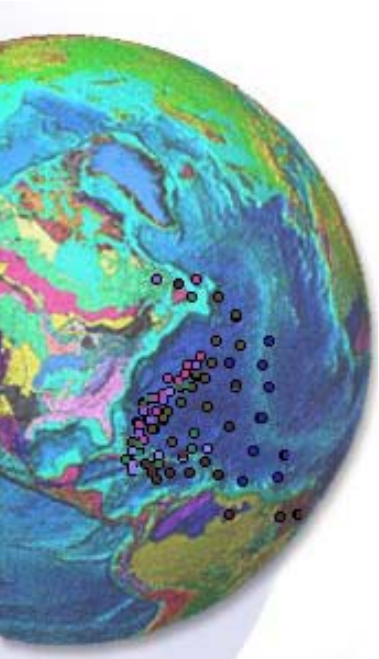


AND Archives: Freeing Ourselves From the "Tyranny of the OR"

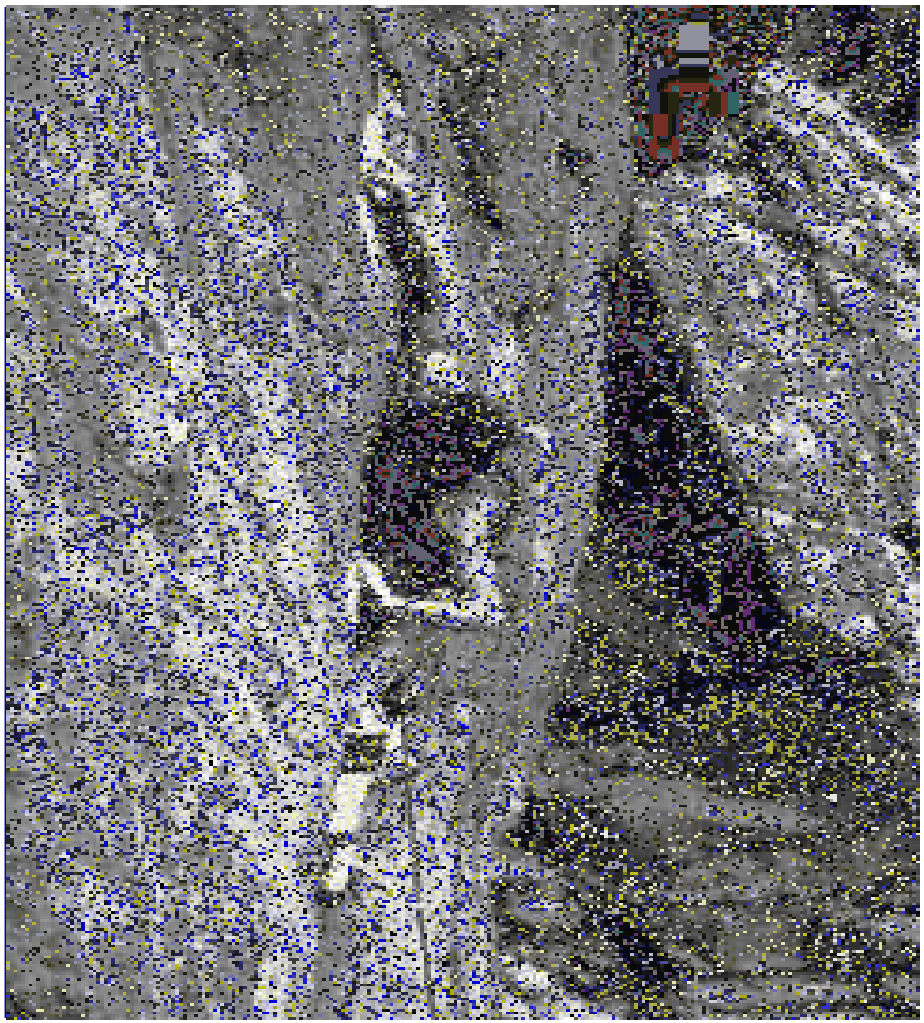
Ted Habermann
NOAA National Data Centers

NOAA Enterprise Geospatial Services

This presentation is designed to be viewed as a PPT slide show.

Built To Last

Jim Collins (famous Boulder climber did first free assent of Genesis) and Jerry Porras did a study of Visionary Companies: *premier institutions in their industries, widely admired by their peers and having a long track record of making a significant impact on the world around them. The key point is that a visionary company is an organization.*



Identified characteristics of visionary companies through comparisons with comparable companies. One characteristic was:

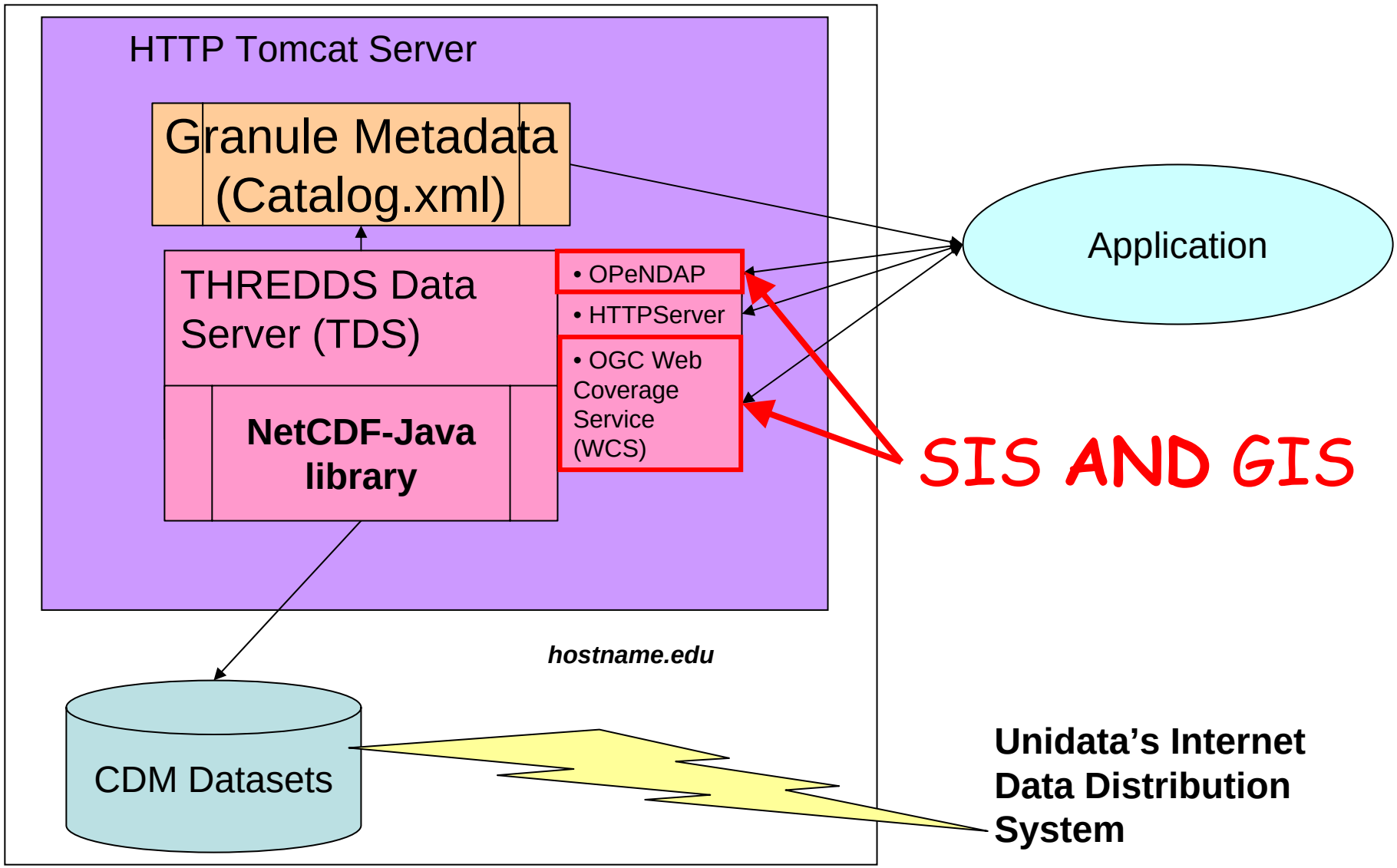
Avoid the "Tyranny of the OR" by embracing the "Genius of the AND".

Climb! The History of Rock Climbing in Colorado, Mountaineers Books, 2002.

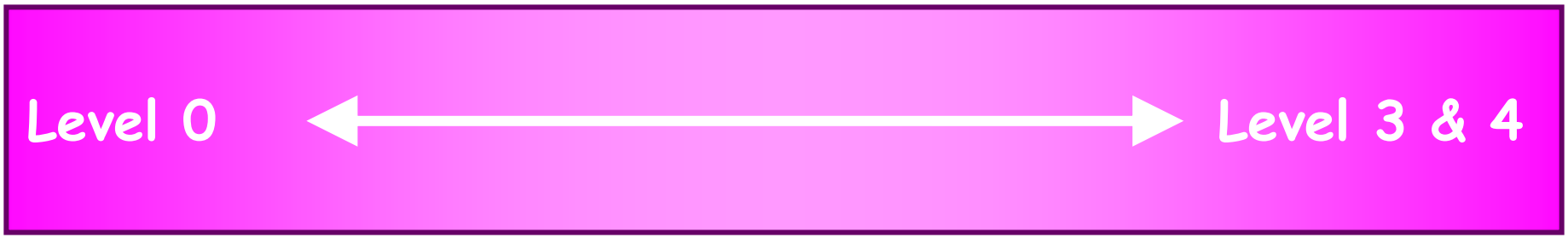
Tyranny of the OR

purpose beyond profit		pragmatic pursuit of profit
a relatively fixed core ideology		vigorous change and movement
conservatism around the core		bold, committing, risky moves
clear vision and sense of direction		opportunistic groping and experimentation
Big Hairy Audacious Goals		incremental evolutionary progress
selection of managers steeped in the core		selection of managers that induce change
ideological control	OR	operational autonomy
extremely tight culture (almost cult-like)		ability to change, move and adapt
investment for the long-term		demands for short-term performance
philosophical, visionary, futuristic		superb daily execution, “nuts and bolts”
organization aligned with a core ideology		organization adapted to its environment
science information systems		**geographic information systems**

THREDDS Data Server



Data Processing Levels



Telemetry information, Swaths

Time and Scan Angle

Complex custom formats (bits)

Large volume

Radiance in instrument units

Complex and Hard

Grids

Latitude & Longitude

Standard formats (bytes)

Small volume

Sea Surface Temp °C

Simple and Easy

POES Level 1b data

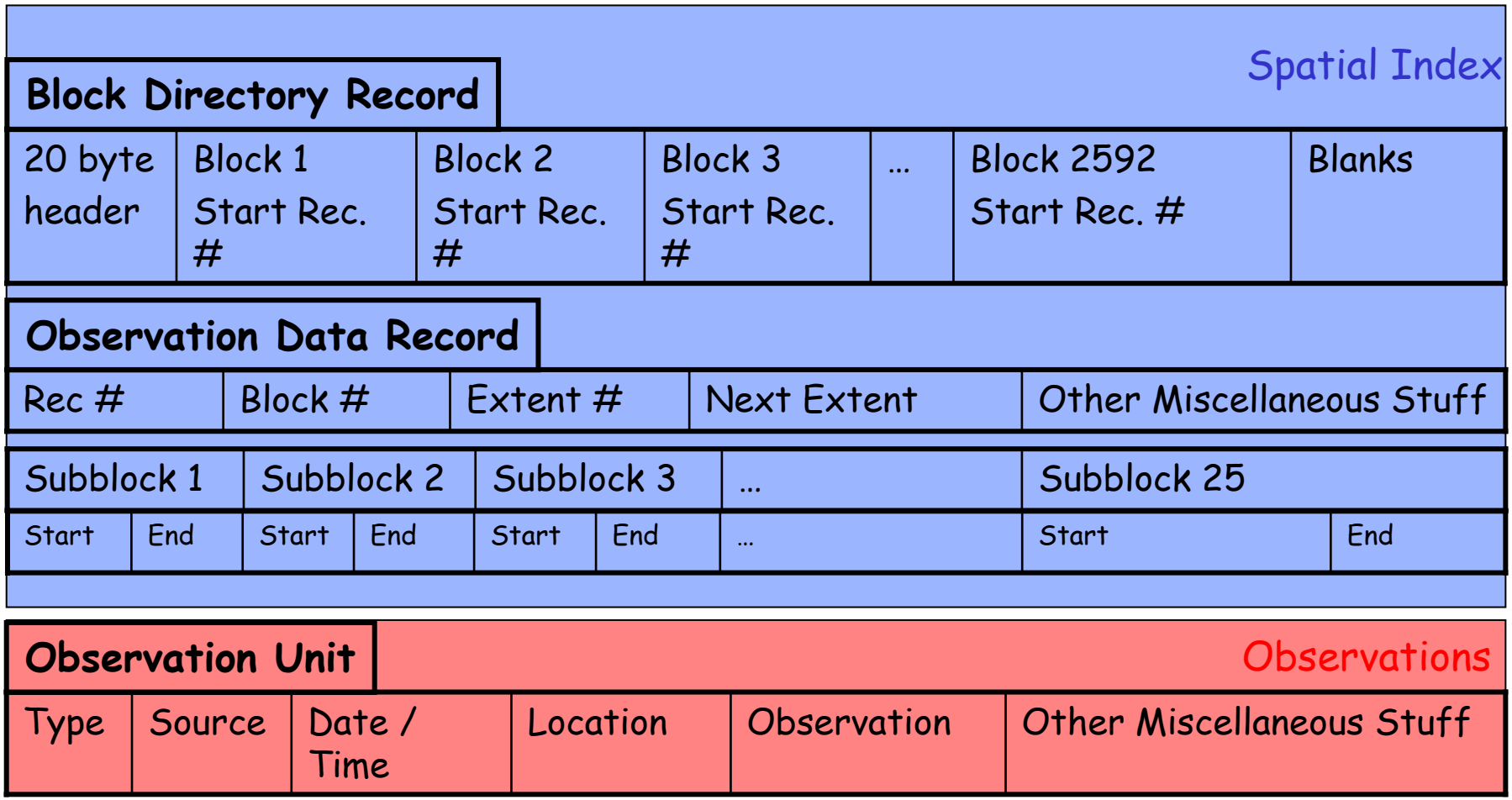
NESDIS Products: 14, 50, 100km grids produced daily/weekly

8km Level 2 SST

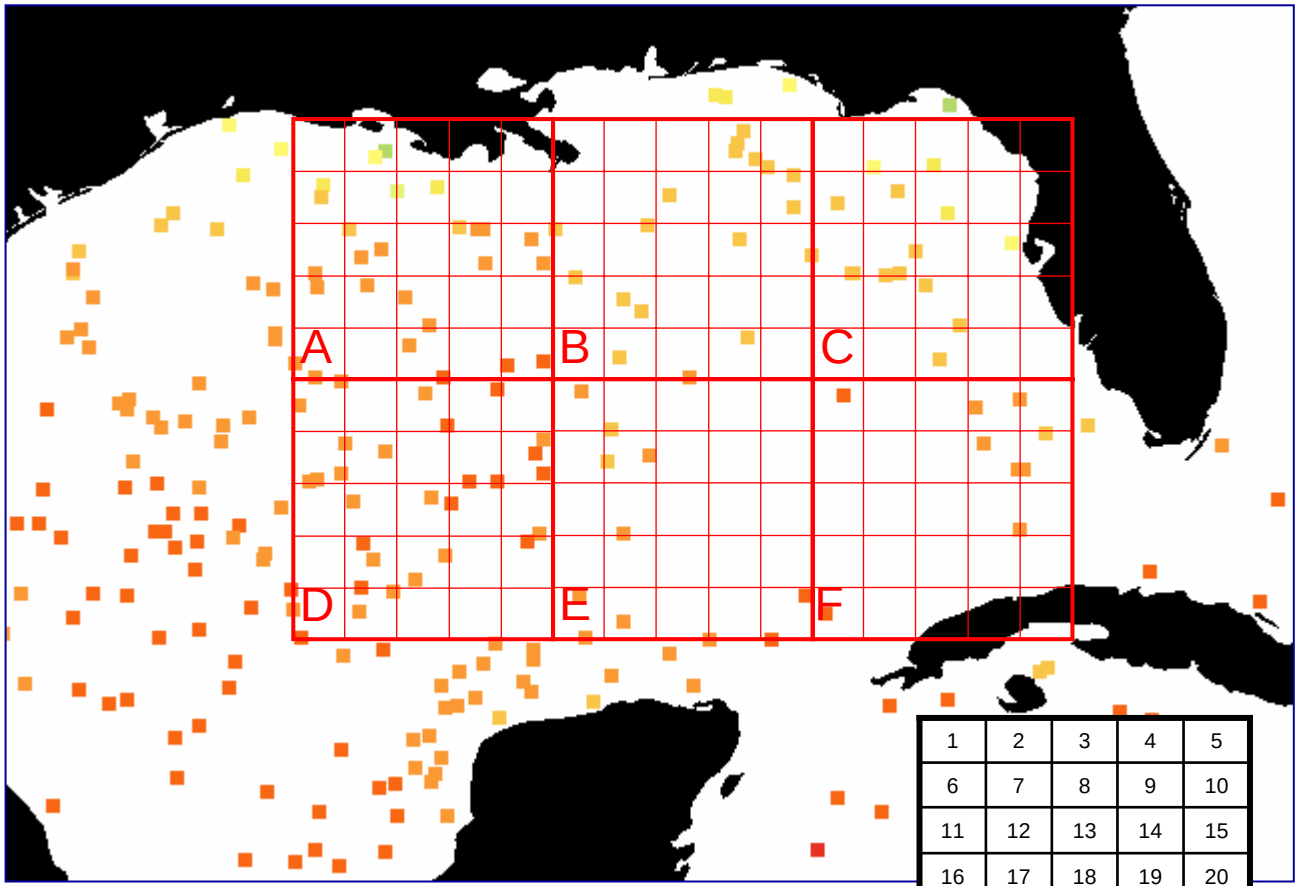
Most primitive useful form??

NESDIS Level 2 Observations

NESDIS (and Navy) Level 2 SST and Aerosol Observations are available via phone call / FTP arrangements with NCDC at present. These observations are in a custom format designed during the 1970's. The format has three major components: 5X5 spatial index, 1X1 spatial index, and the observations.



Spatial Sorting and Indexing Point Data



1	2	3	4	5
6	7	8	9	10
11	12	13	14	15
16	17	18	19	20
21	22	23	24	25

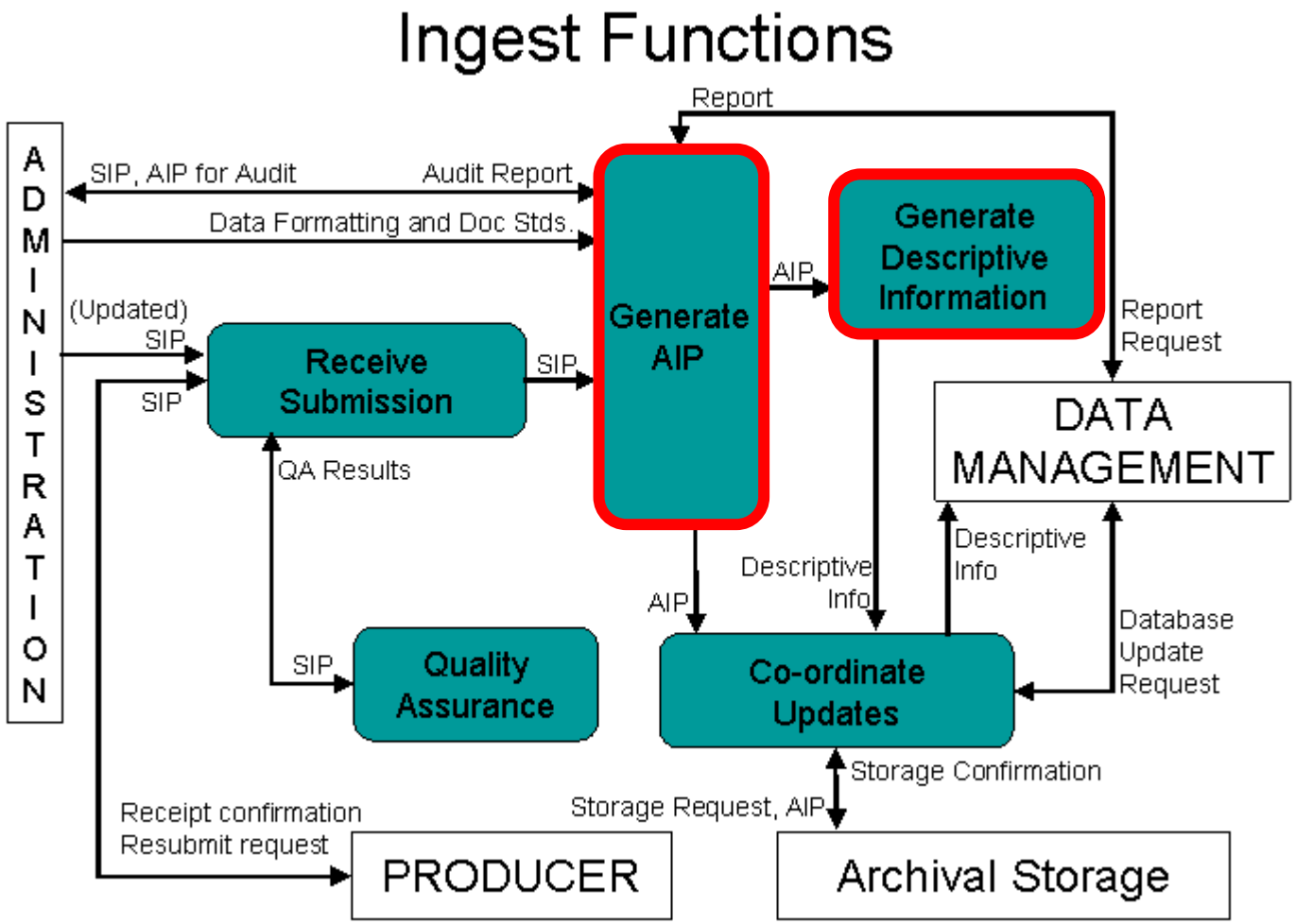
Sub-block Numbering

- Block Directory
- Block A
  - Sub-block 1
    - No Data
  - Sub-block 2
    - 2 Observations
  - Sub-block 6
    - 2 Observations
  - Sub-block 7
    - 1 Observations
  - …
- Block B
  - Sub-block 1-3
    - No Observations
  - Sub-block 4
    - 4 Observations
  - Sub-block 5
    - 1 Observations
  - …
- Block C
- Block D
- Next block …

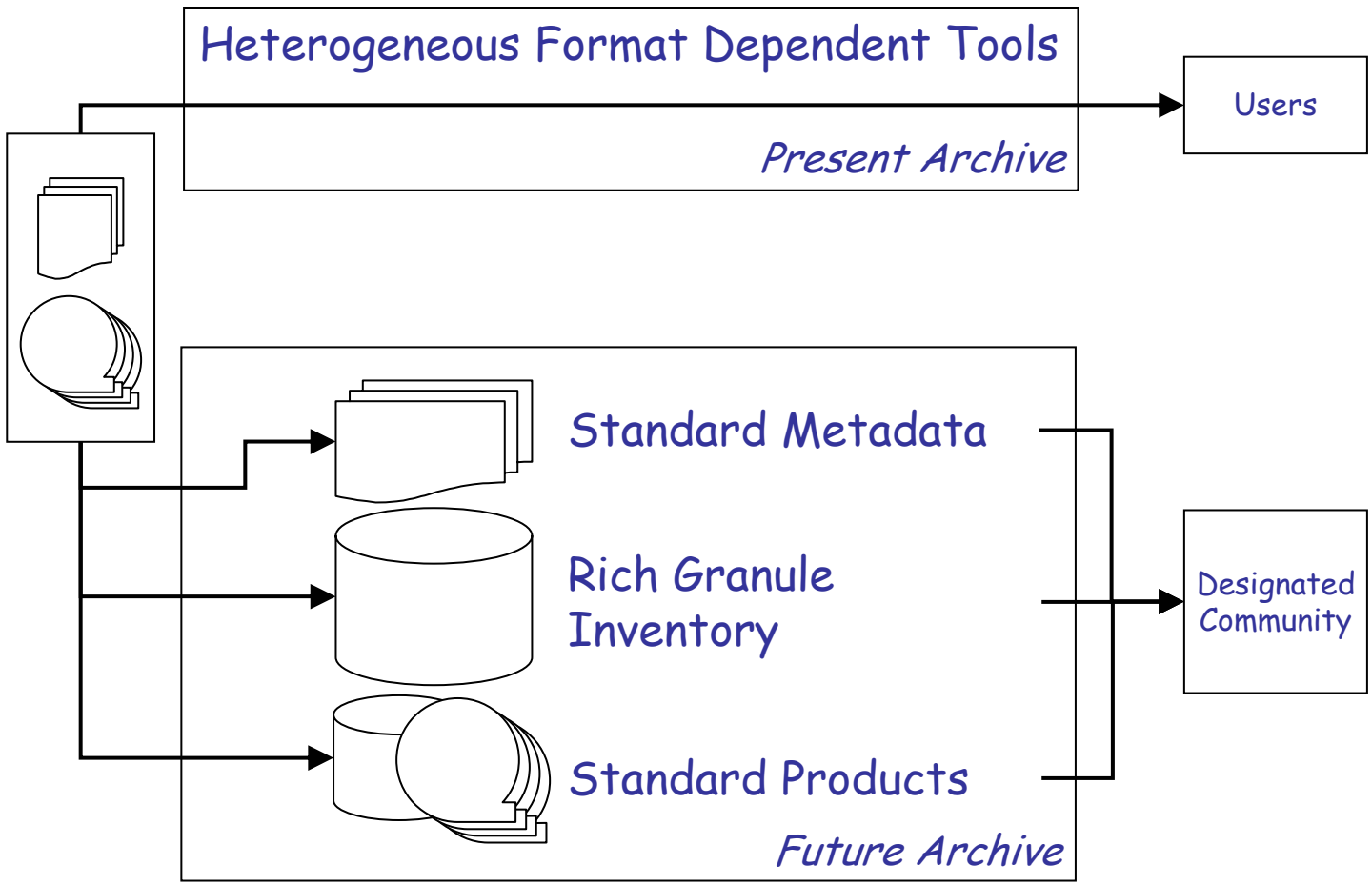
Satellite Data as points: Andy Pursch, Scott Shipley and someone @ NESDIS

Over the last decade commercial databases have developed the built-in capability to do this kind of spatial indexing. They bring many other capabilities to the table as well.

OAIS Ingest Functions



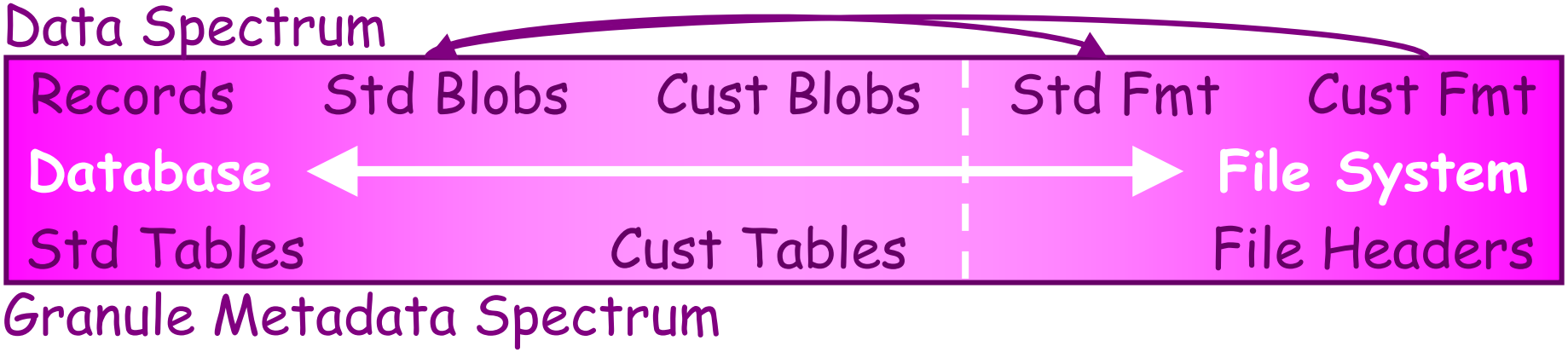
Archive Process Evolution

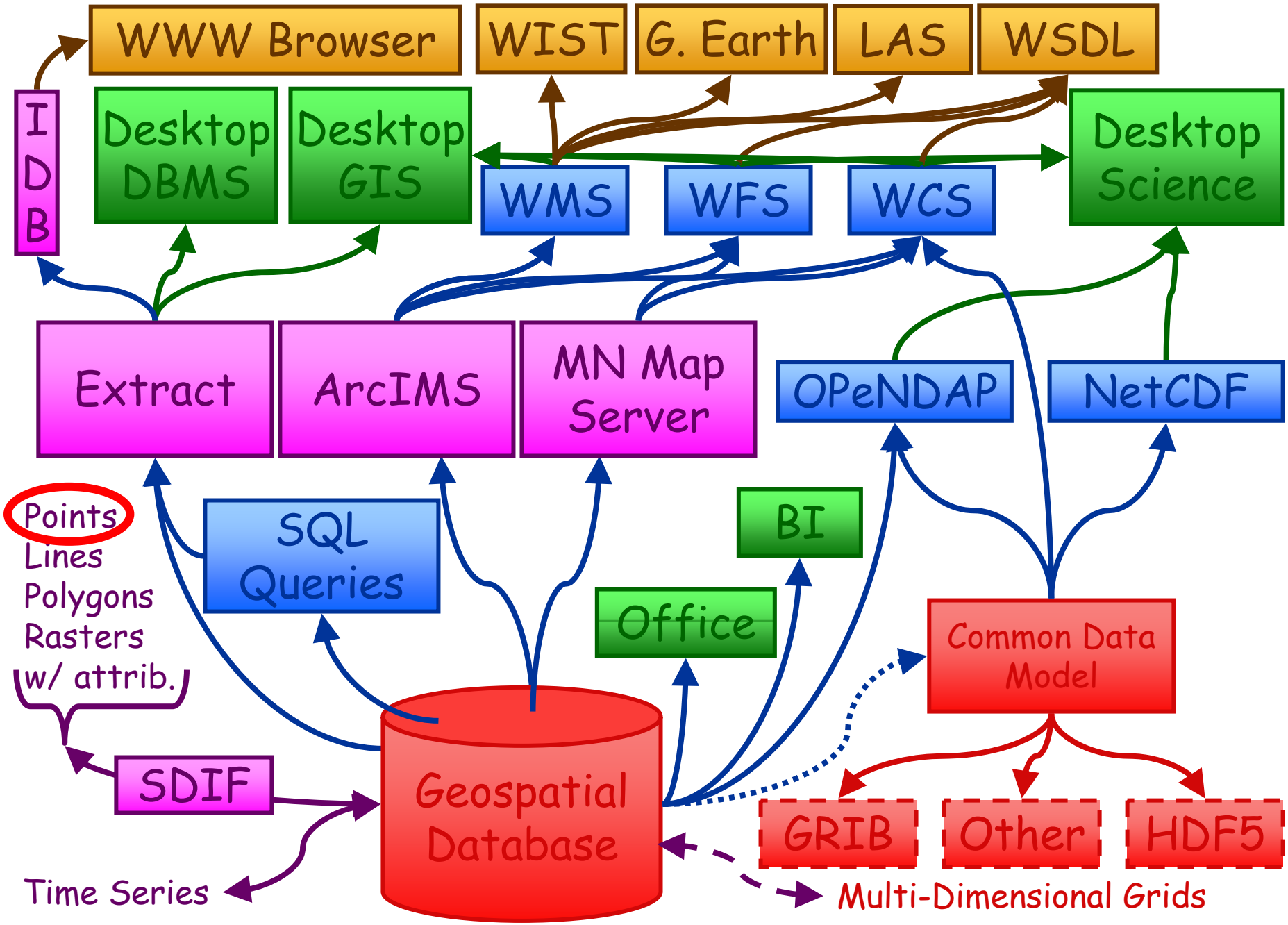


Homogeneous Data and Metadata, Standard Tools

Step 1: Migrate the observations from a custom file format into a standard spatial database.

Step 2: Output a standard file format from the database.

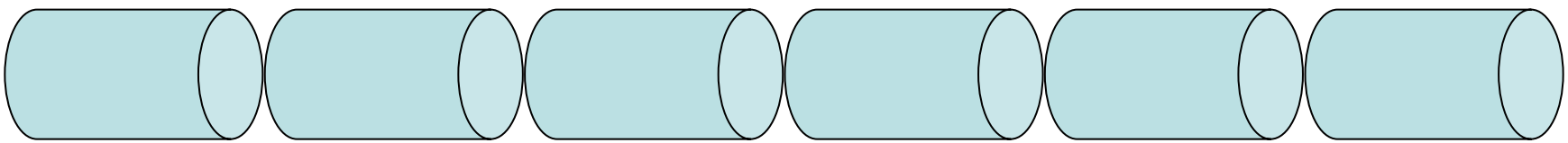



WWW Browser
WIST
G. Earth
LAS
WSDL
IDB
Desktop DBMS
Desktop GIS
WMS
WFS
WCS
Desktop Science
Extract
ArcIMS
MN Map Server
OPeNDAP
NetCDF
Points
Lines
Polygons
Rasters
w/ attrib.
SQL Queries
BI
Office
Common Data Model
SDIF
Geospatial Database
GRIB
Other
HDF5
Time Series
Multi-Dimensional Grids

Processing Pipeline

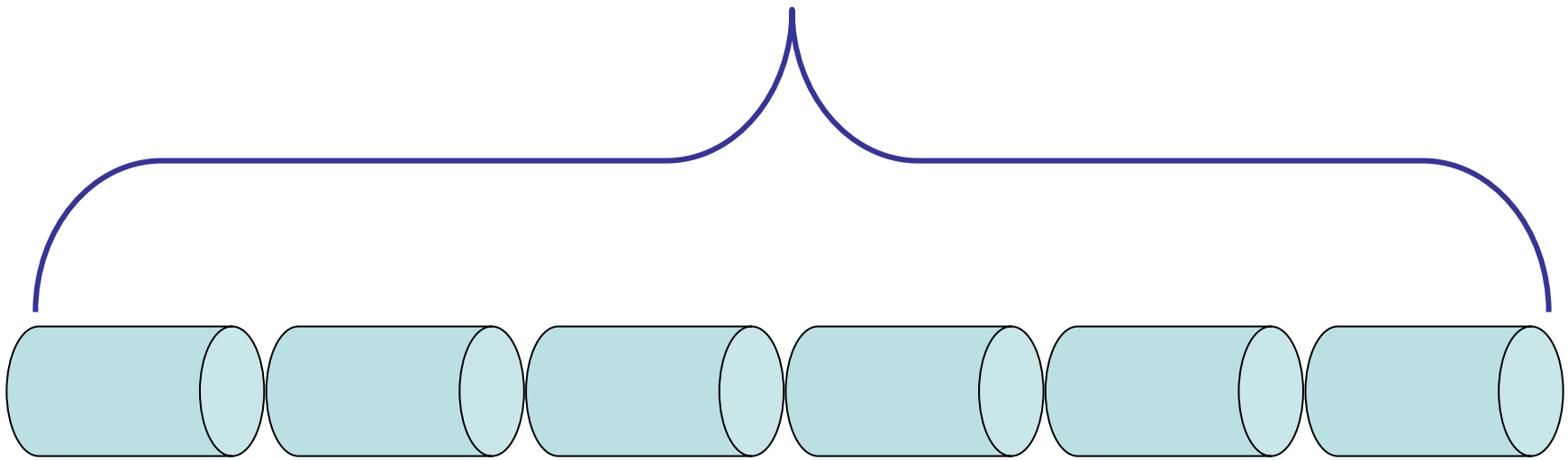
A pipeline provides a description of a sequence of data processing tasks. The NGDC data processing pipeline provides a set of pipeline utilities designed around work queues that run in parallel to sequentially process data objects. The pipeline is an open source project hosted in the **Jakarta Commons Sandbox** (http://jakarta.apache.org/commons/sandbox/pipeline/).

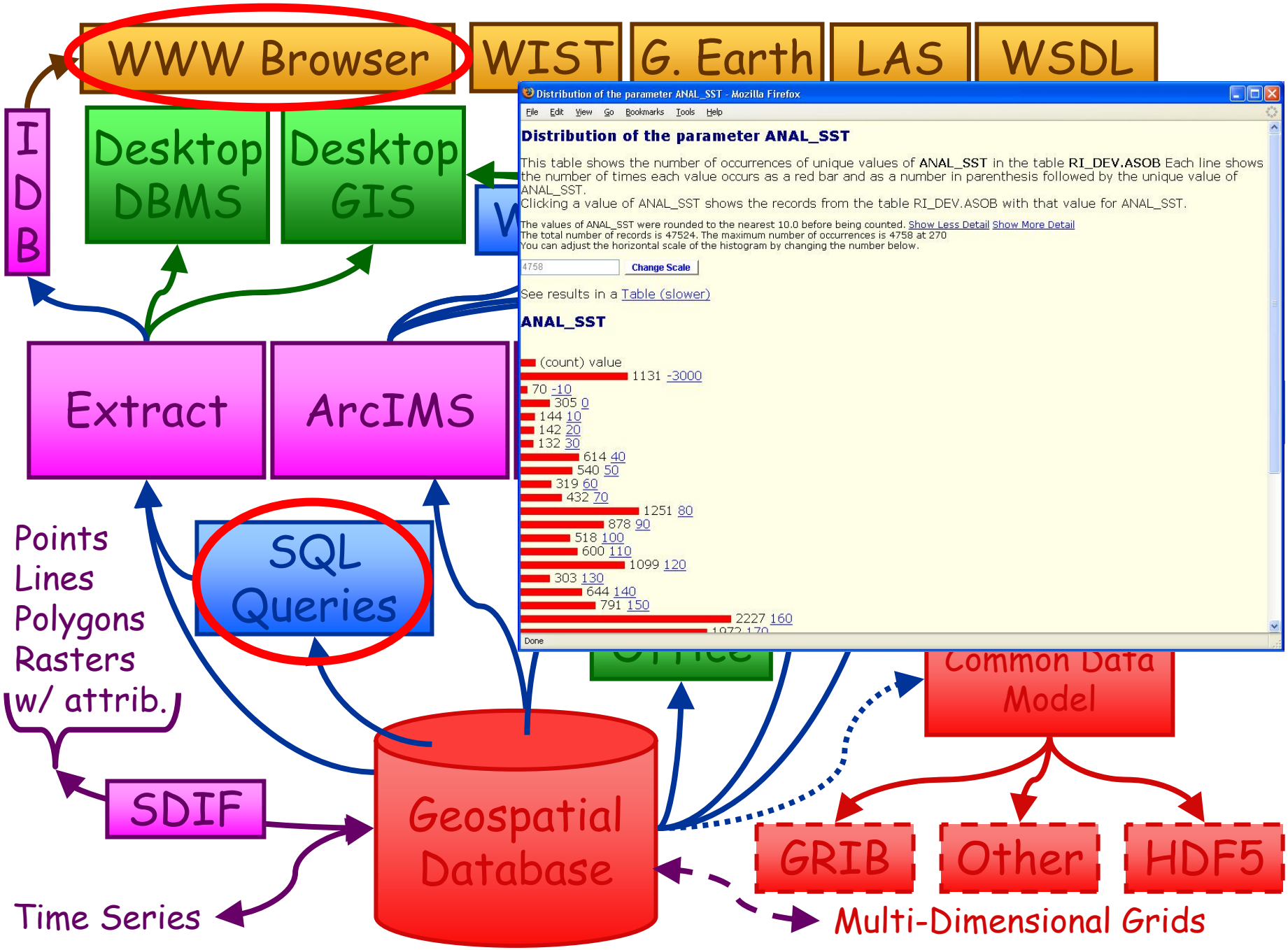
Processing steps are specified as a series of stages in an XML configuration file.

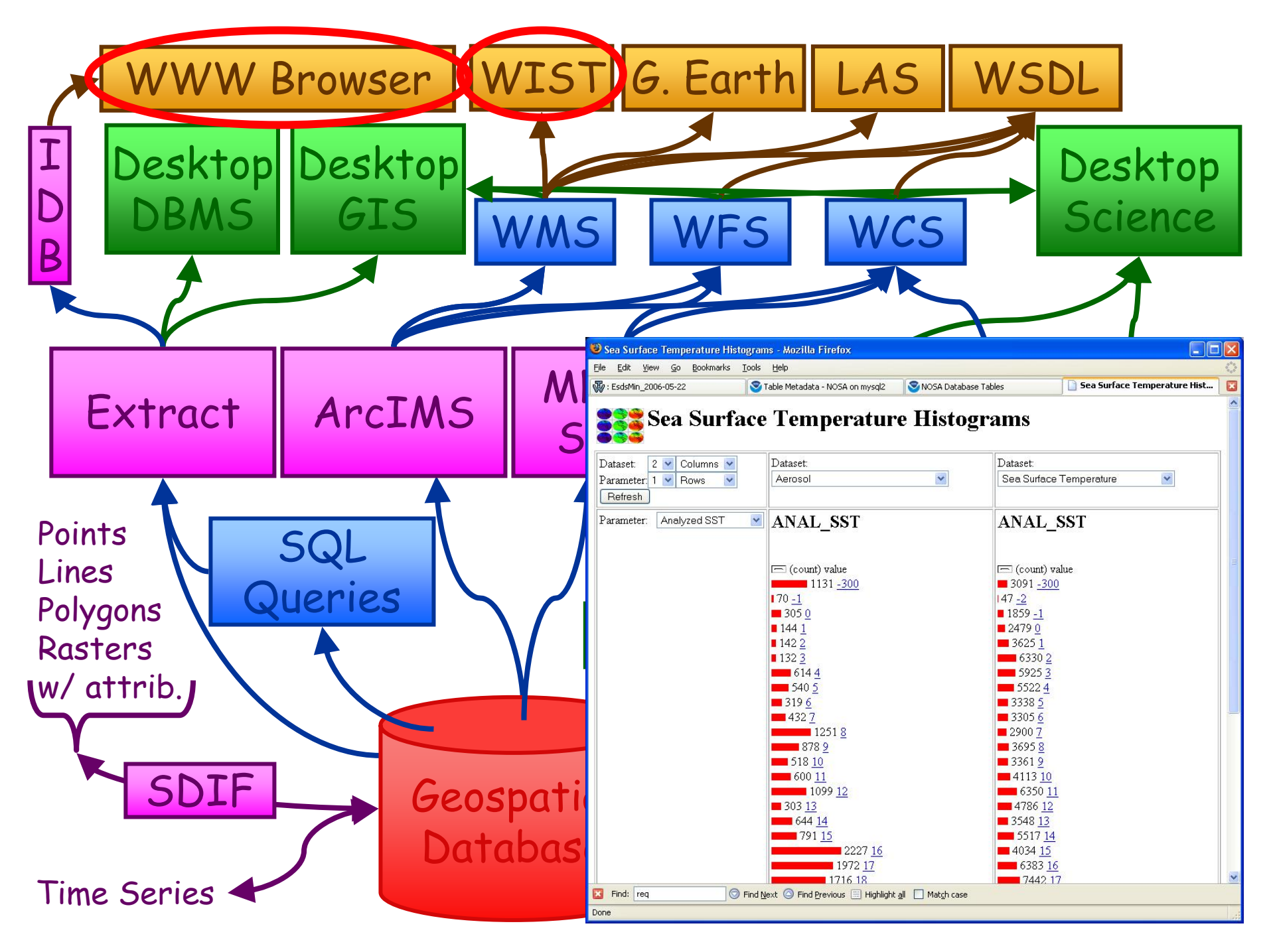


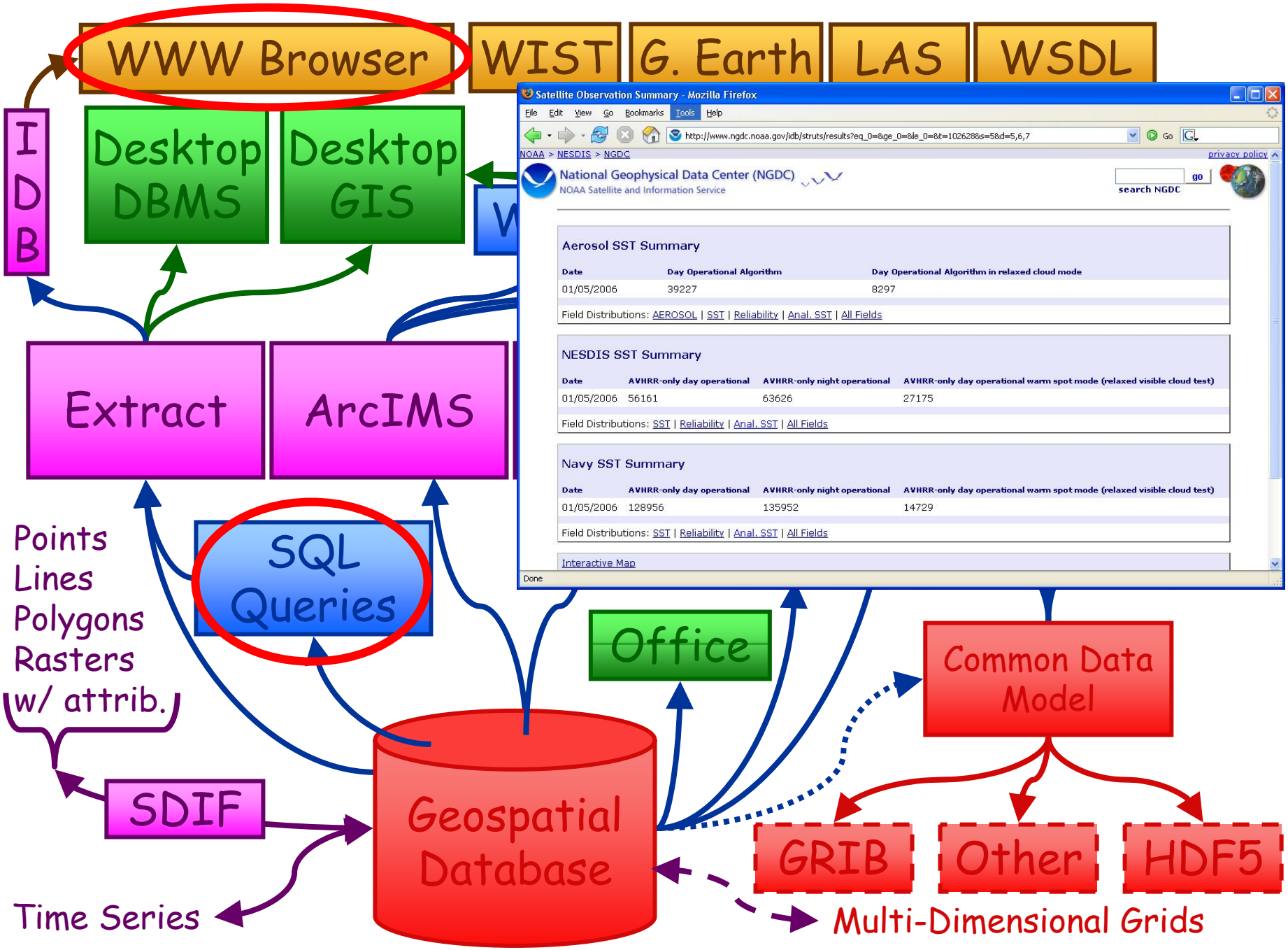
SST Ingest Processing

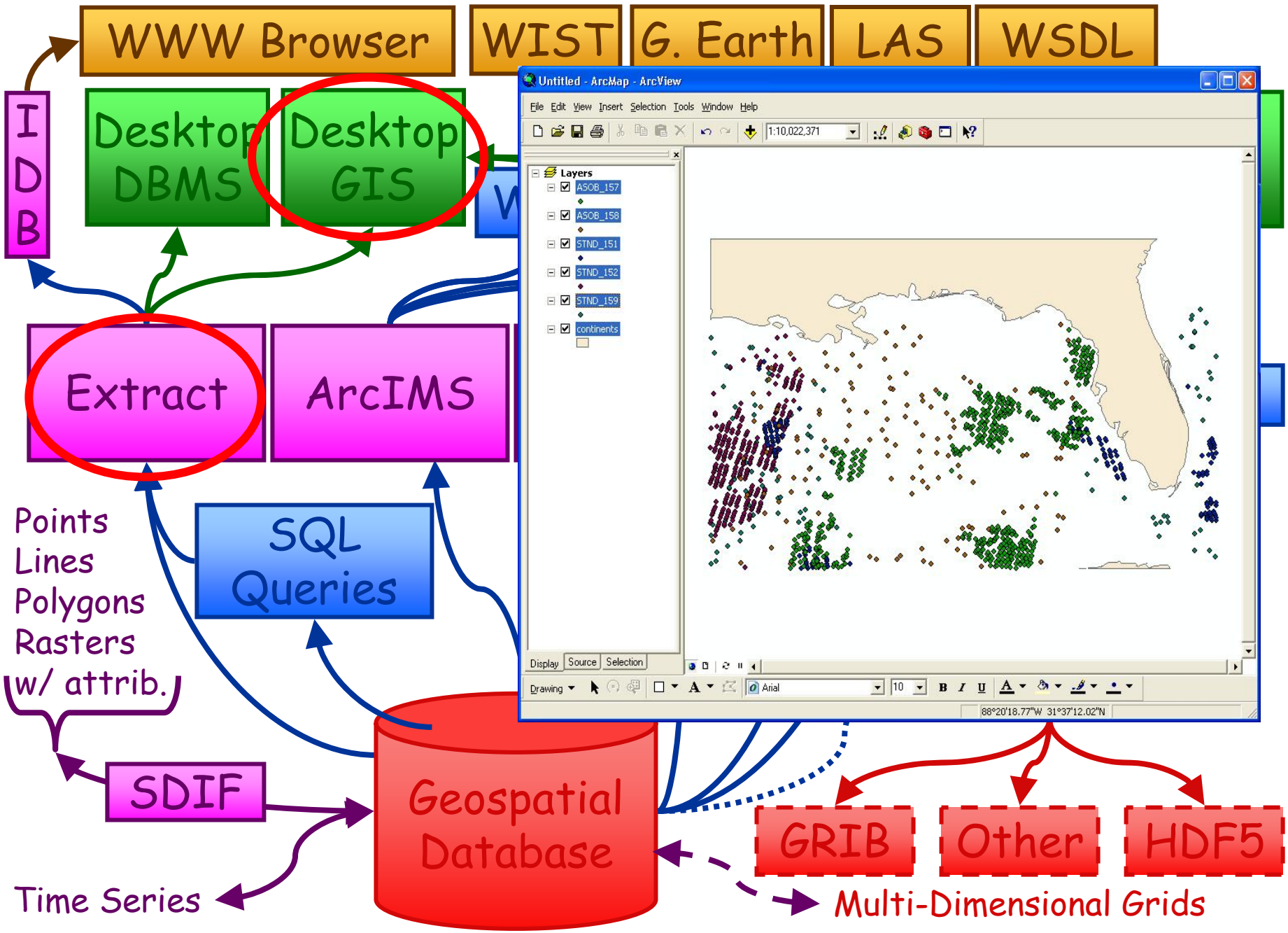
Stage 1. Find Matching Files
Stage 2. Avoid Duplicate Processing
Stage 3. Read Data / Create Spatial Objects
Stage 4. Write Thinned Layer (10%) to DB & CDM
Stage 5. Write Complete Layer to DB & CDM
Stage 6. Create Summary (Grid) Table to DB & CDM
Stage 7. Create Rich Inventory Record

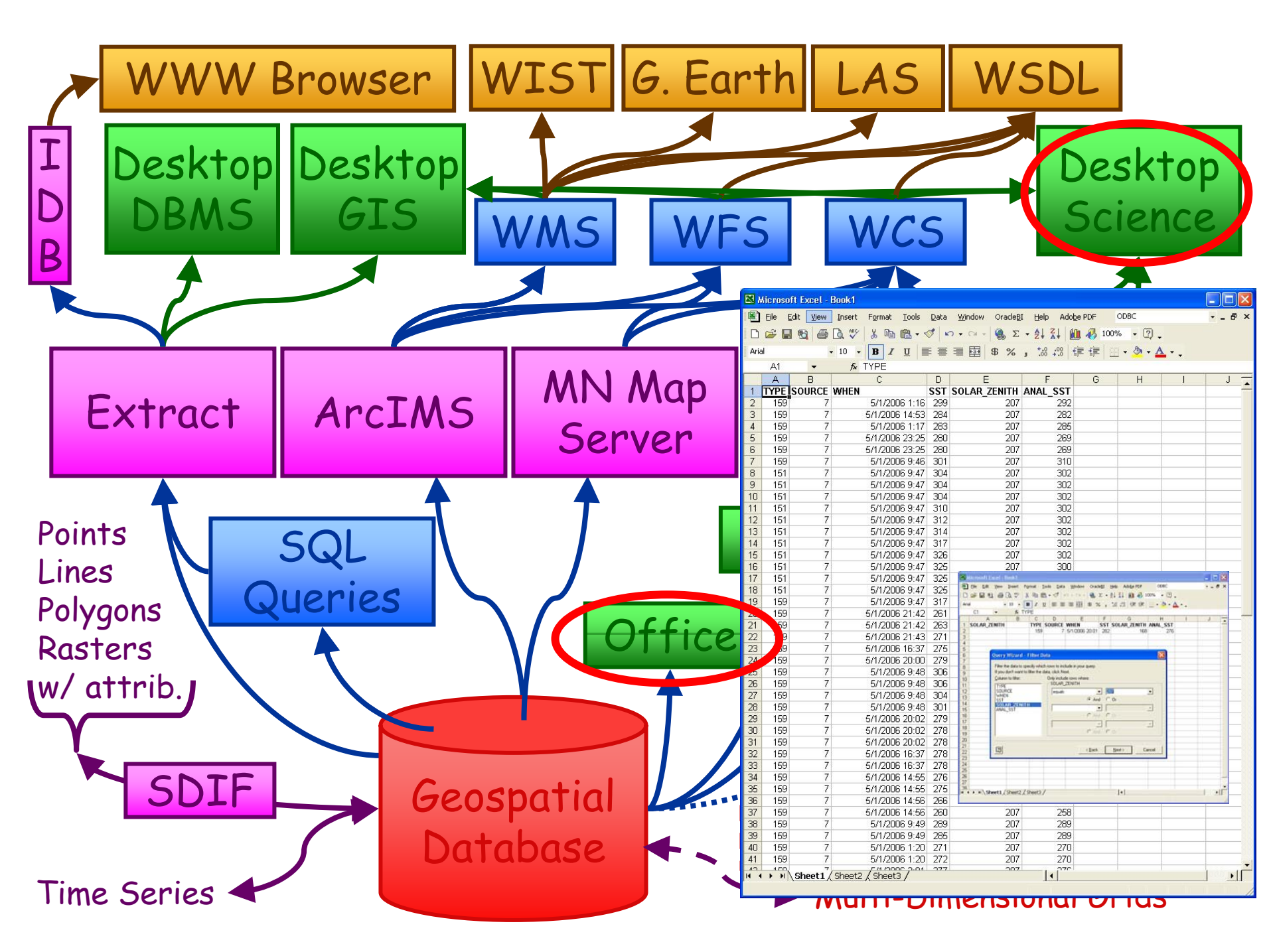


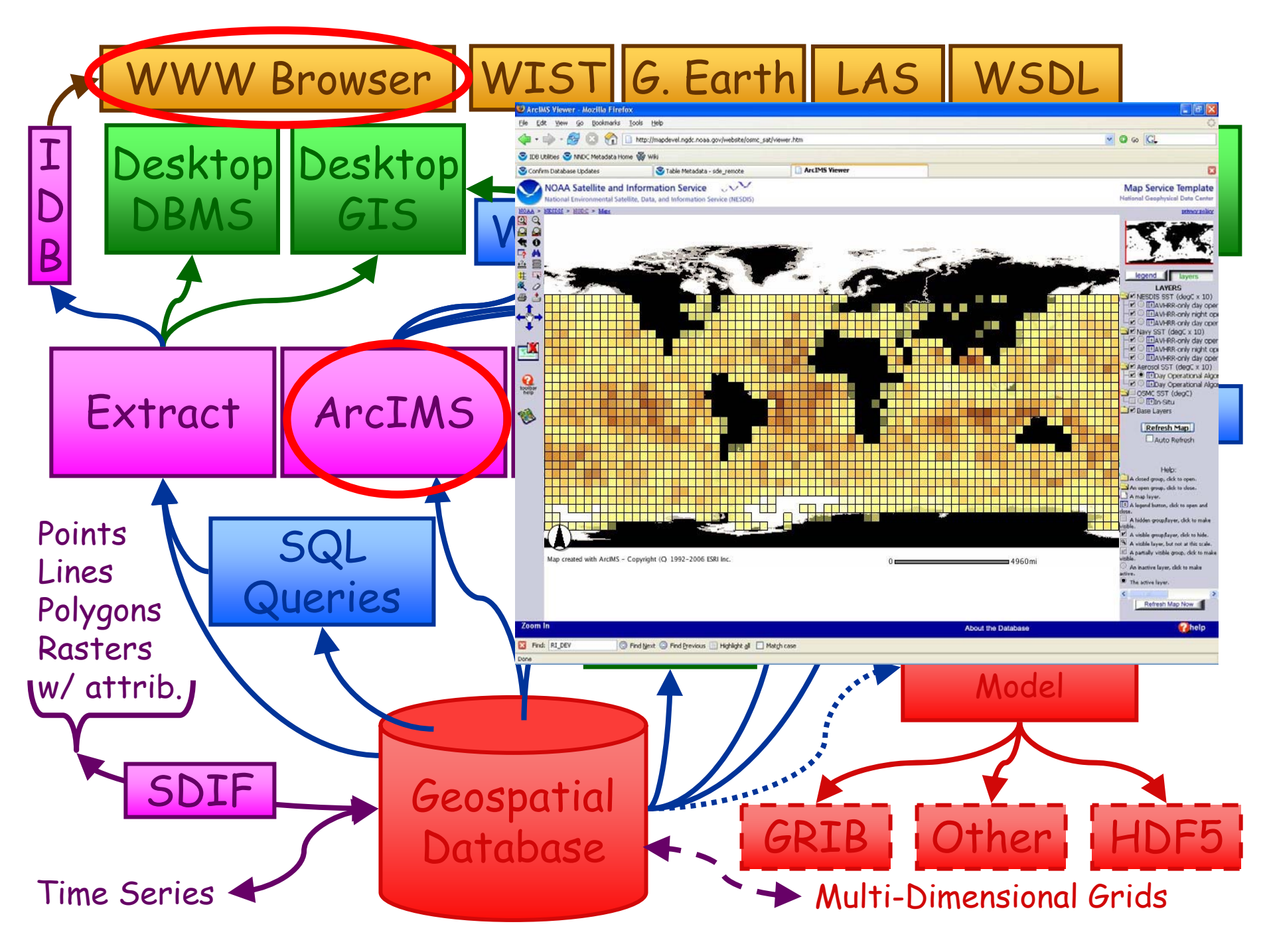

WWW Browser
WIST
G. Earth
LAS
WSDL
IDB
Desktop DBMS
Desktop GIS
Distribution of the parameter ANAL_SST - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
Distribution of the parameter ANAL_SST
This table shows the number of occurrences of unique values of ANAL_SST in the table RI_DEV.ASOB Each line shows the number of times each value occurs as a red bar and as a number in parenthesis followed by the unique value of ANAL_SST.
Clicking a value of ANAL_SST shows the records from the table RI_DEV.ASOB with that value for ANAL_SST.
The values of ANAL_SST were rounded to the nearest 10.0 before being counted. Show Less Detail Show More Detail
The total number of records is 47524. The maximum number of occurrences is 4758 at 270
You can adjust the horizontal scale of the histogram by changing the number below.
4758
Change Scale
See results in a Table (slower)
ANAL_SST
(count) value
1131 -3000
70 -10
305 0
144 10
142 20
132 30
614 40
540 50
319 60
432 70
1251 80
878 90
518 100
600 110
1099 120
303 130
644 140
791 150
2227 160
Done
Extract
ArcIMS
SQL Queries
Points
Lines
Polygons
Rasters
w/ attrib.
SDIF
Geospatial Database
Time Series
Common Data Model
GRIB
Other
HDF5
Multi-Dimensional Grids

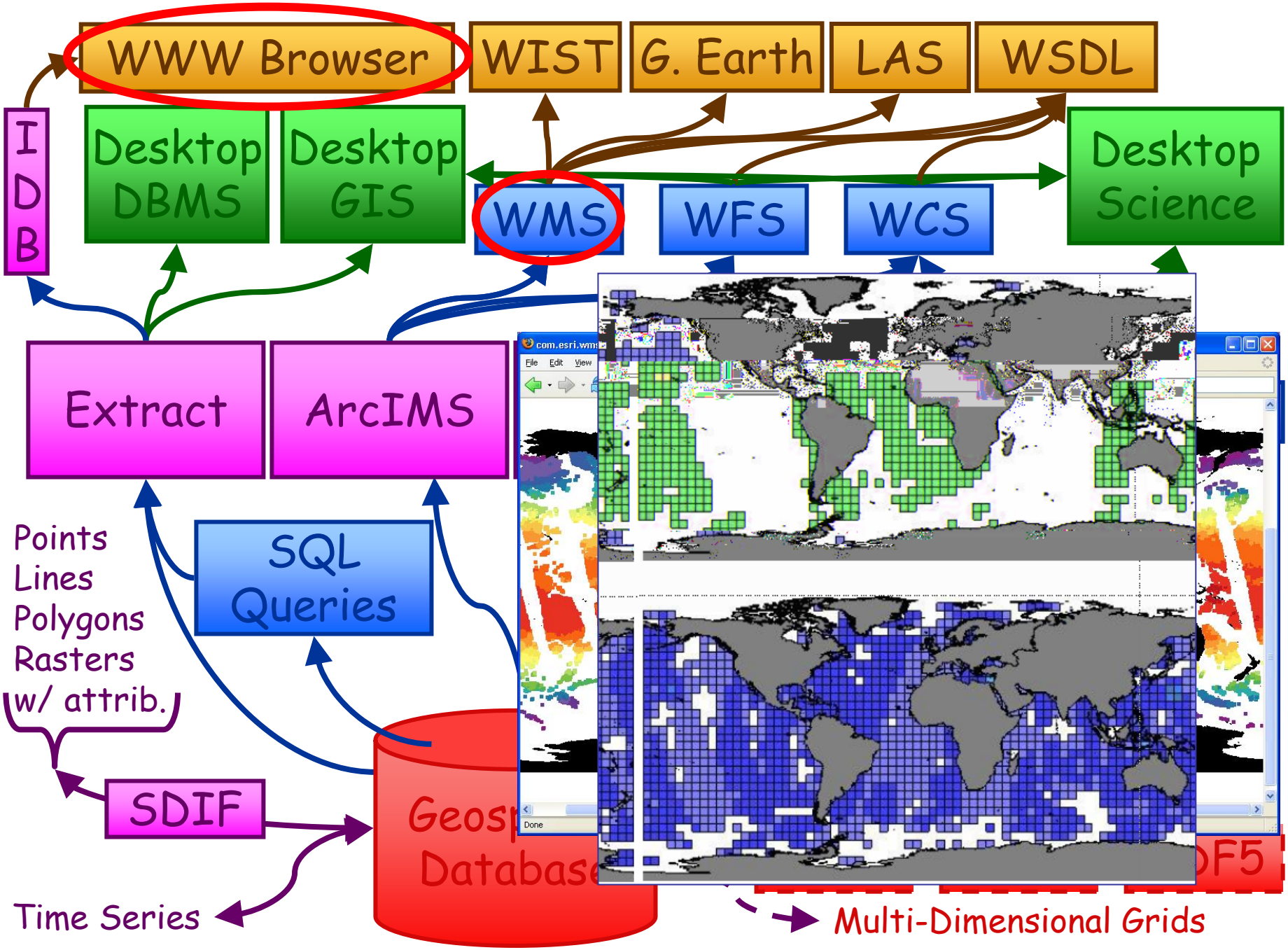

WWW Browser
WIST
G. Earth
LAS
WSDL
IDB
Desktop DBMS
Desktop GIS
WMS
WFS
WCS
Desktop Science
Extract
ArcIMS
SQL Queries
Points
Lines
Polygons
Rasters
w/ attrib.
SDIF
Time Series
Sea Surface Temperature Histograms - Mozilla Firefox
Sea Surface Temperature Histograms
Refresh
Aerosol
Sea Surface Temperature
Analyzed SST
ANAL_SST
ANAL_SST
(count) value
(count) value
Find:
Done

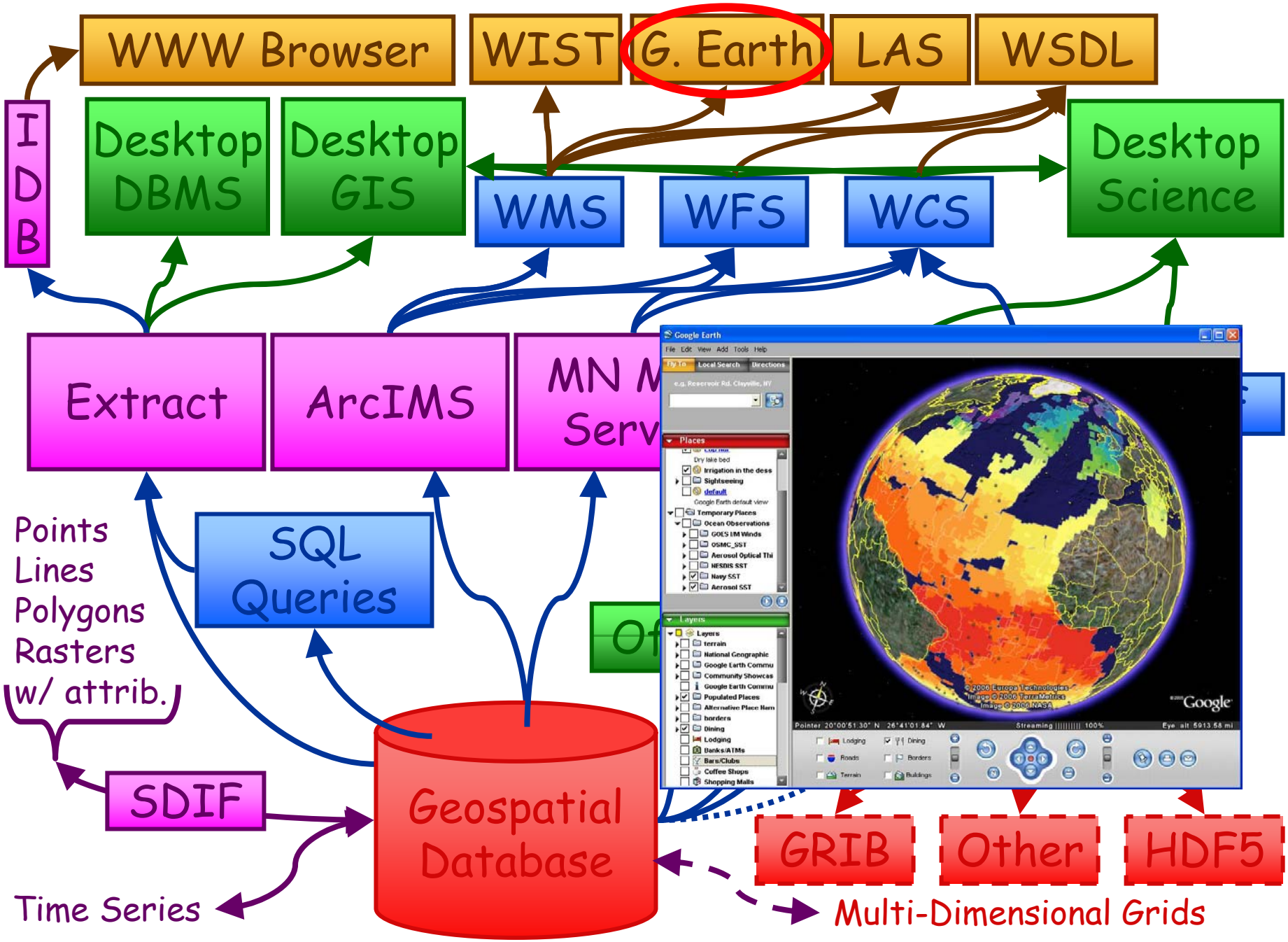

WWW Browser
WIST
G. Earth
LAS
WSDL
IDB
Desktop DBMS
Desktop GIS
Extract
ArcIMS
SQL Queries
Points
Lines
Polygons
Rasters
w/ attrib.
SDIF
Time Series
Office
Geospatial Database
Common Data Model
GRIB
Other
HDF5
Multi-Dimensional Grids
Satellite Observation Summary - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
http://www.ngdc.noaa.gov/idb/struts/results?eq_0=&ge_0=&le_0=&t=1026288s=5&d=5,6,7
NOAA > NESDIS > NGDC
privacy policy
National Geophysical Data Center (NGDC)
NOAA Satellite and Information Service
go
search NGDC
Aerosol SST Summary
Date
Day Operational Algorithm
Day Operational Algorithm in relaxed cloud mode
01/05/2006
39227
8297
Field Distributions: AEROSOL | SST | Reliability | Anal. SST | All Fields
NESDIS SST Summary
Date
AVHRR-only day operational
AVHRR-only night operational
AVHRR-only day operational warm spot mode (relaxed visible cloud test)
01/05/2006
56161
63626
27175
Field Distributions: SST | Reliability | Anal. SST | All Fields
Navy SST Summary
Date
AVHRR-only day operational
AVHRR-only night operational
AVHRR-only day operational warm spot mode (relaxed visible cloud test)
01/05/2006
128956
135952
14729
Field Distributions: SST | Reliability | Anal. SST | All Fields
Interactive Map
Done

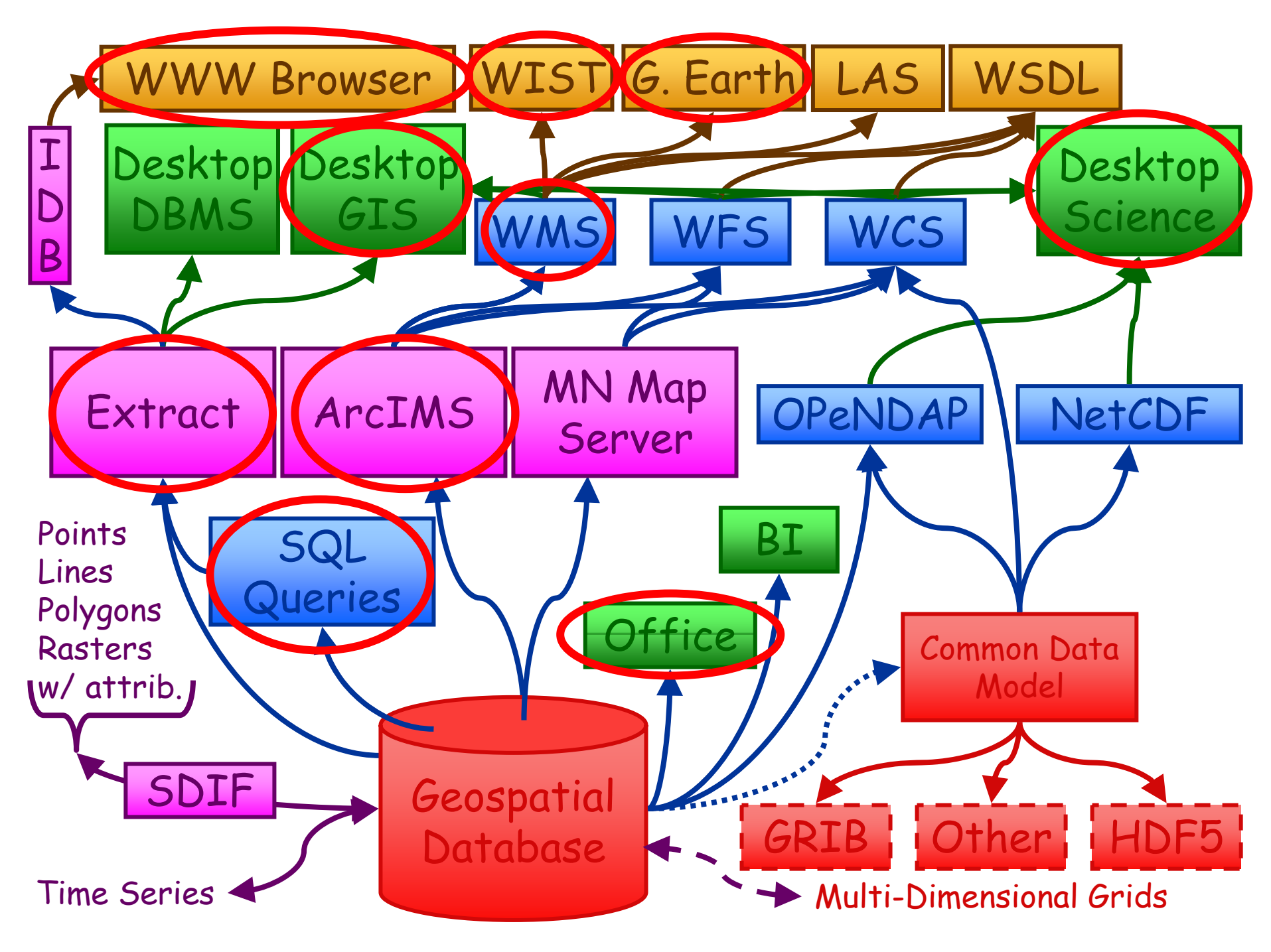

WWW Browser
WIST
G. Earth
LAS
WSDL
IDB
Desktop DBMS
Desktop GIS
Extract
ArcIMS
SQL Queries
Points
Lines
Polygons
Rasters
w/ attrib.
SDIF
Time Series
Geospatial Database
GRIB
Other
HDF5
Multi-Dimensional Grids
Untitled - ArcMap - ArcView
File Edit View Insert Selection Tools Window Help
Layers
ASOB_157
ASOB_158
STND_151
STND_152
STND_159
continents
Display
Source
Selection
Drawing
Arial
88°20'18.77"W 31°37'12.02"N


WWW Browser
WIST
G. Earth
LAS
WSDL
IDB
Desktop DBMS
Desktop GIS
WMS
WFS
WCS
Desktop Science
Extract
ArcIMS
MN Map Server
Points
Lines
Polygons
Rasters
w/ attrib.
SQL Queries
Office
SDIF
Geospatial Database
Time Series
Multi-Dimensional Grids
Microsoft Excel - Book1
File Edit View Insert Format Tools Data Window OracleBI Help Adobe PDF ODBC
Arial 10
A1 TYPE
TYPE SOURCE WHEN SST SOLAR_ZENITH ANAL_SST
159 7 5/1/2006 1:16 299 207 292
159 7 5/1/2006 14:53 284 207 282
159 7 5/1/2006 1:17 283 207 285
159 7 5/1/2006 23:25 280 207 269
159 7 5/1/2006 23:25 280 207 269
159 7 5/1/2006 9:46 301 207 310
151 7 5/1/2006 9:47 304 207 302
151 7 5/1/2006 9:47 304 207 302
151 7 5/1/2006 9:47 304 207 302
151 7 5/1/2006 9:47 310 207 302
151 7 5/1/2006 9:47 312 207 302
151 7 5/1/2006 9:47 314 207 302
151 7 5/1/2006 9:47 317 207 302
151 7 5/1/2006 9:47 326 207 302
151 7 5/1/2006 9:47 325 207 300
151 7 5/1/2006 9:47 325
151 7 5/1/2006 9:47 325
151 7 5/1/2006 9:47 317
159 7 5/1/2006 21:42 261
159 7 5/1/2006 21:42 263
159 7 5/1/2006 21:43 271
159 7 5/1/2006 16:37 275
159 7 5/1/2006 20:00 279
159 7 5/1/2006 9:48 306
159 7 5/1/2006 9:48 306
159 7 5/1/2006 9:48 304
159 7 5/1/2006 9:48 301
159 7 5/1/2006 20:02 279
159 7 5/1/2006 20:02 278
159 7 5/1/2006 20:02 278
159 7 5/1/2006 16:37 278
159 7 5/1/2006 16:37 278
159 7 5/1/2006 14:55 276
159 7 5/1/2006 14:55 275
159 7 5/1/2006 14:56 266
159 7 5/1/2006 14:56 260 207 258
159 7 5/1/2006 9:49 289 207 289
159 7 5/1/2006 9:49 285 207 289
159 7 5/1/2006 1:20 271 207 270
159 7 5/1/2006 1:20 272 207 270
Sheet1 Sheet2 Sheet3

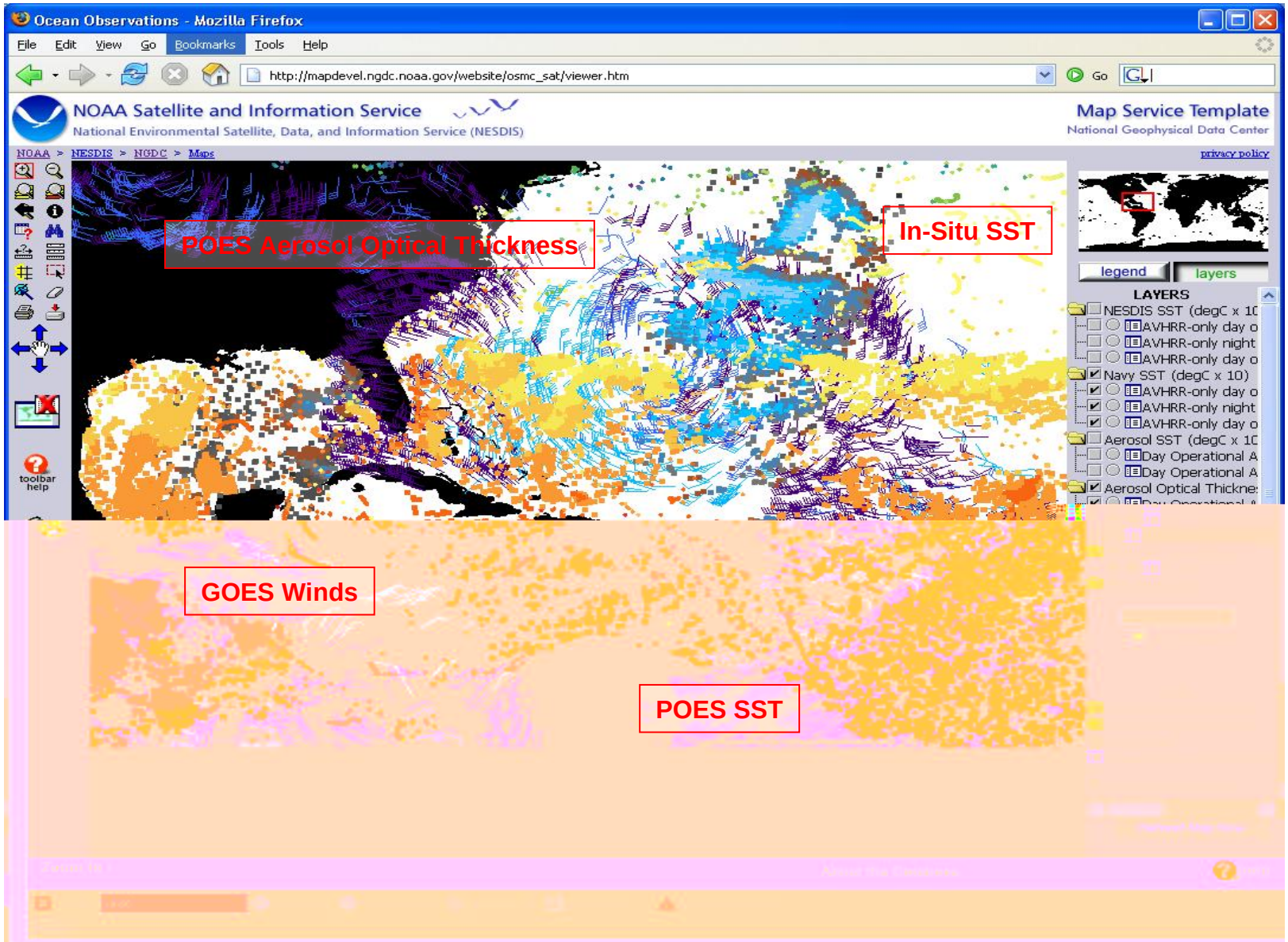

WWW Browser
WIST
G. Earth
LAS
WSDL
IDB
Desktop DBMS
Desktop GIS
Extract
ArcIMS
SQL Queries
Points
Lines
Polygons
Rasters
w/ attrib.
SDIF
Time Series
Geospatial Database
Model
GRIB
Other
HDF5
Multi-Dimensional Grids
ArcIMS Viewer - Mozilla Firefox
NOAA Satellite and Information Service
Map Service Template
Refresh Map
Zoom In
About the Database

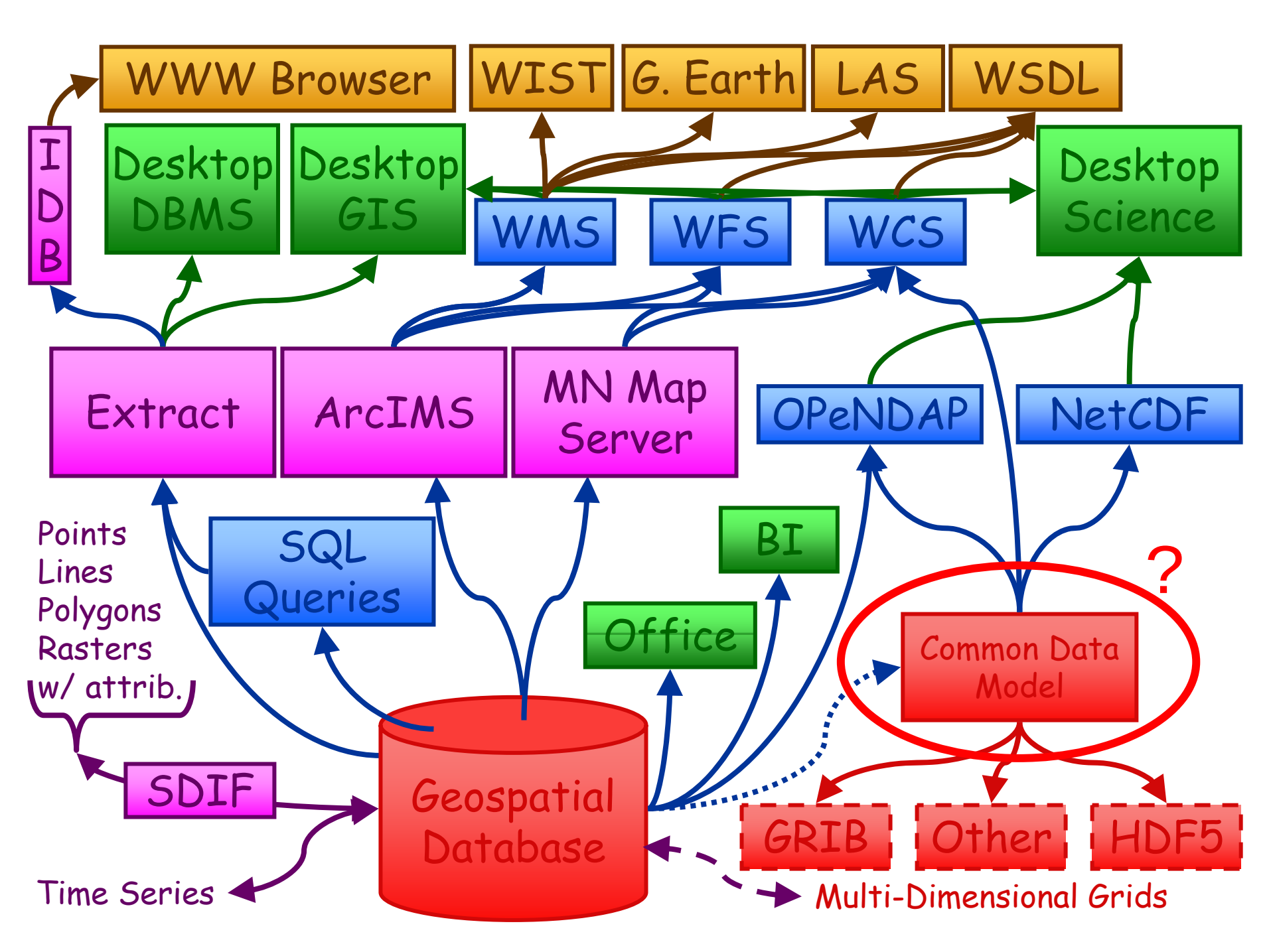

WWW Browser
WIST
G. Earth
LAS
WSDL
IDB
Desktop DBMS
Desktop GIS
WMS
WFS
WCS
Desktop Science
Extract
ArcIMS
Points
Lines
Polygons
Rasters
w/ attrib.
SQL Queries
SDIF
Time Series
Multi-Dimensional Grids


WWW Browser
WIST
G. Earth
LAS
WSDL
IDB
Desktop DBMS
Desktop GIS
WMS
WFS
WCS
Desktop Science
Extract
ArcIMS
MN M Serv
Google Earth
Points
Lines
Polygons
Rasters
w/ attrib.
SQL Queries
SDIF
Geospatial Database
GRIB
Other
HDF5
Time Series
Multi-Dimensional Grids


WWW Browser
WIST
G. Earth
LAS
WSDL
IDB
Desktop DBMS
Desktop GIS
WMS
WFS
WCS
Desktop Science
Extract
ArcIMS
MN Map Server
OPeNDAP
NetCDF
Points
Lines
Polygons
Rasters
w/ attrib.
SQL Queries
BI
Office
Common Data Model
SDIF
Geospatial Database
GRIB
Other
HDF5
Time Series
Multi-Dimensional Grids

Integrated Visualization (GIS)




WWW Browser
WIST
G. Earth
LAS
WSDL
IDB
Desktop DBMS
Desktop GIS
WMS
WFS
WCS
Desktop Science
Extract
ArcIMS
MN Map Server
OPeNDAP
NetCDF
Points
Lines
Polygons
Rasters
w/ attrib.
SQL Queries
BI
Office
?
Common Data Model
SDIF
Geospatial Database
GRIB
Other
HDF5
Time Series
Multi-Dimensional Grids

Partnership?

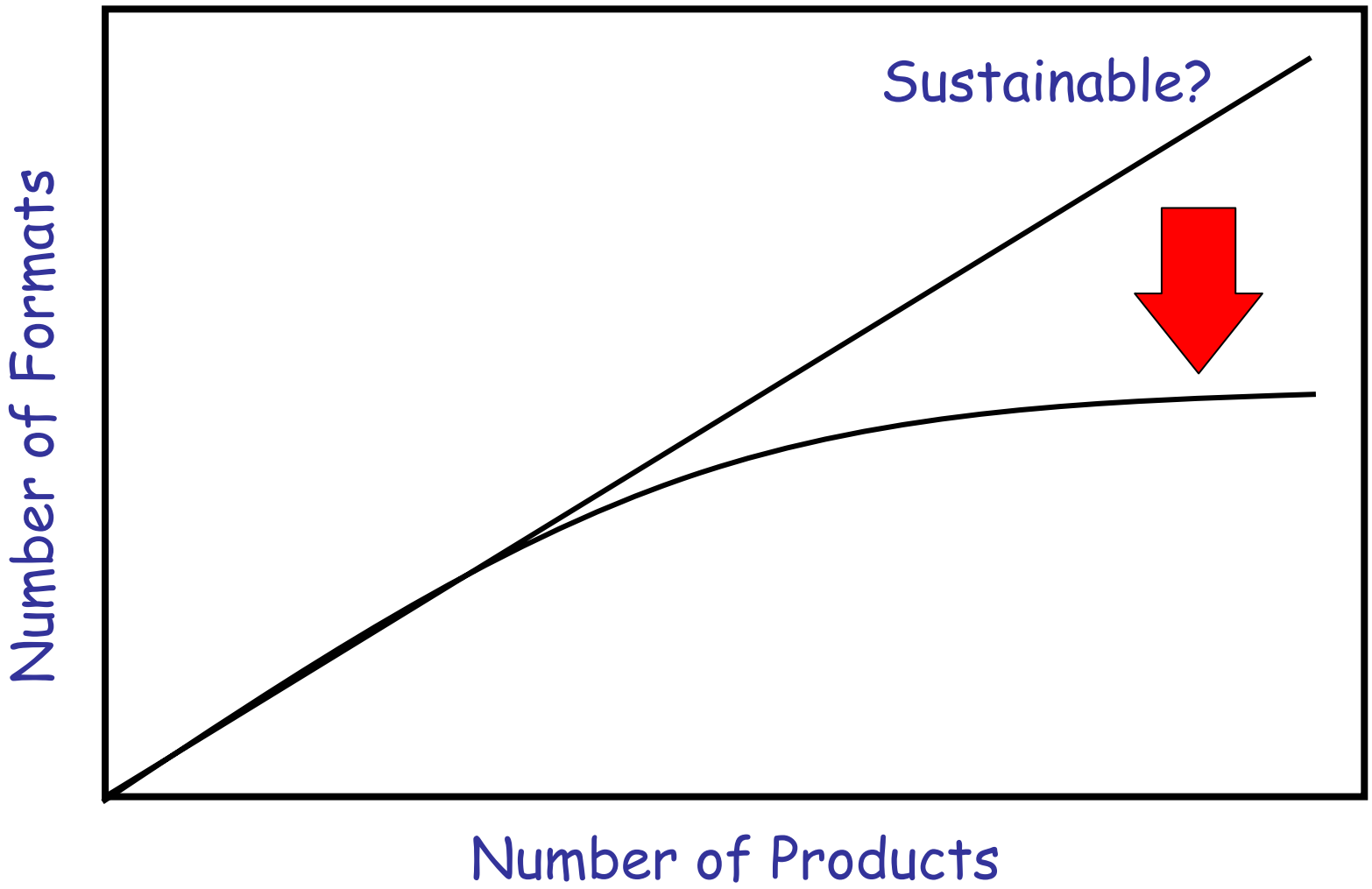
NOAA is a very different kind of organization than Unidata, but there are good signs:

NOAA Data Management Integration Team (DMIT) voted "Support for Common Data Model" as the #1 recommendation to IOOS for work that is consistent with the NOAA GEO-Integrated Data Environment Plan.

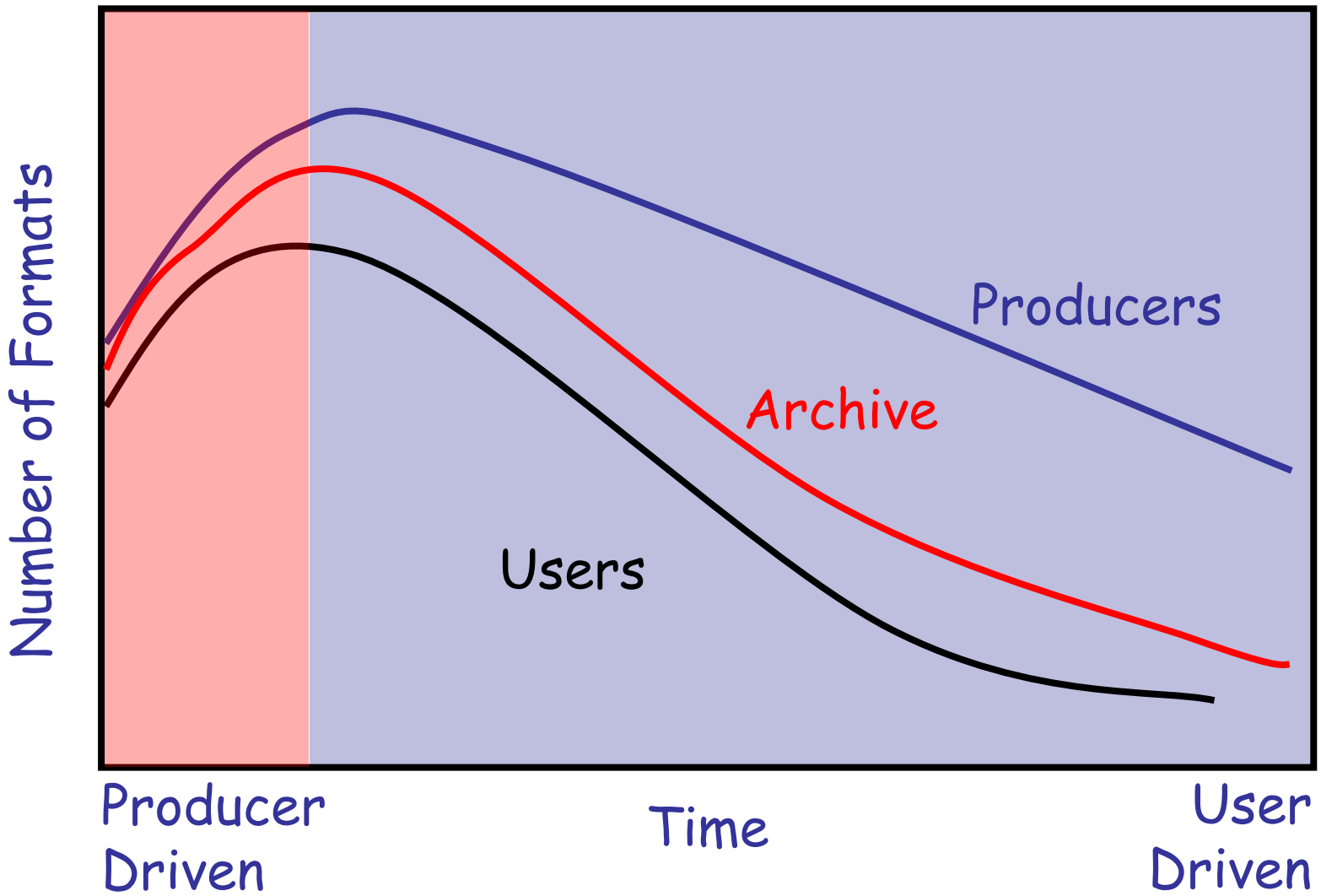
10 NOAA people attended Unidata training.

8 CLASS developers and others attending HDF Conference.

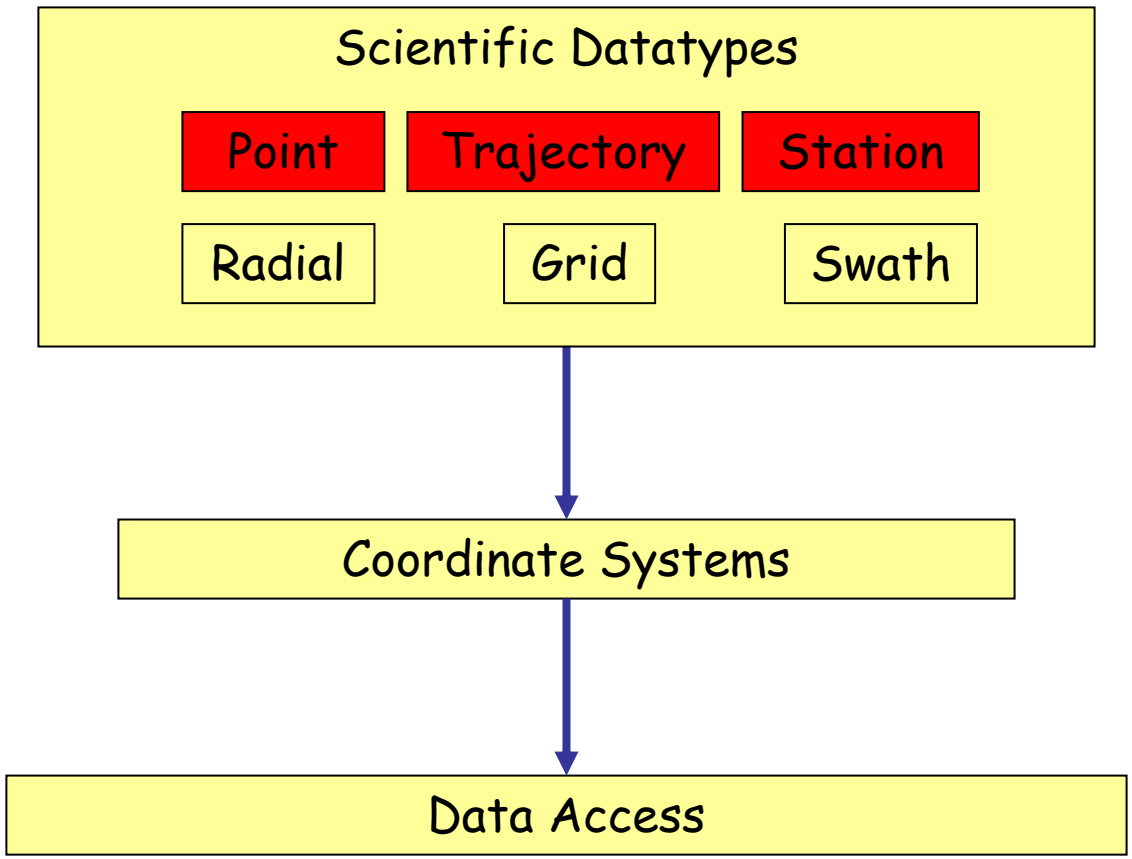
Formats and Products



Format Evolution



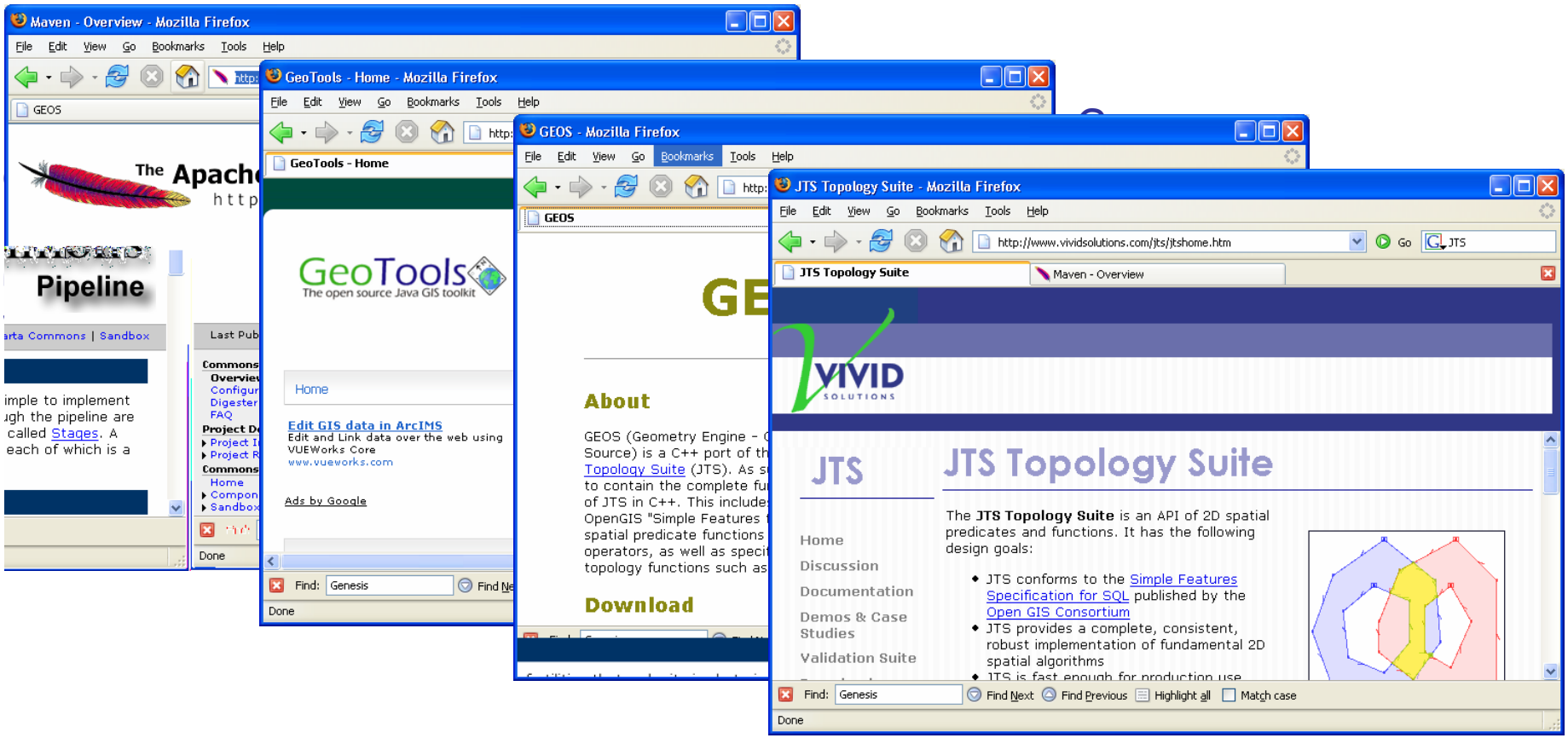
Common Data Model



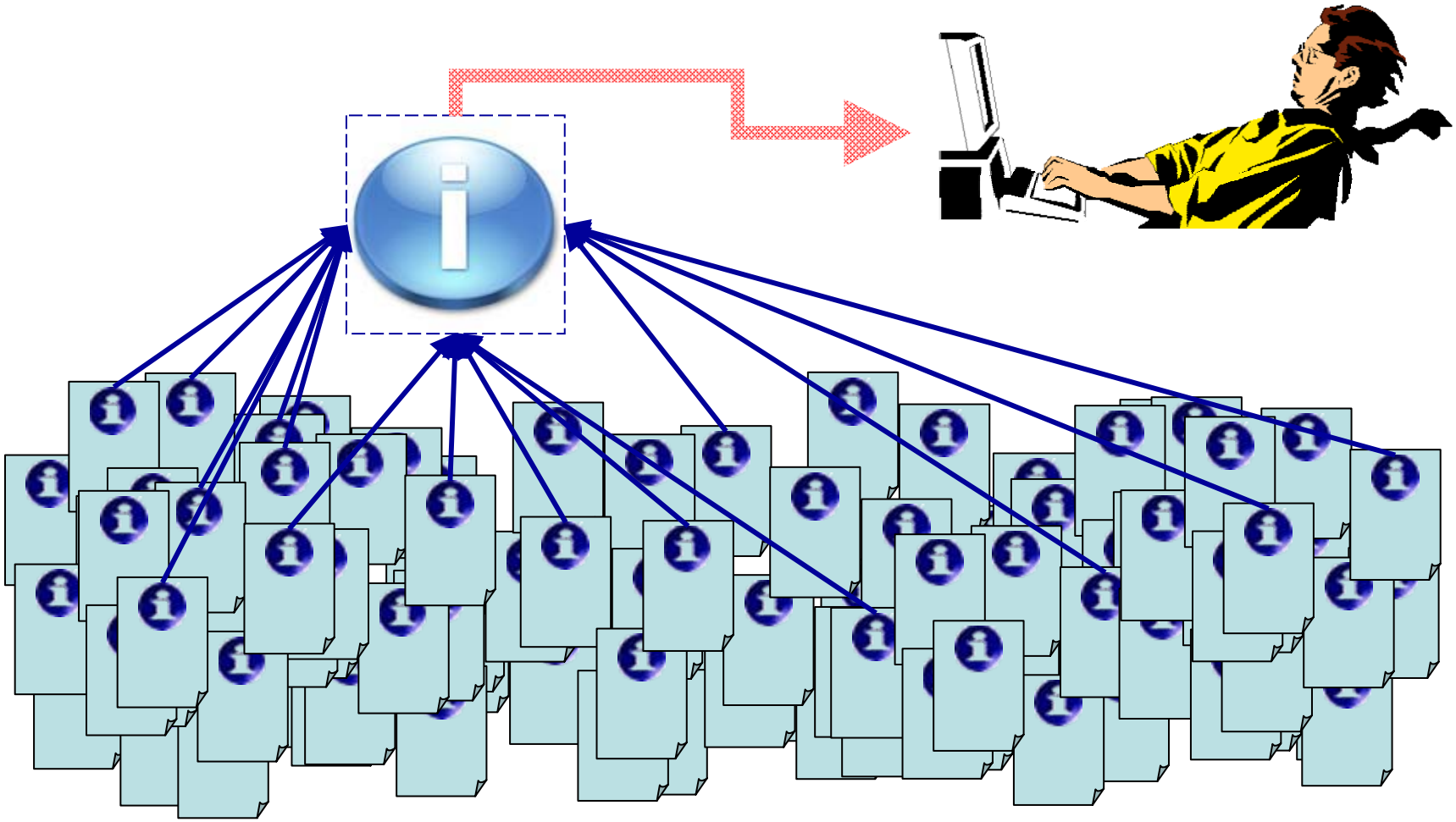
Open Geospatial Consortium Simple Features

Simple Features Spec

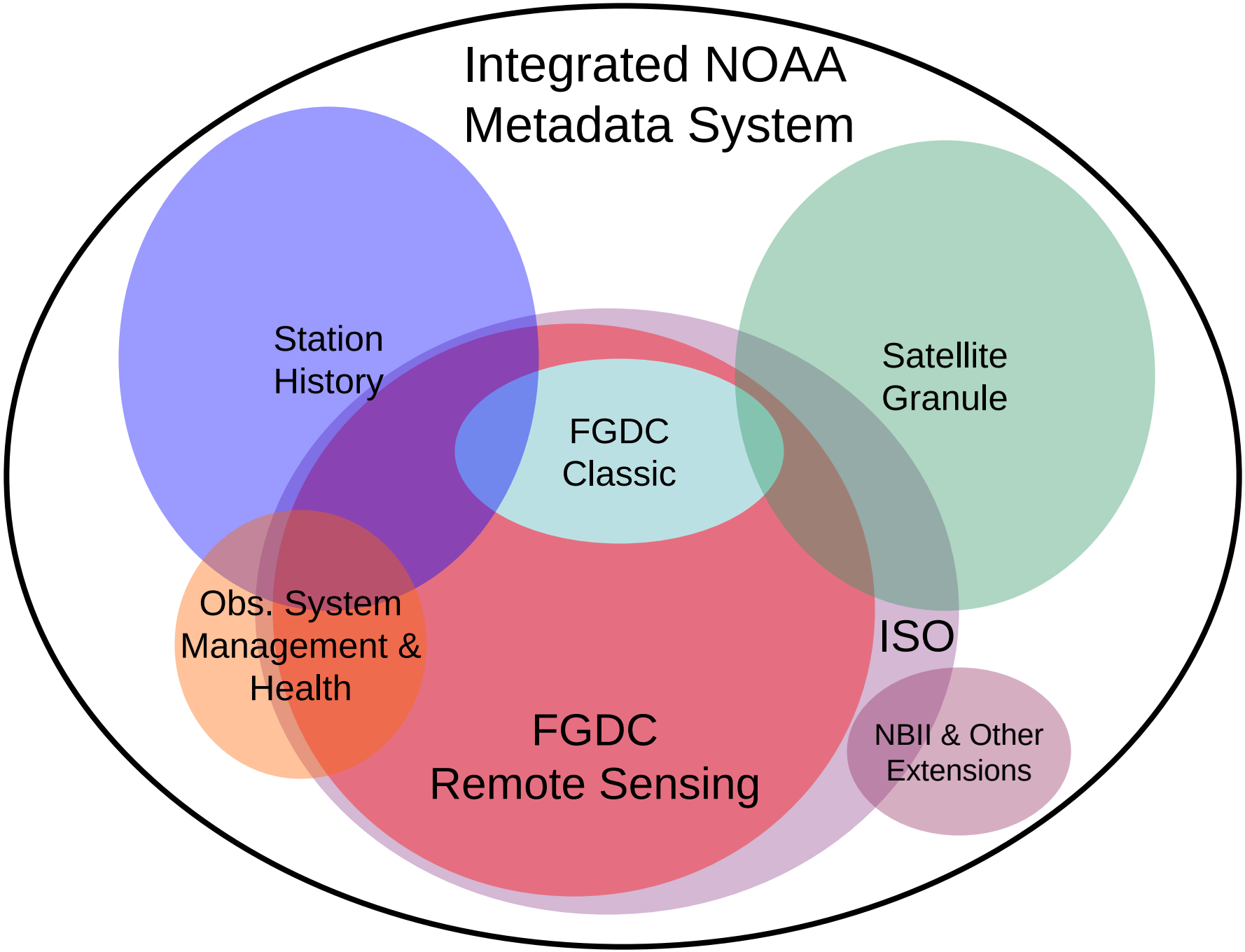
The Simple Feature Specification application programming interfaces (APIs) provide for publishing, storage, access, and simple operations on Simple Features (point, line, polygon, multi-point, etc). The purpose of these specifications is to describe interfaces to allow GIS software engineers to develop applications that expose functionality required to access and manipulate geospatial information comprising features with 'simple' geometry using different technologies.



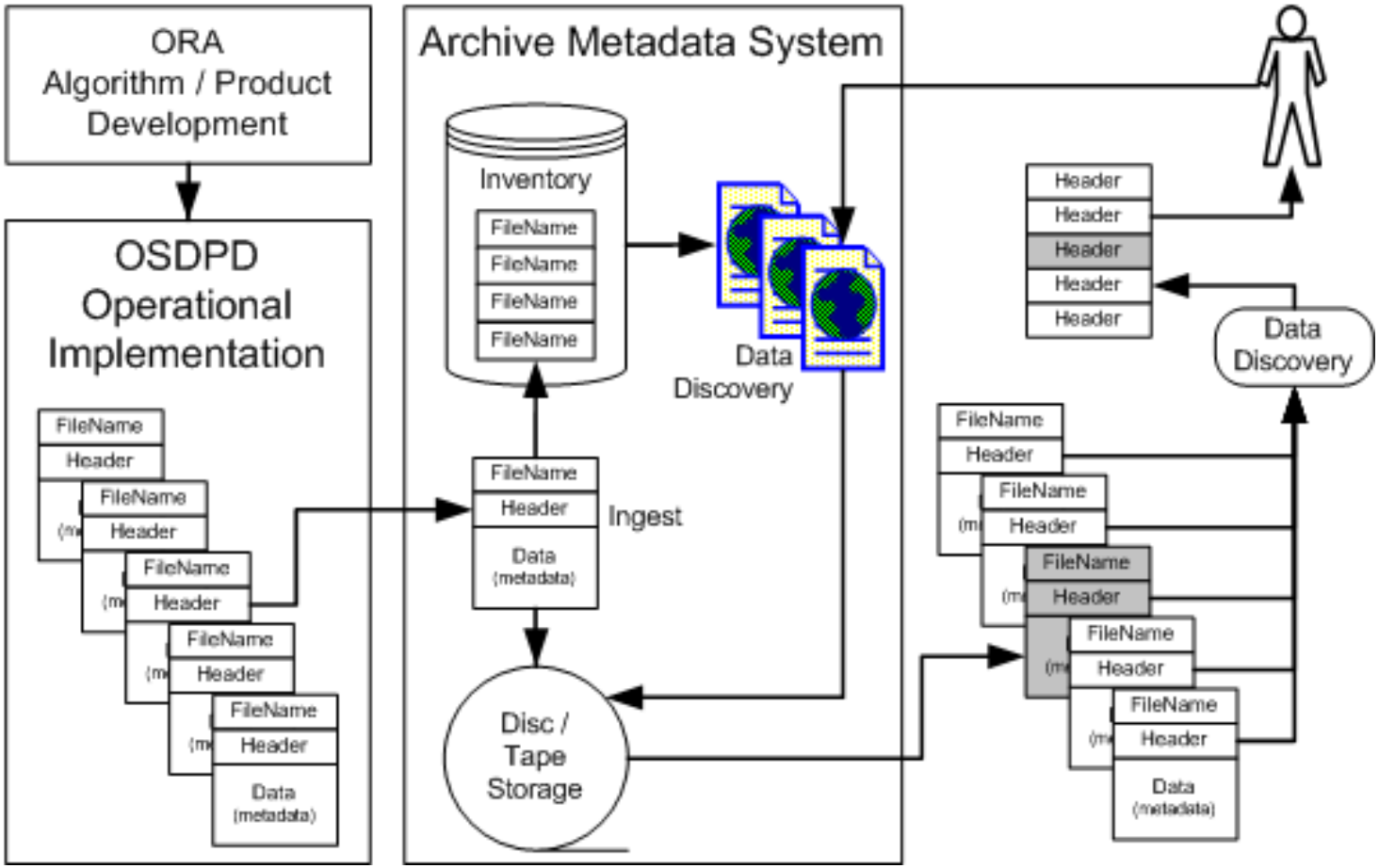
The Rich Inventory Concept



Very similar to "file content metadata" at NCAR

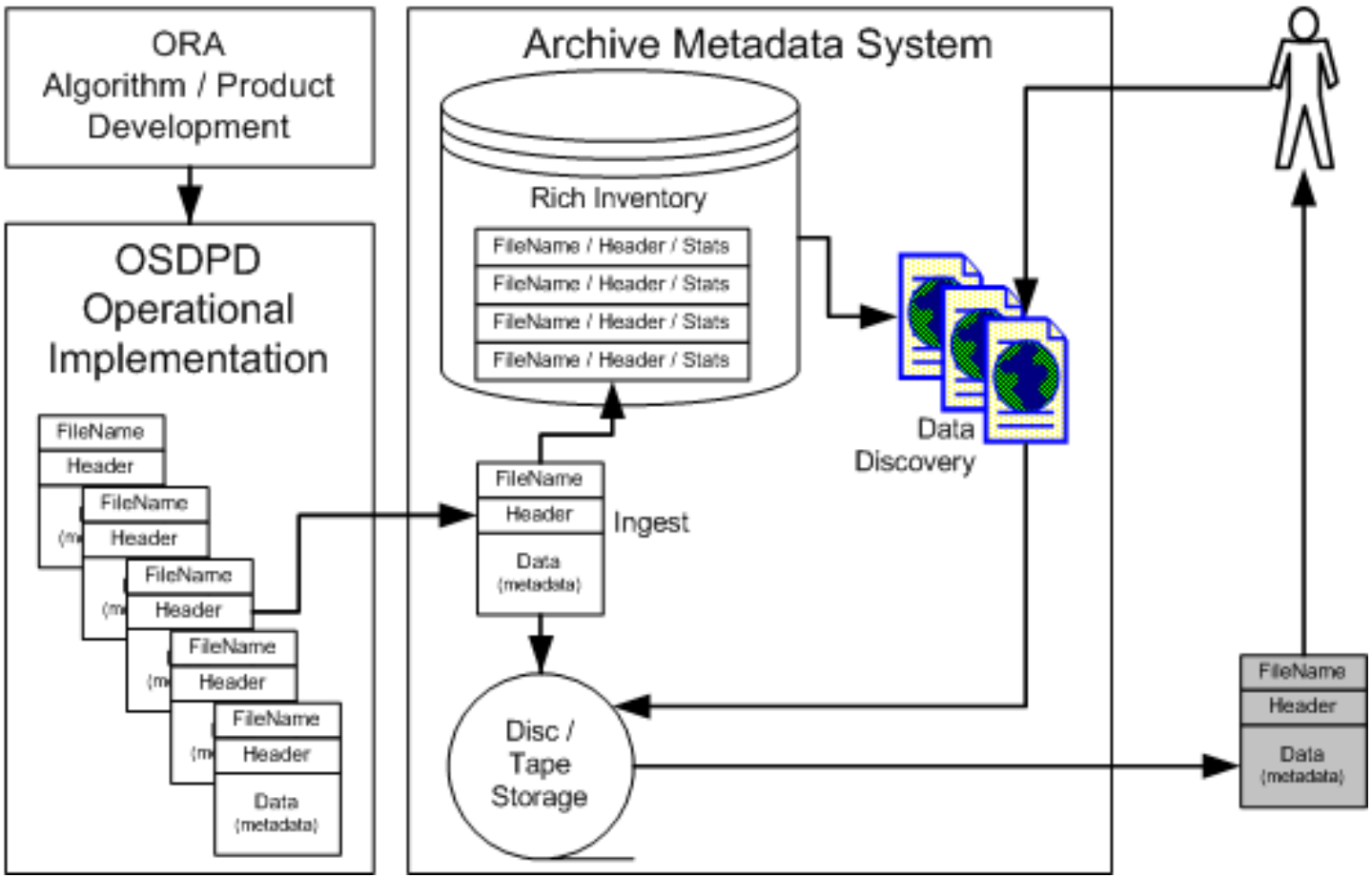

Integrated NOAA
Metadata System
Station
History
Satellite
Granule
FGDC
Classic
Obs. System
Management &
Health
ISO
FGDC
Remote Sensing
NBII & Other
Extensions

Granule Metadata Management - Present



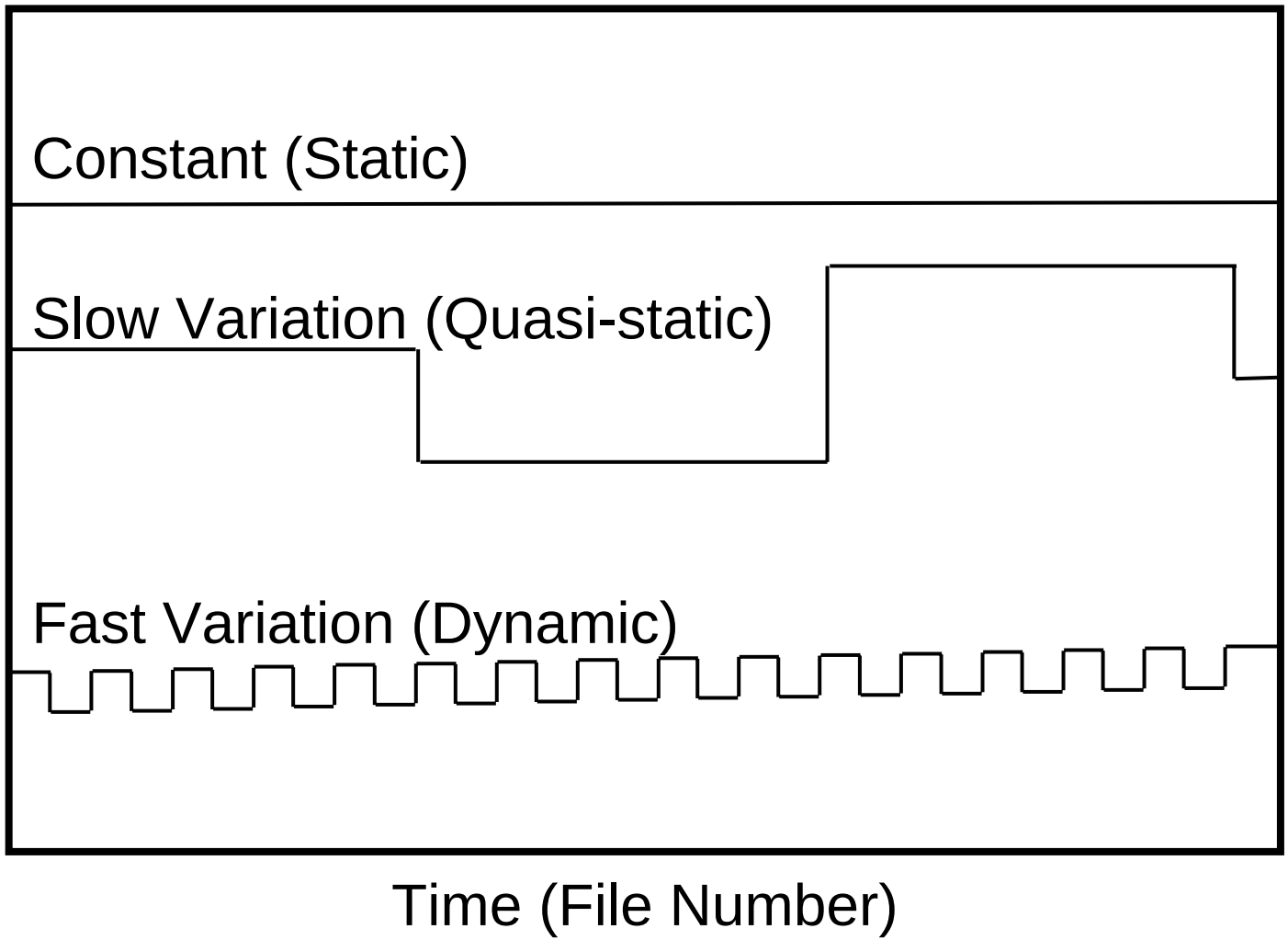
1. Files come to CLASS and filename metadata is ingested into inventory.
2. Fileheader metadata is stored and is not available to data discovery system.
3. Descriptive Statistics are not calculated.
4. Users need to develop their own data discovery systems.

Granule Metadata Management - Rich Inventory

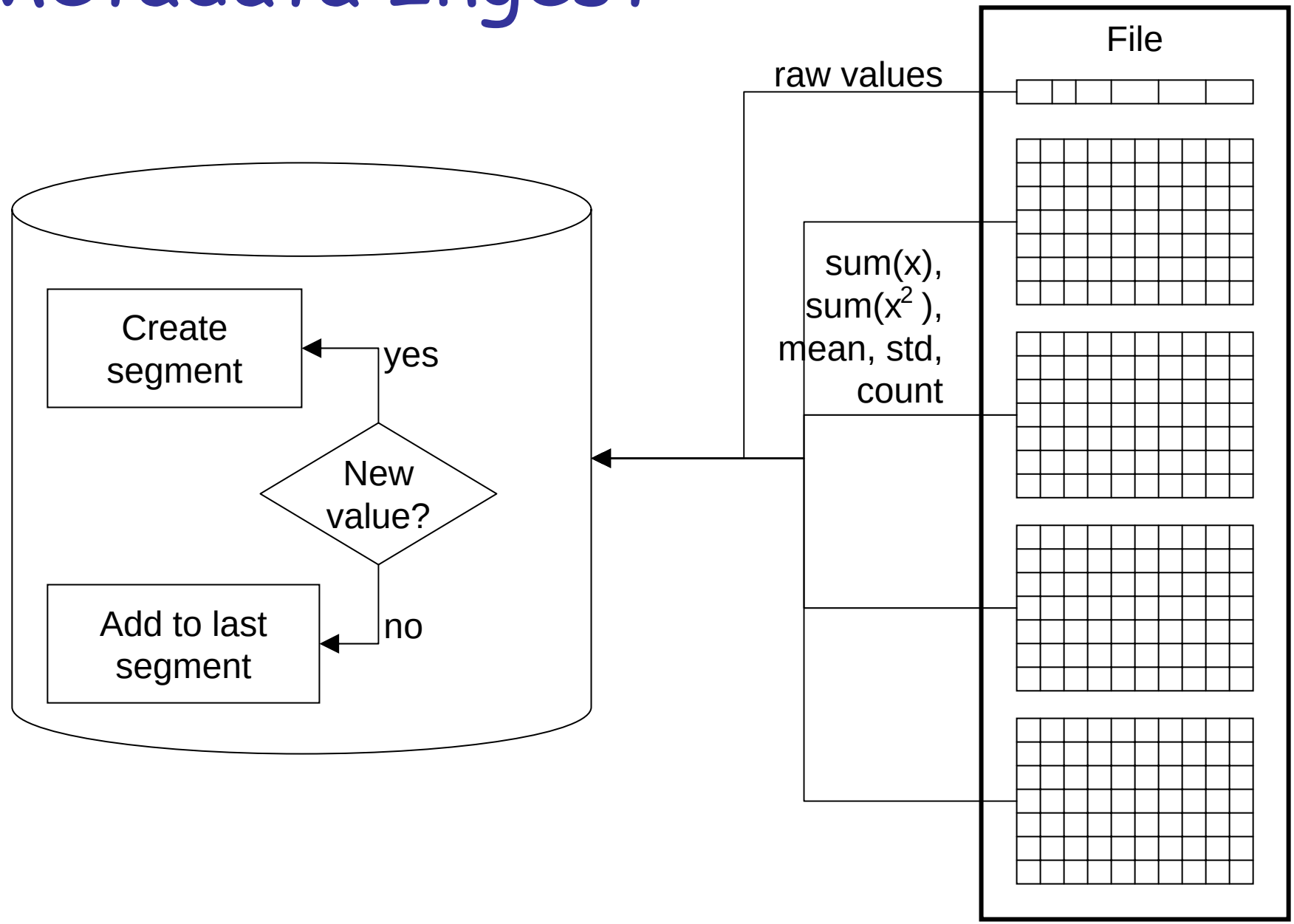


1. Files come to CLASS
2. Filename and fileheader metadata are added to inventory.
3. Descriptive Statistics are calculated and added to inventory.
4. All metadata is available to the data discovery system and users get the data they need without secondary data discovery.

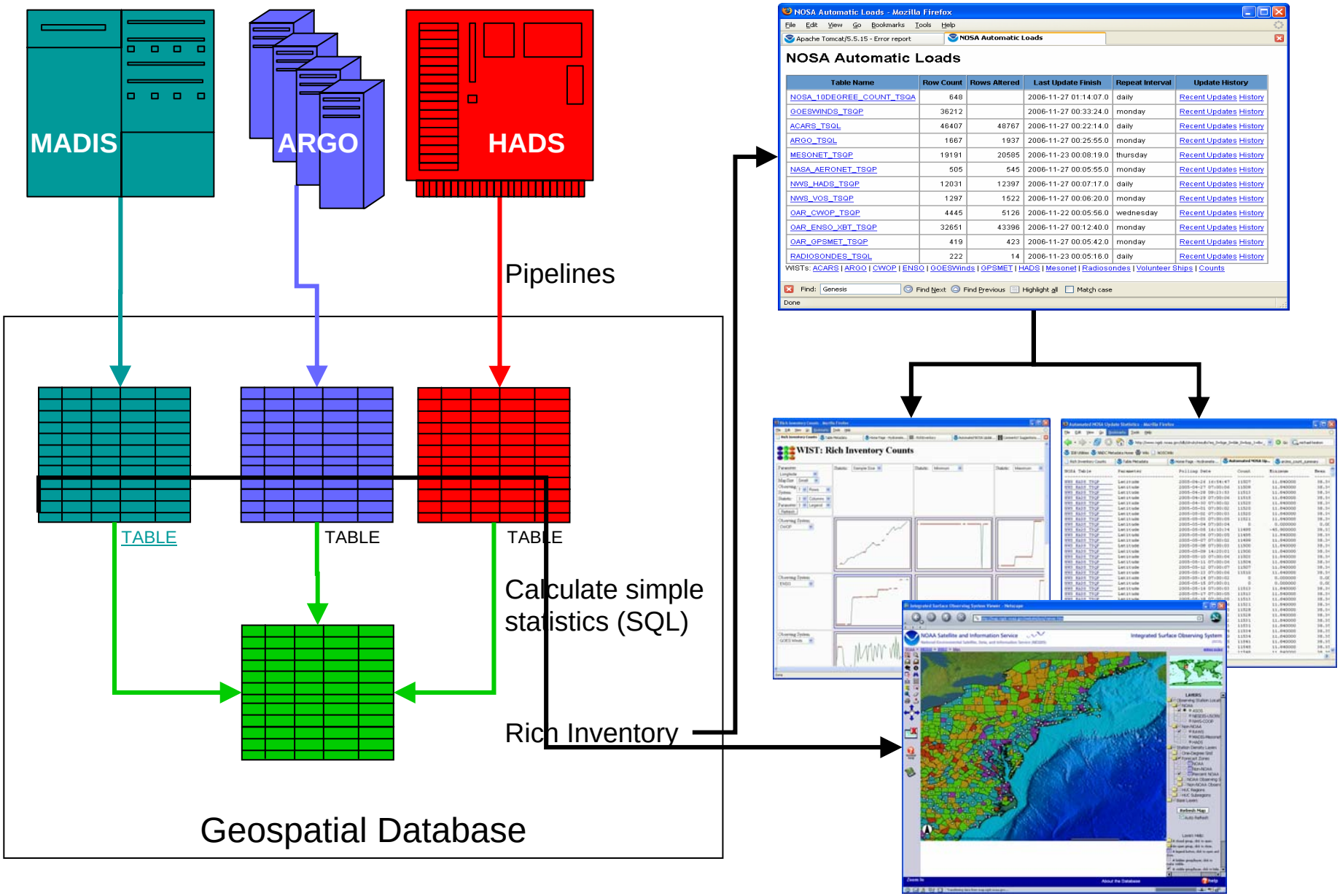
Segment Model



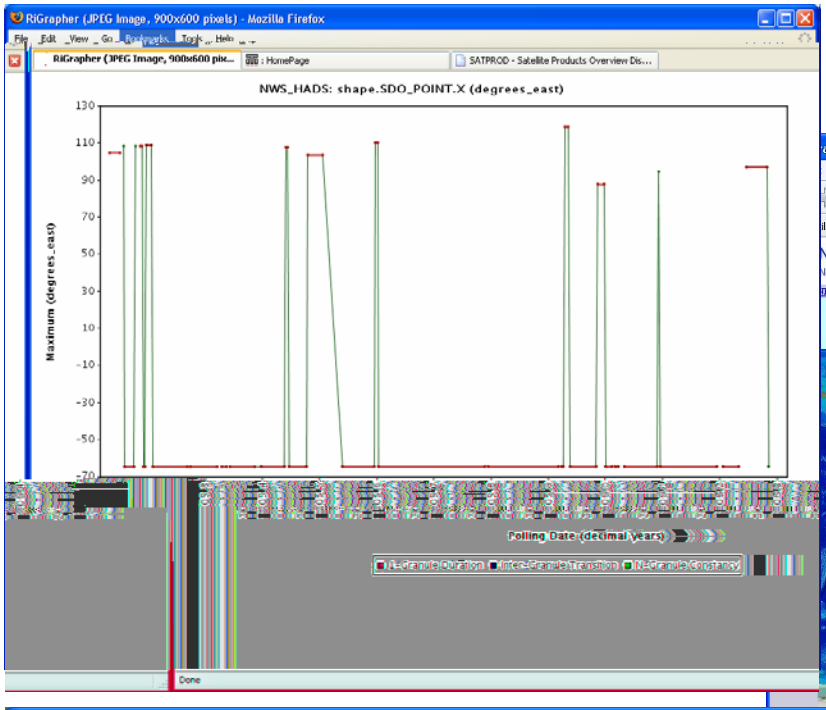
Metadata Ingest

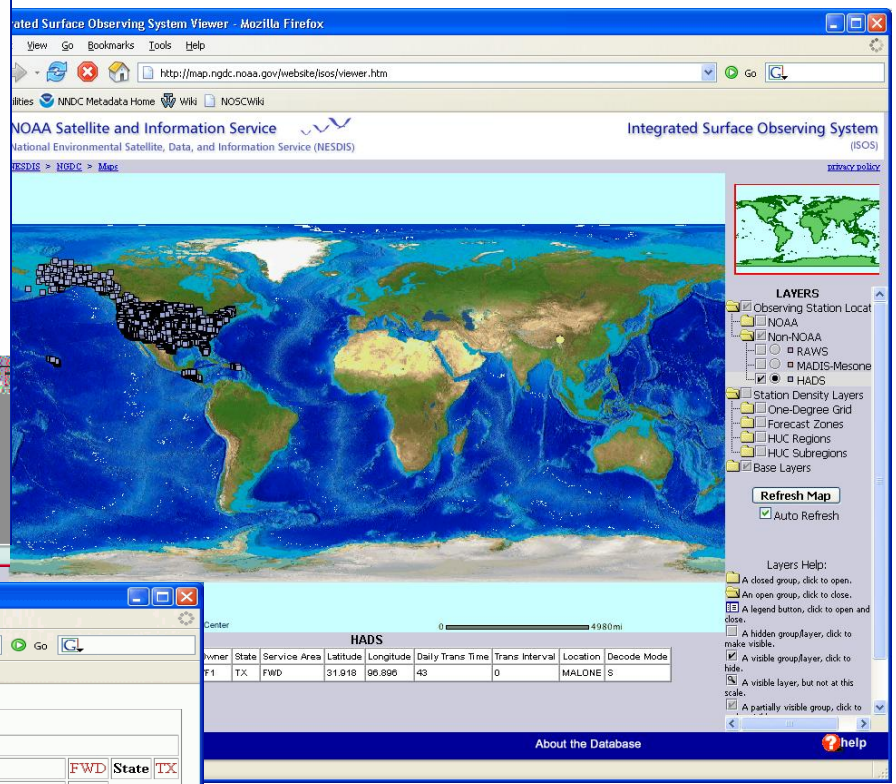


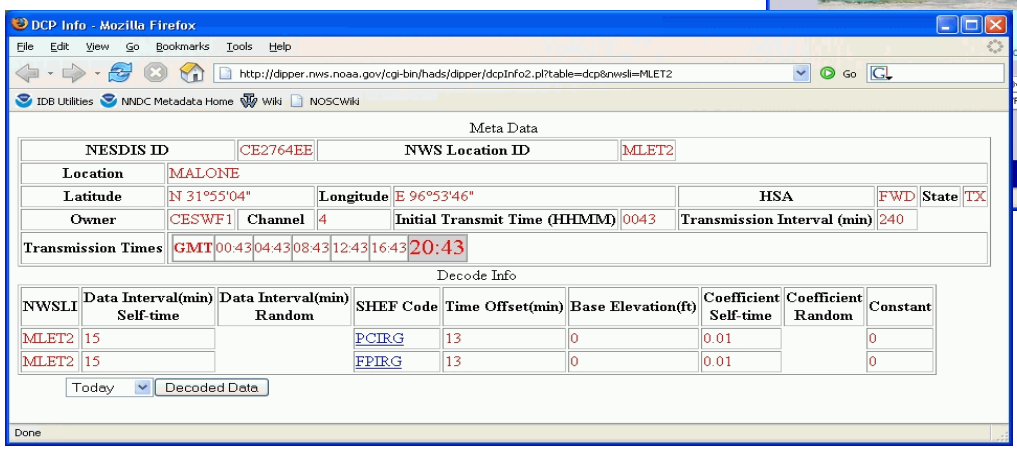
Automated Observing System Ingest



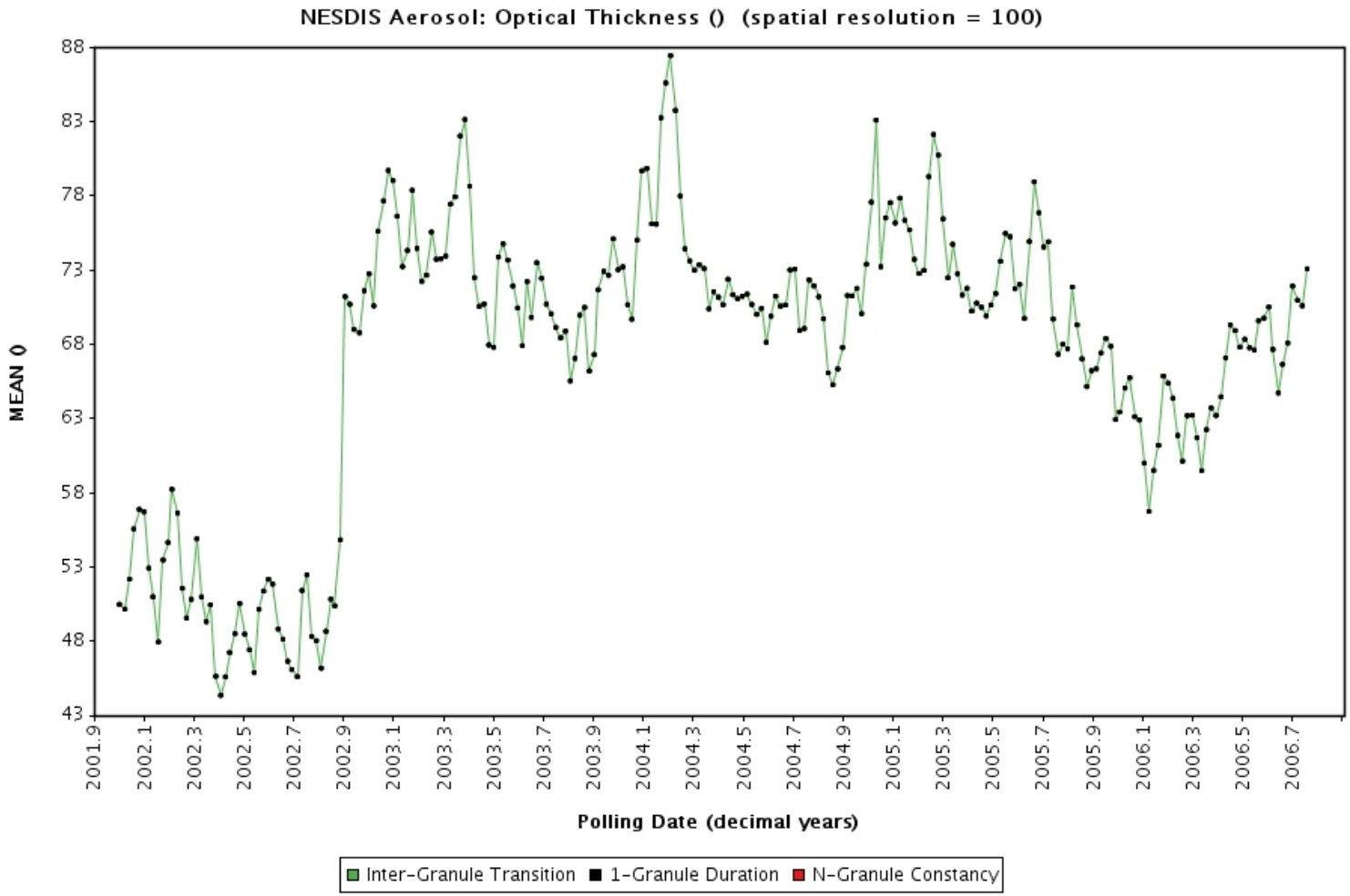
HADS Network Monitoring







Algorithm Change: Aerosol



Algorithm Change: Aerosol

Hi Ted,
Dr. Ignatov and I did some digging and this is the result. Sasha's conclusion is the most pertinent info we could find from logs or email archives. Here it is:

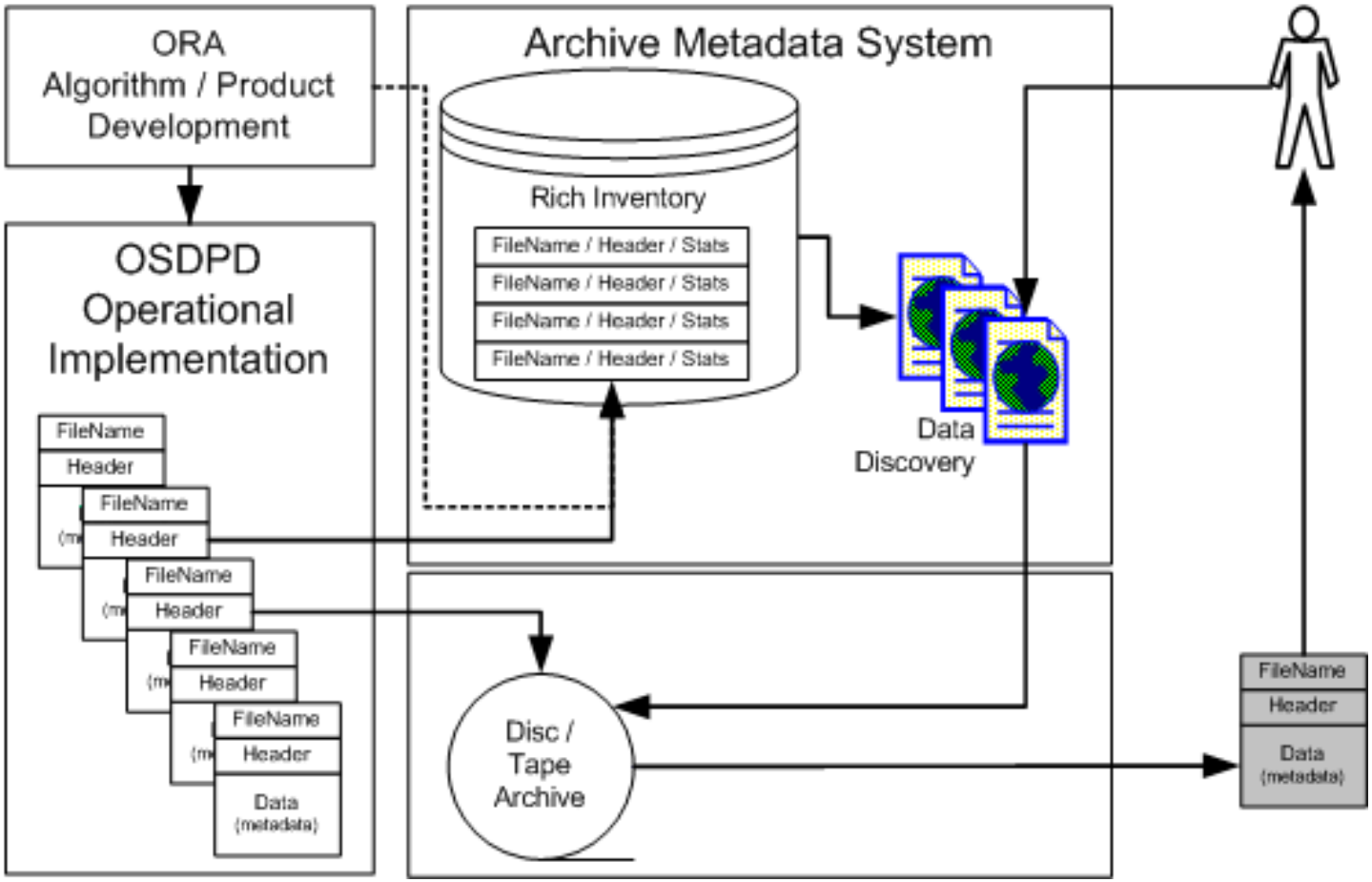
Hi John,
i checked my 2002 email archives, and here is what i found out:

it appears that the current 3rd generation aerosol algorithm was implemented into operations around Oct-Nov 2002 time frame. cannot say more precisely, as all email correspondence i am looking at, talks about this indirectly. (maybe it's what Steve refers to as the Phase II aerosol-SST algorithm.) At the same time, Steve had implemented quite a few other changes fixing data bugs and formats: view angle problem in AEROBS, increased digitization in all channel's reflectances and AODs, etc.

The jump in AOD1 is deemed due to introducing 3rd generation algorithm, which replaced the 2nd generation. The new numbers (~0.08) look more realistic than the previous ones (~0.05 or so). The changes seen in the data is close to the expected effect of this change. the 3rd gen alg takes into account the exact spectral response of N16 AVHRR, whereas the 2nd gen was using a generic set of LUTs for all AVHRRs ("one size fits all").

hopefully this settles the issue..
cheers, sasha

Granule Metadata Management - Distributed Ingest



1. Product generation algorithms write all metadata to inventory directly instead of file headers.
2. Files are archived somewhere with pointers from Inventory.
3. Users get the data they need from distributed system without secondary data discovery.


ted.habermann@noaa.gov