

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

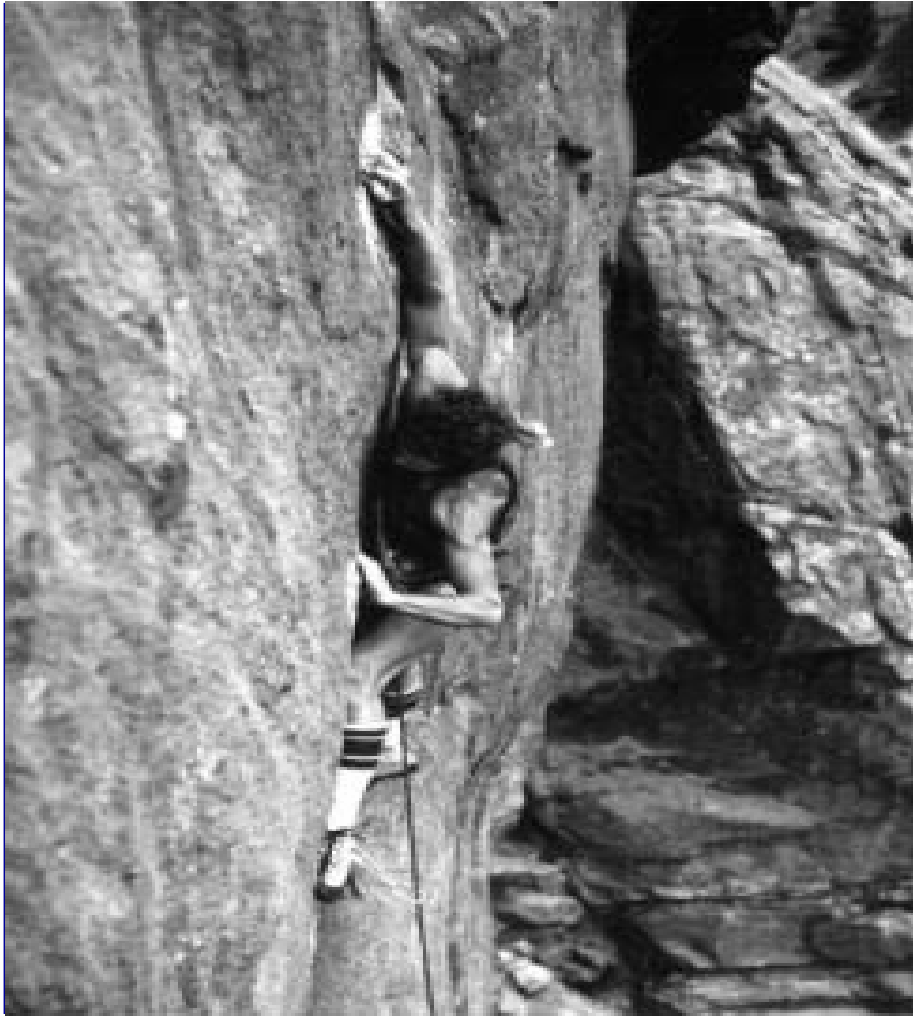
Ted Habermann  
NOAA National Data Centers

**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 ascent 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.*



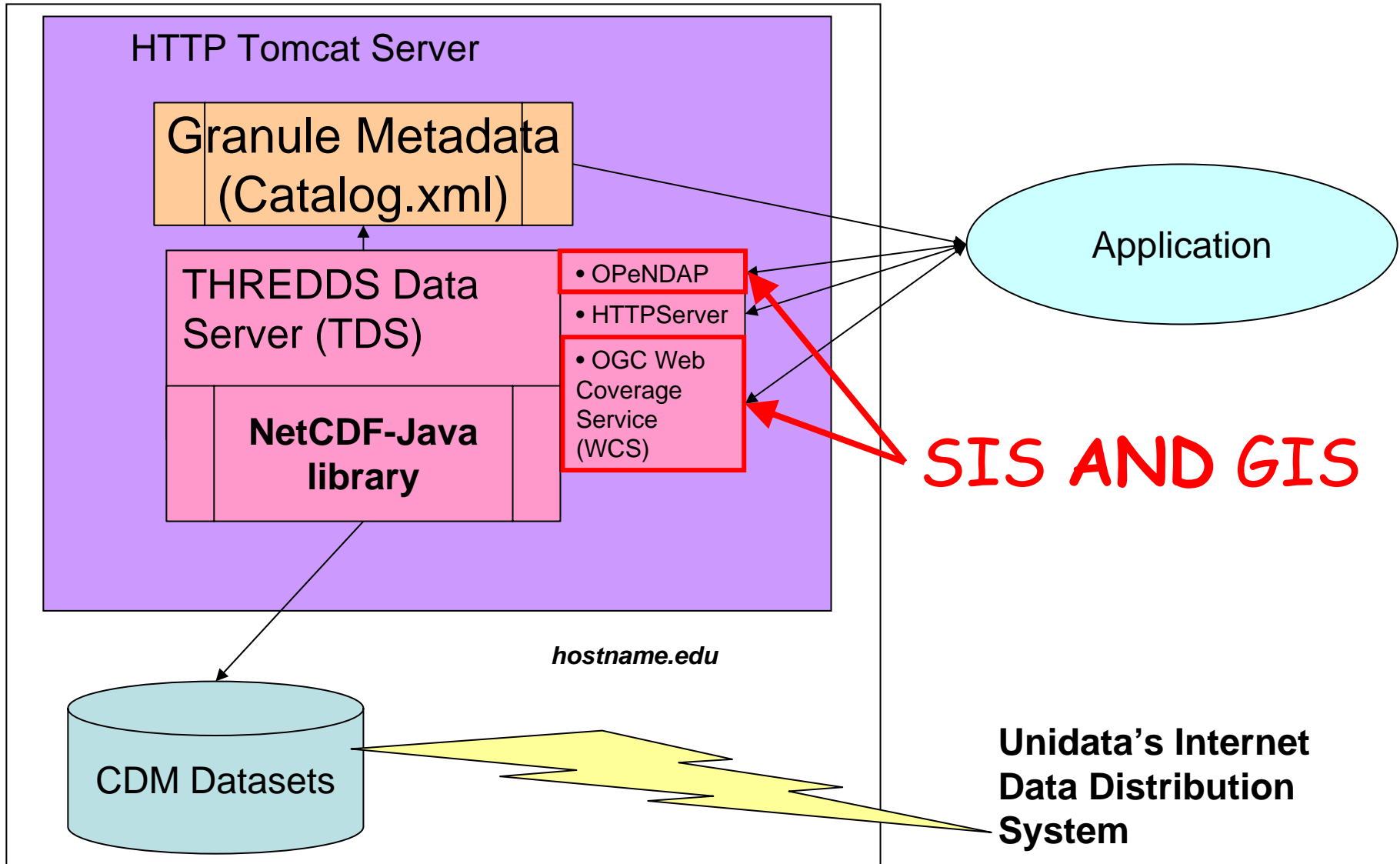
*Identified characteristics of visionary companies through comparisons with comparable companies. One characteristic was:*

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

# 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   |
| <b>science information systems</b>         |    | <b>geographic information systems</b>     |

# THREDDS Data Server



# Data Processing Levels



Telemetry information, Swaths

Grids

Time and Scan Angle

Latitude & Longitude

Complex custom formats (bits)

Standard formats (bytes)

Large volume

Small volume

Radiance in instrument units

Sea Surface Temp °C

**Complex and Hard**

**Simple and Easy**

POES Level 1b data

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

8km Level 2 SST

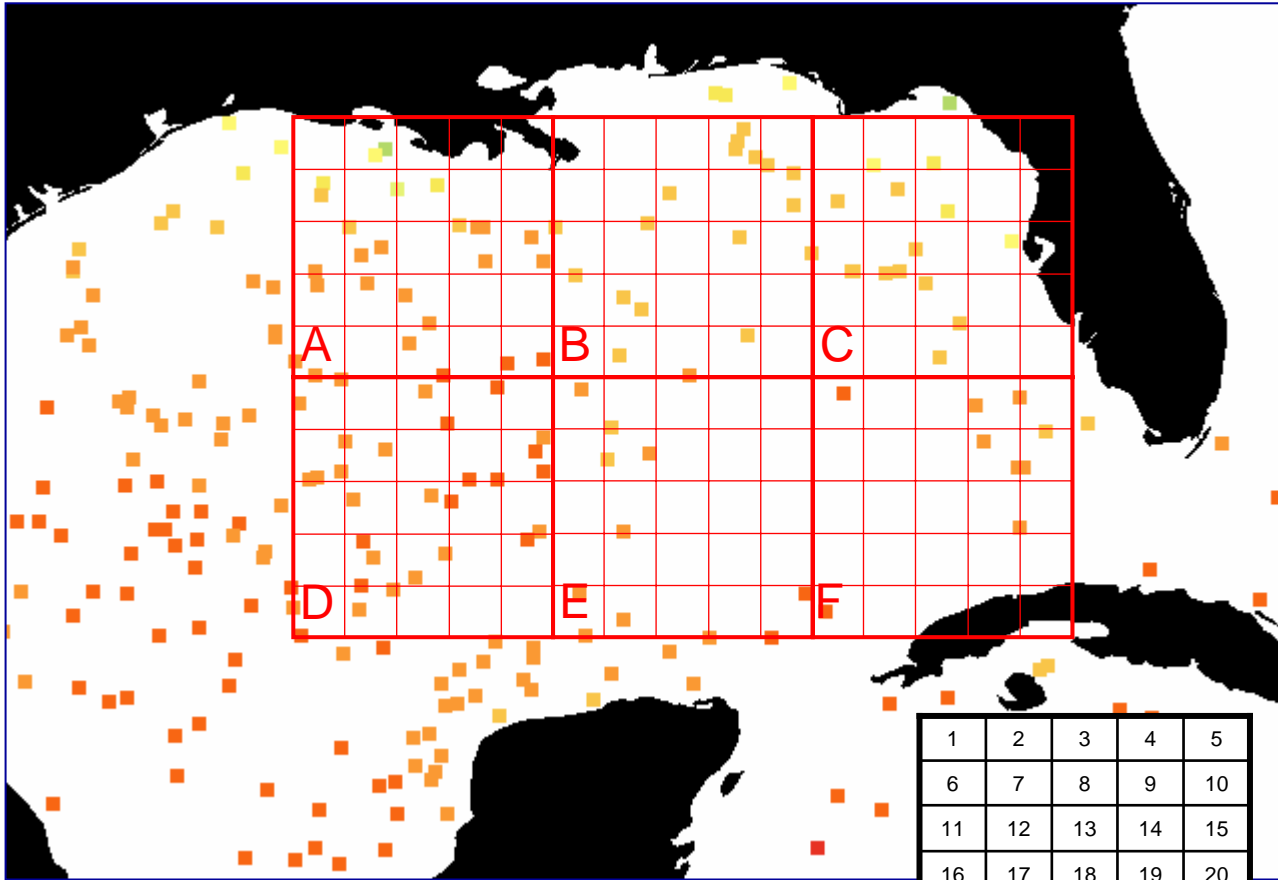
Most primitive useful form??

# 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.

|                                |                      |             |                      |            |                      |             |     |                           |     |                     |
|--------------------------------|----------------------|-------------|----------------------|------------|----------------------|-------------|-----|---------------------------|-----|---------------------|
| <b>Spatial Index</b>           |                      |             |                      |            |                      |             |     |                           |     |                     |
| <b>Block Directory Record</b>  |                      |             |                      |            |                      |             |     |                           |     |                     |
| 20 byte header                 | Block 1 Start Rec. # |             | Block 2 Start Rec. # |            | Block 3 Start Rec. # |             | ... | Block 2592 Start Rec. #   |     | Blanks              |
| <b>Observation Data Record</b> |                      |             |                      |            |                      |             |     |                           |     |                     |
| Rec #                          |                      | Block #     |                      | Extent #   |                      | Next Extent |     | Other Miscellaneous Stuff |     |                     |
| Subblock 1                     |                      | Subblock 2  |                      | Subblock 3 |                      | ...         |     | Subblock 25               |     |                     |
| Start                          | End                  | Start       | End                  | Start      | End                  | ...         |     | Start                     | End |                     |
| <b>Observation Unit</b>        |                      |             |                      |            |                      |             |     |                           |     |                     |
|                                |                      |             |                      |            |                      |             |     |                           |     | <b>Observations</b> |
| Type                           | Source               | Date / Time |                      | Location   |                      | Observation |     | Other Miscellaneous Stuff |     |                     |

# 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

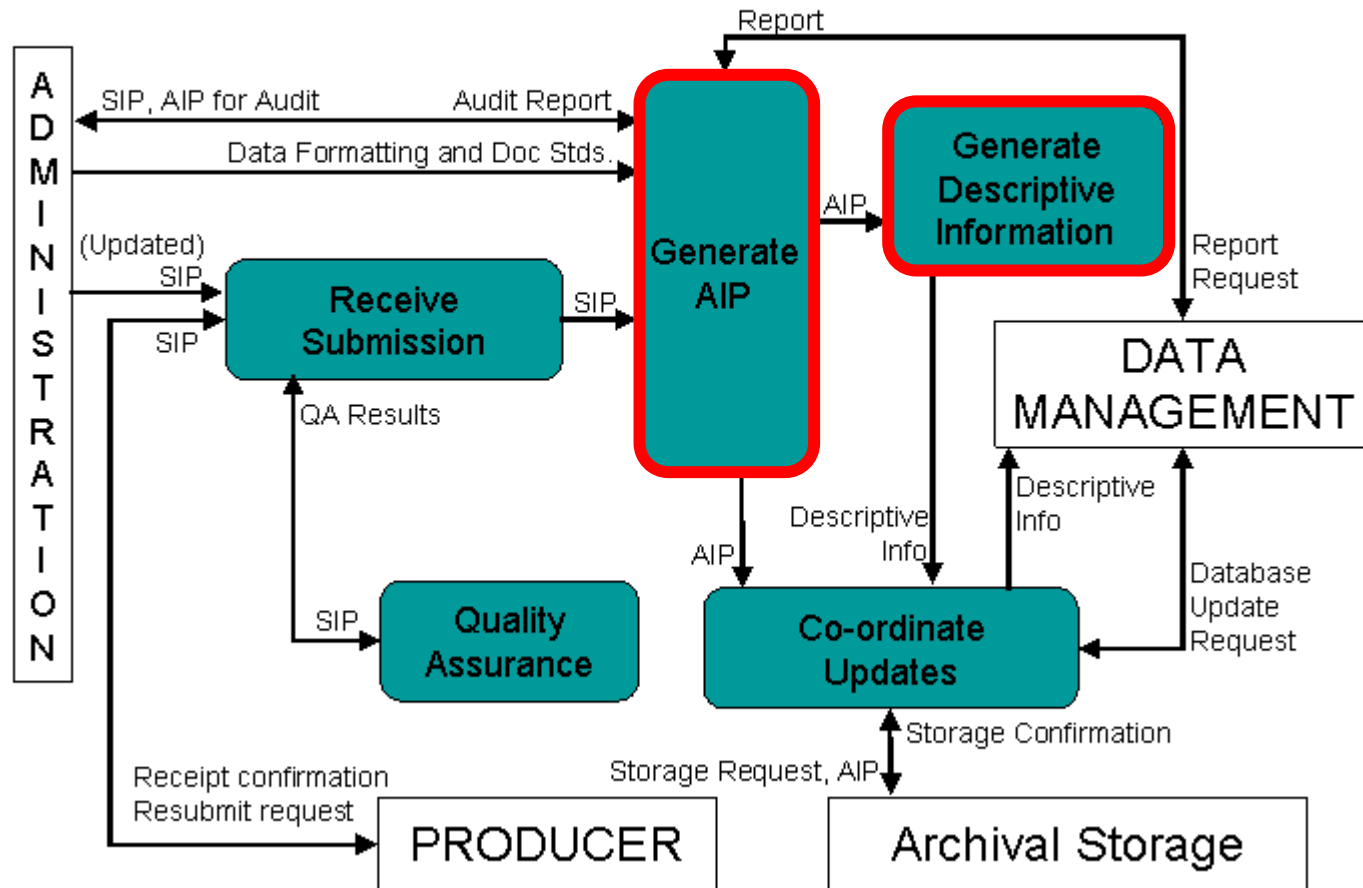
Next block ...

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.

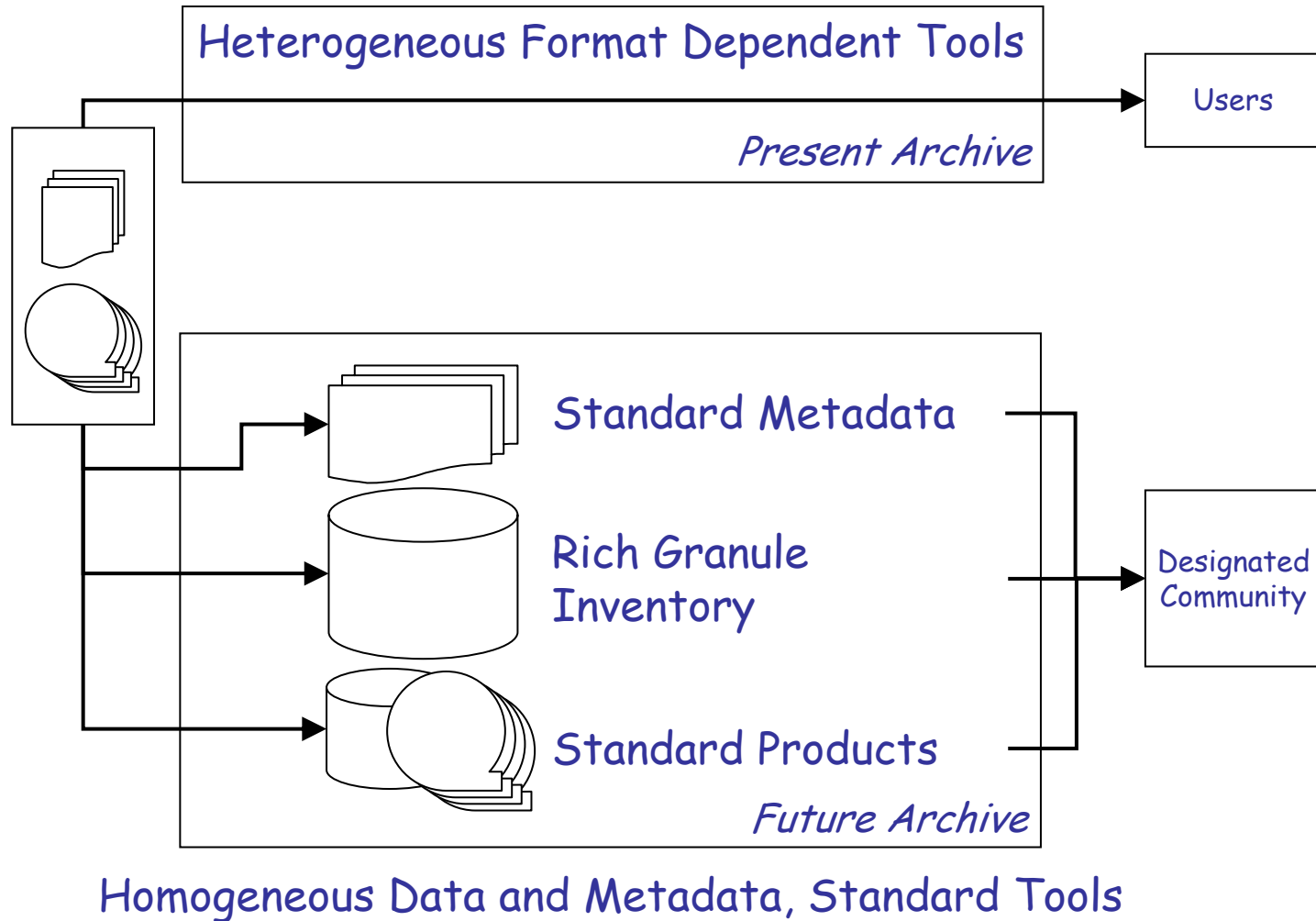
# OAIS Ingest Functions

## Ingest Functions



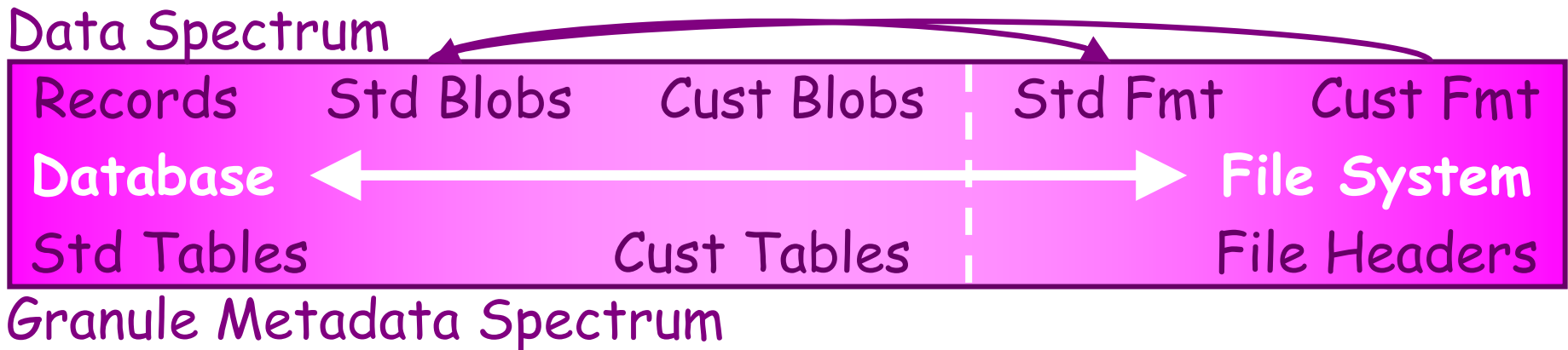


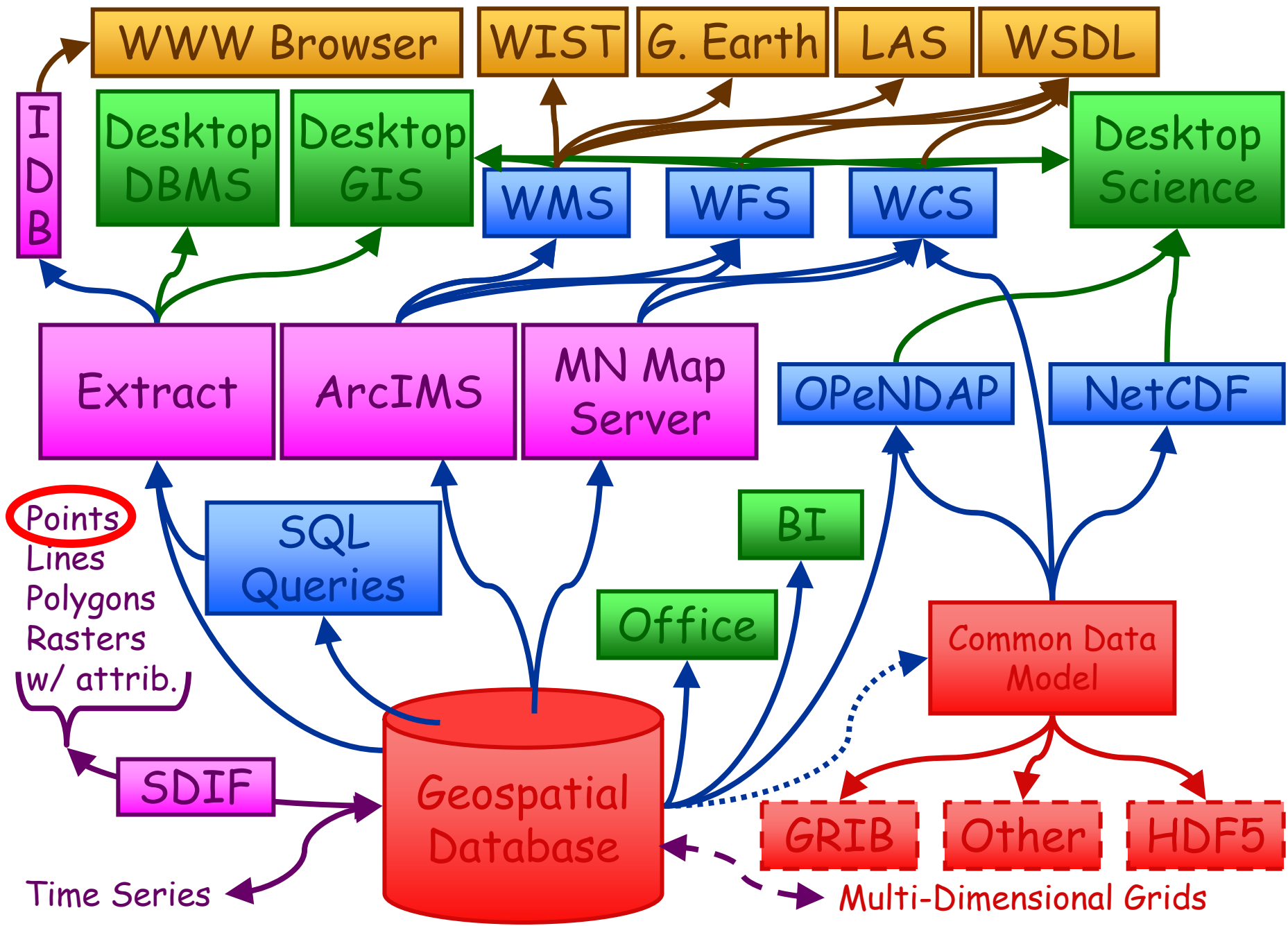
# Archive Process Evolution



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.



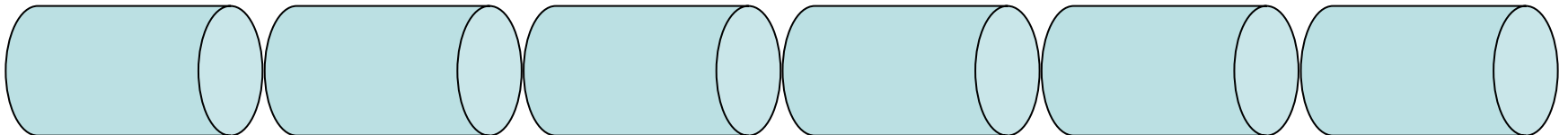


# Processing Pipeline

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

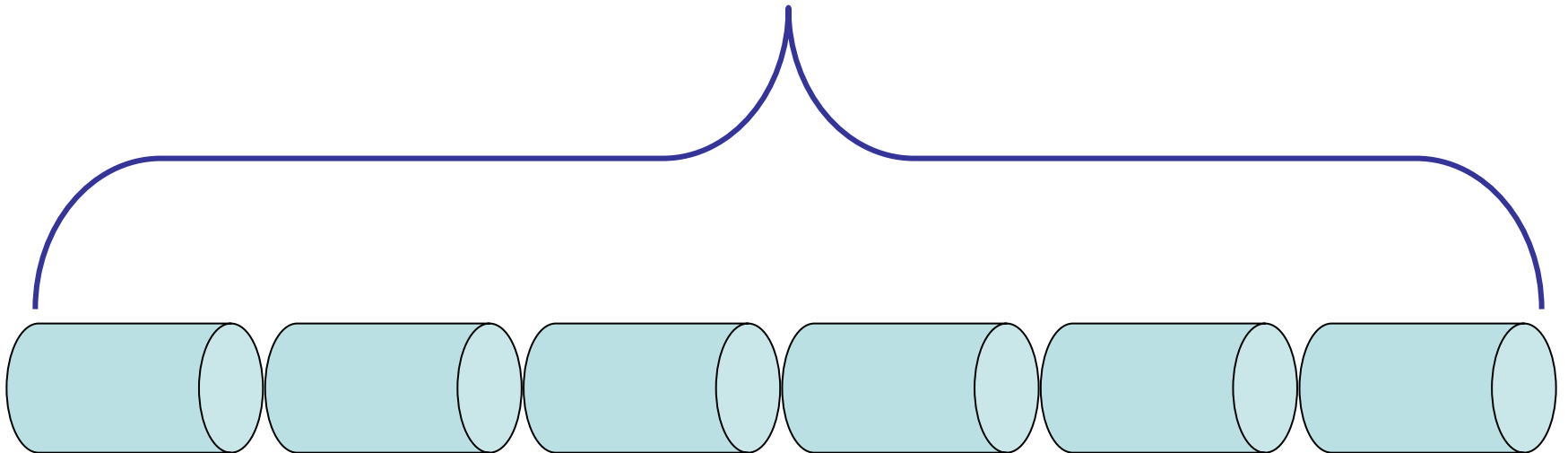
Stage 3. Read Data / Create Spatial Objects

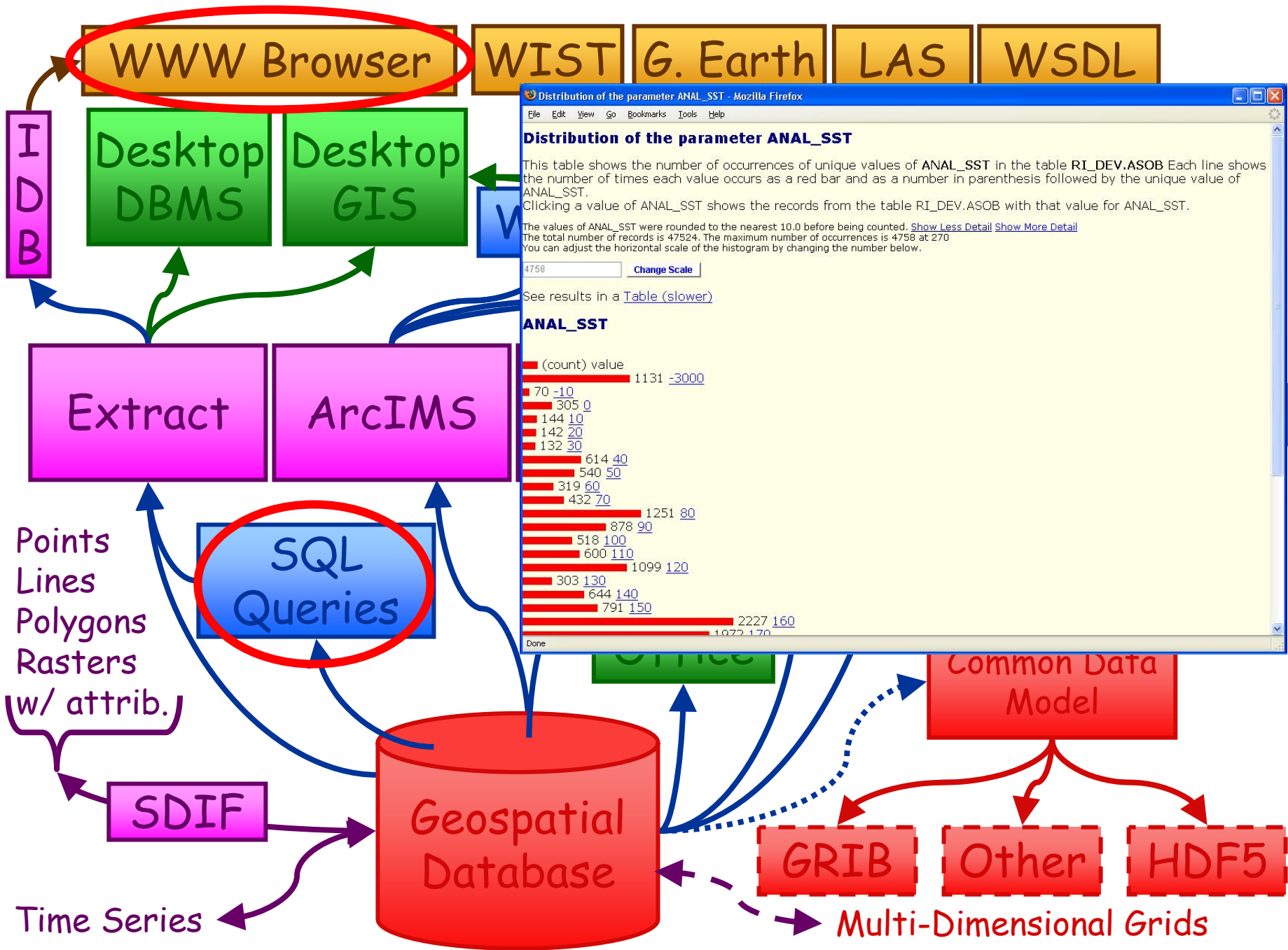
Stage 4. Write Thinned Layer (10%) to DB & CDM

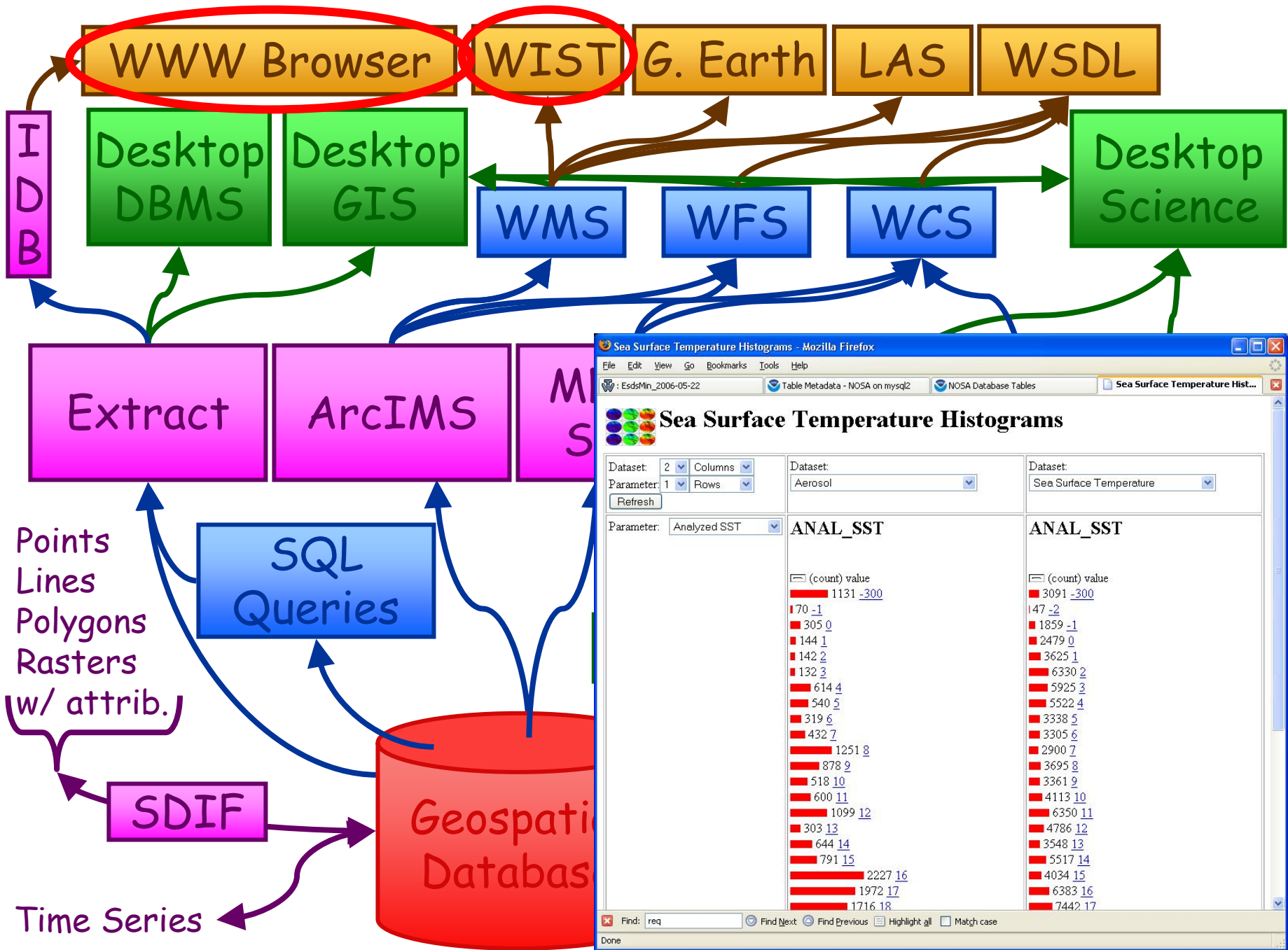
Stage 5. Write Complete Layer to DB & CDM

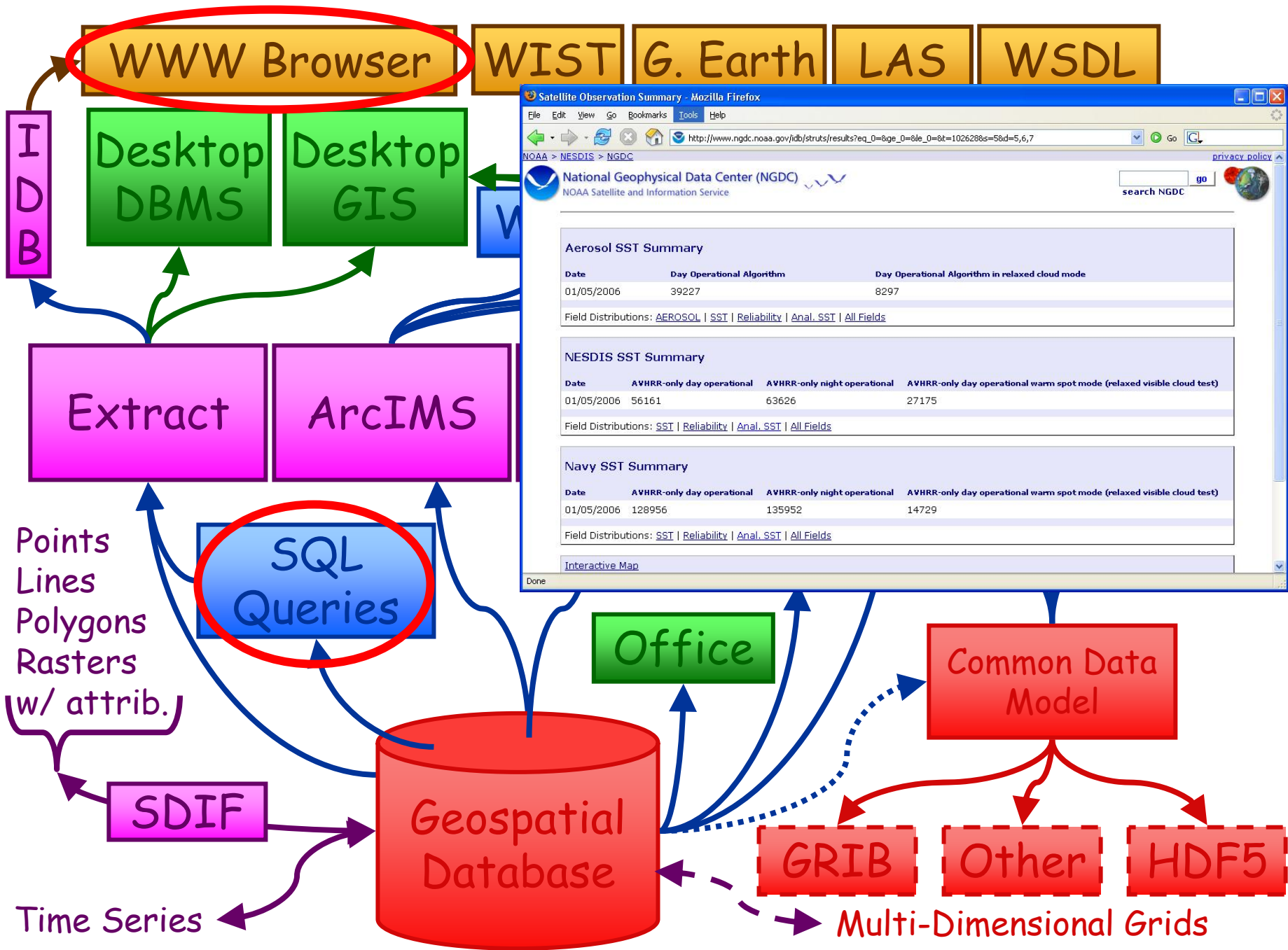
Stage 6. Create Summary (Grid) Table to DB & CDM

Stage 7. Create Rich Inventory Record

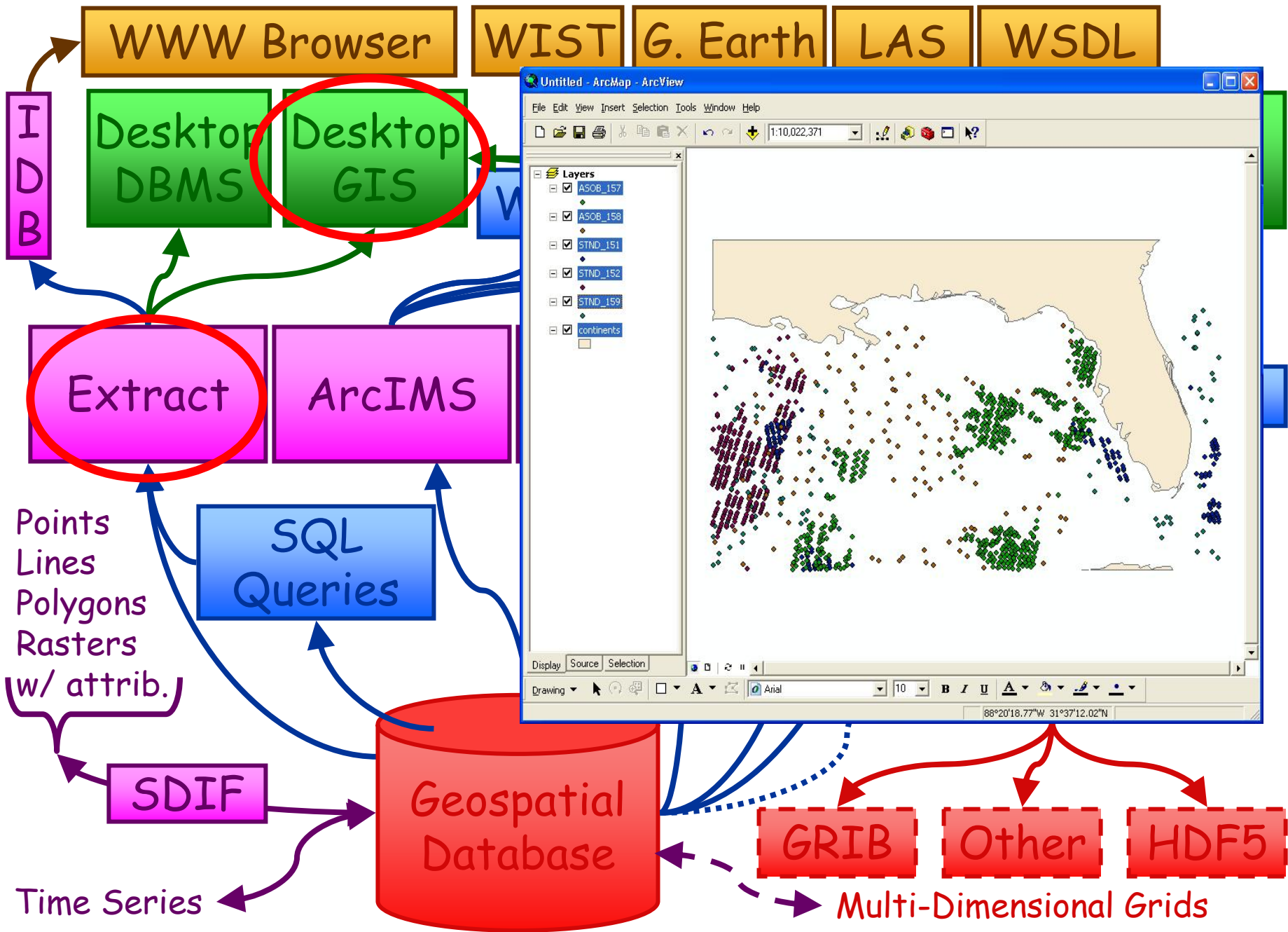


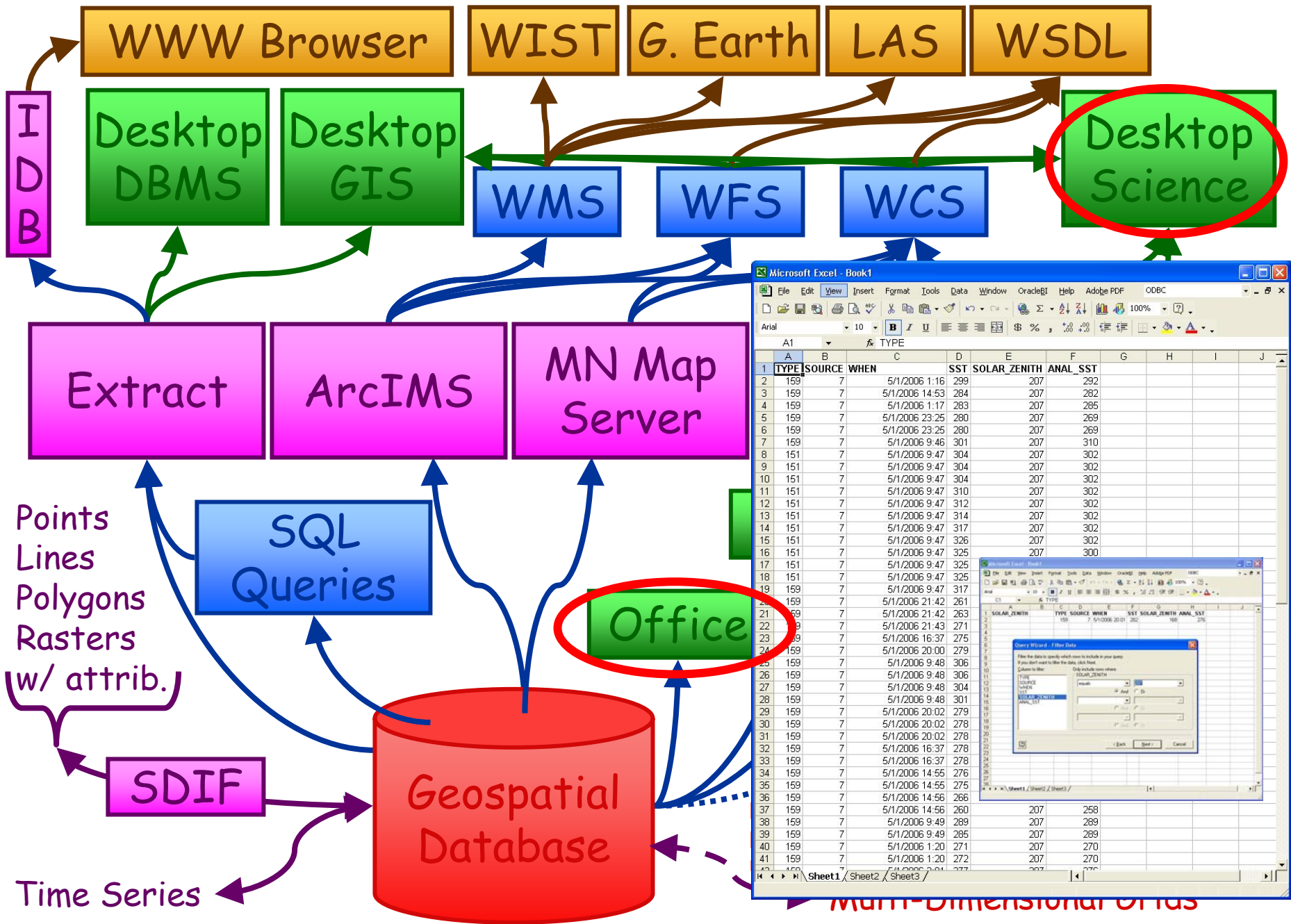








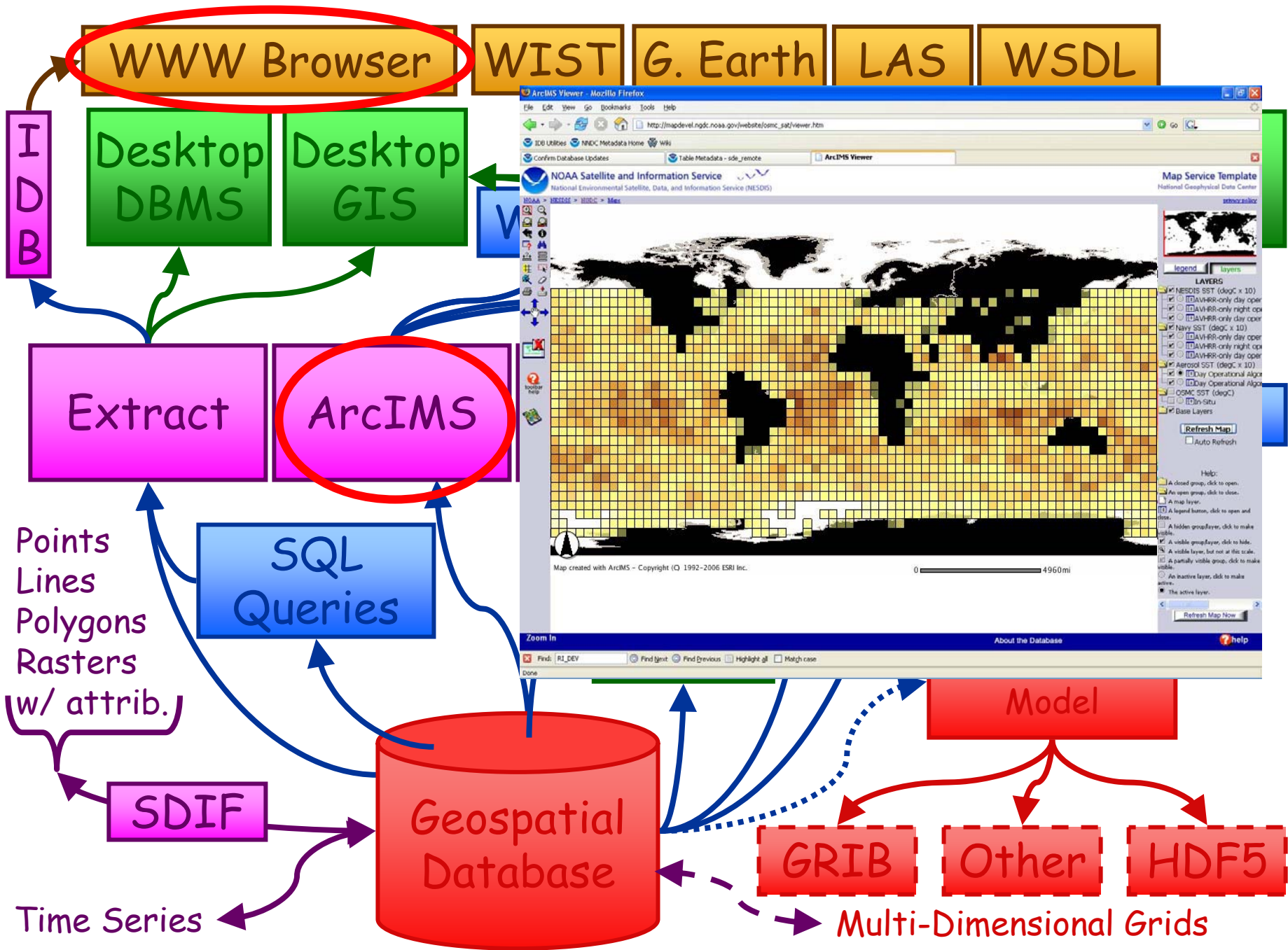


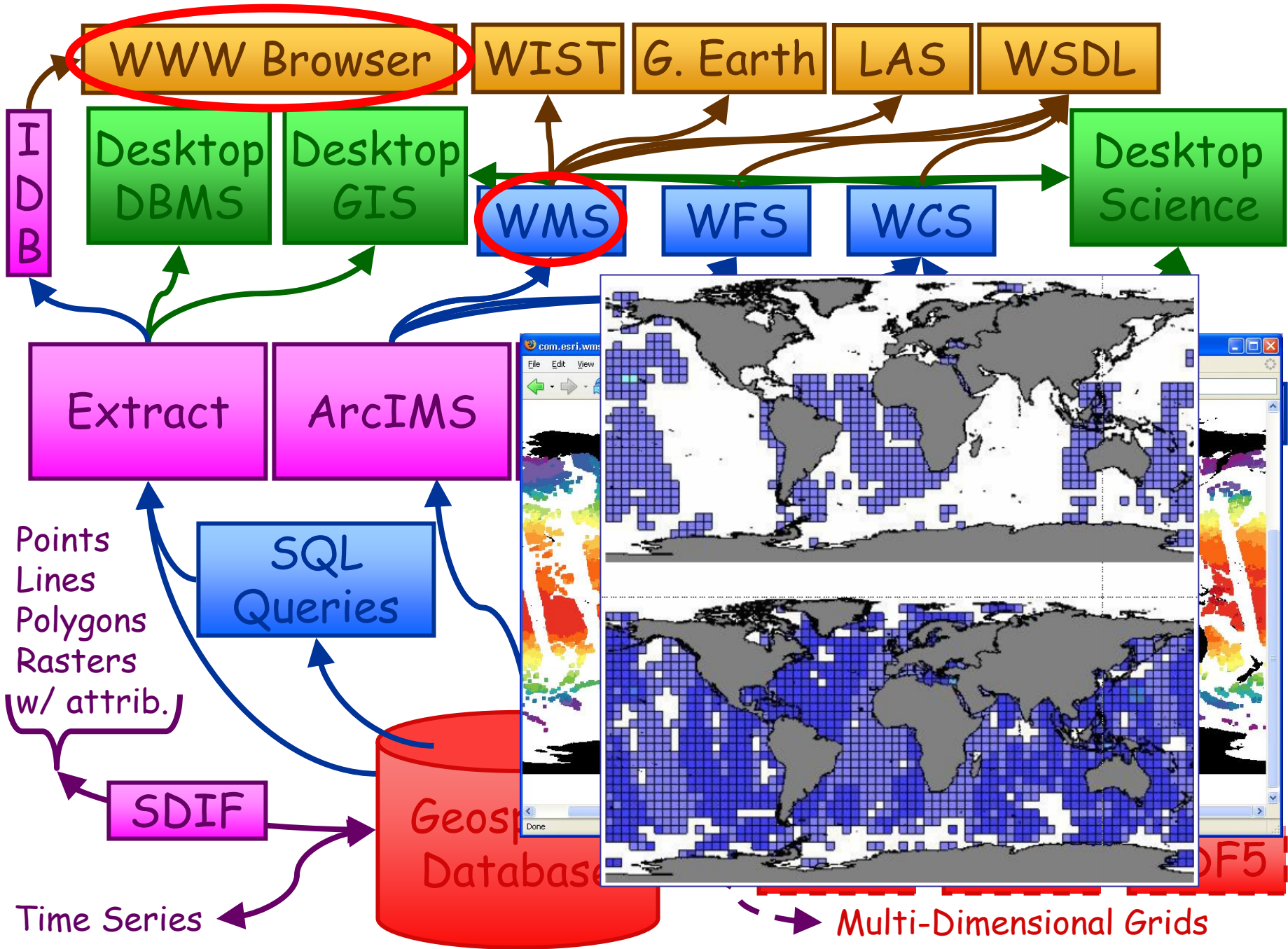


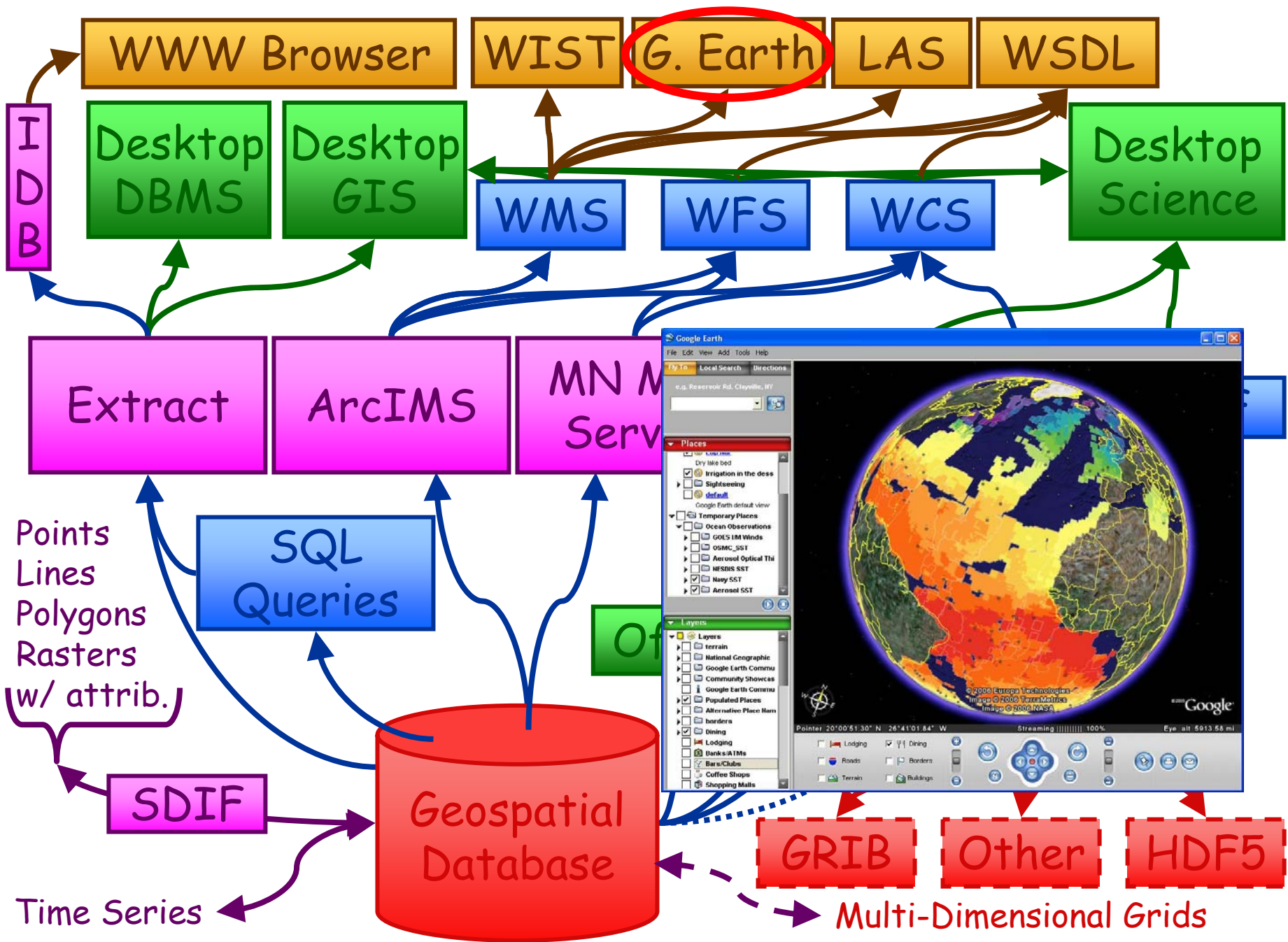
Microsoft Excel - Book1

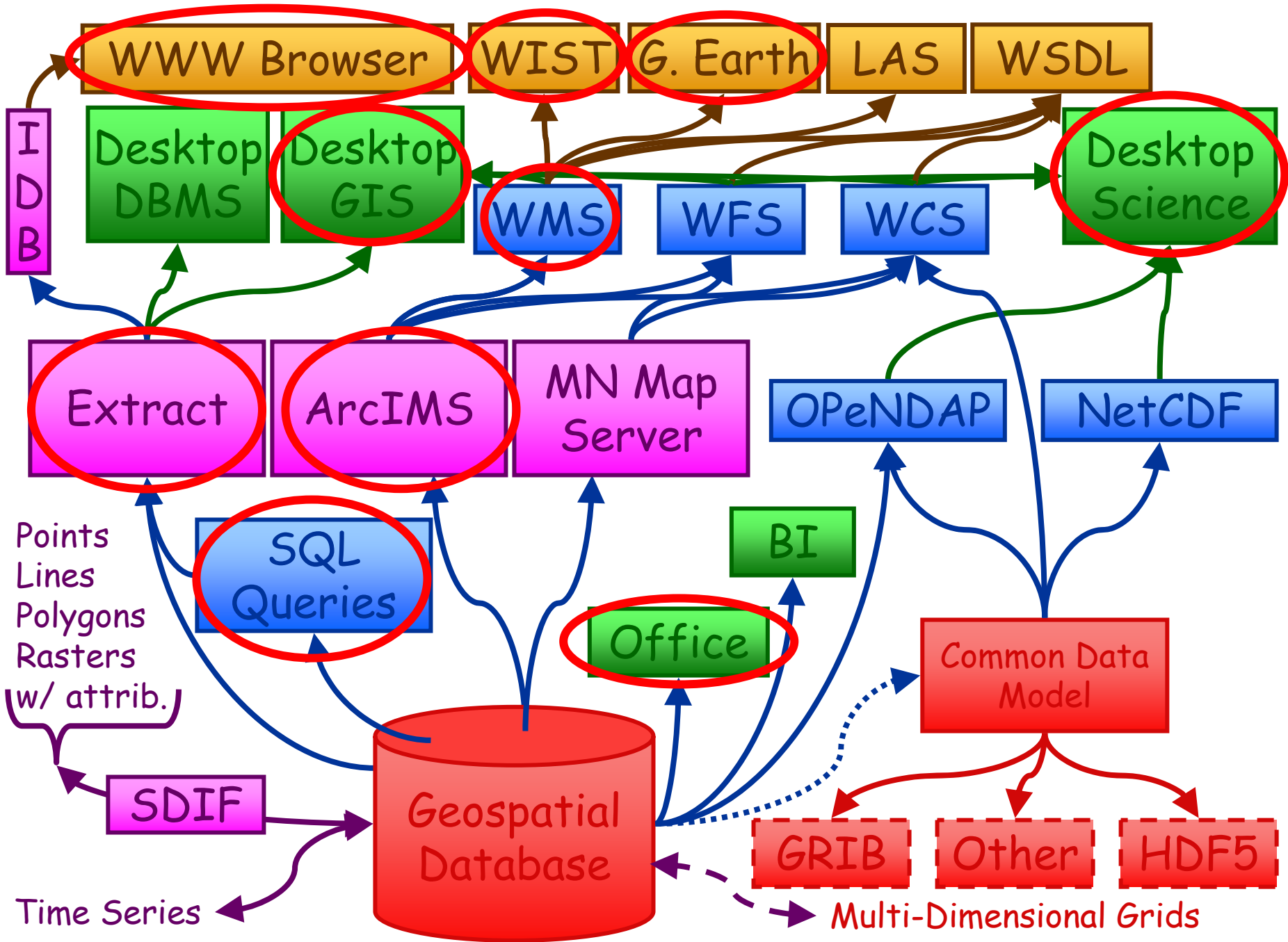
| A1   | B      | C    | D              | E            | F        | G   | H | I | J |
|------|--------|------|----------------|--------------|----------|-----|---|---|---|
| TYPE | SOURCE | WHEN | SST            | SOLAR ZENITH | ANAL_SST |     |   |   |   |
| 1    | 159    | 7    | 5/1/2006 1:16  | 299          | 207      | 292 |   |   |   |
| 2    | 159    | 7    | 5/1/2006 14:53 | 284          | 207      | 282 |   |   |   |
| 3    | 159    | 7    | 5/1/2006 1:17  | 283          | 207      | 285 |   |   |   |
| 4    | 159    | 7    | 5/1/2006 23:25 | 280          | 207      | 269 |   |   |   |
| 5    | 159    | 7    | 5/1/2006 23:25 | 280          | 207      | 269 |   |   |   |
| 6    | 159    | 7    | 5/1/2006 9:46  | 301          | 207      | 310 |   |   |   |
| 7    | 151    | 7    | 5/1/2006 9:47  | 304          | 207      | 302 |   |   |   |
| 8    | 151    | 7    | 5/1/2006 9:47  | 304          | 207      | 302 |   |   |   |
| 9    | 151    | 7    | 5/1/2006 9:47  | 304          | 207      | 302 |   |   |   |
| 10   | 151    | 7    | 5/1/2006 9:47  | 304          | 207      | 302 |   |   |   |
| 11   | 151    | 7    | 5/1/2006 9:47  | 310          | 207      | 302 |   |   |   |
| 12   | 151    | 7    | 5/1/2006 9:47  | 312          | 207      | 302 |   |   |   |
| 13   | 151    | 7    | 5/1/2006 9:47  | 314          | 207      | 302 |   |   |   |
| 14   | 151    | 7    | 5/1/2006 9:47  | 317          | 207      | 302 |   |   |   |
| 15   | 151    | 7    | 5/1/2006 9:47  | 326          | 207      | 302 |   |   |   |
| 16   | 151    | 7    | 5/1/2006 9:47  | 325          | 207      | 300 |   |   |   |
| 17   | 151    | 7    | 5/1/2006 9:47  | 325          | 207      | 302 |   |   |   |
| 18   | 151    | 7    | 5/1/2006 9:47  | 325          | 207      | 302 |   |   |   |
| 19   | 159    | 7    | 5/1/2006 9:47  | 317          | 207      | 302 |   |   |   |
| 20   | 159    | 7    | 5/1/2006 21:42 | 261          | 207      | 278 |   |   |   |
| 21   | 159    | 7    | 5/1/2006 21:42 | 263          | 207      | 278 |   |   |   |
| 22   | 159    | 7    | 5/1/2006 21:43 | 271          | 207      | 278 |   |   |   |
| 23   | 159    | 7    | 5/1/2006 16:37 | 275          | 207      | 278 |   |   |   |
| 24   | 159    | 7    | 5/1/2006 20:00 | 279          | 207      | 278 |   |   |   |
| 25   | 159    | 7    | 5/1/2006 9:48  | 306          | 207      | 278 |   |   |   |
| 26   | 159    | 7    | 5/1/2006 9:48  | 306          | 207      | 278 |   |   |   |
| 27   | 159    | 7    | 5/1/2006 9:48  | 304          | 207      | 278 |   |   |   |
| 28   | 159    | 7    | 5/1/2006 9:48  | 301          | 207      | 278 |   |   |   |
| 29   | 159    | 7    | 5/1/2006 20:02 | 279          | 207      | 278 |   |   |   |
| 30   | 159    | 7    | 5/1/2006 20:02 | 278          | 207      | 278 |   |   |   |
| 31   | 159    | 7    | 5/1/2006 20:02 | 278          | 207      | 278 |   |   |   |
| 32   | 159    | 7    | 5/1/2006 16:37 | 278          | 207      | 278 |   |   |   |
| 33   | 159    | 7    | 5/1/2006 16:37 | 278          | 207      | 278 |   |   |   |
| 34   | 159    | 7    | 5/1/2006 14:55 | 276          | 207      | 278 |   |   |   |
| 35   | 159    | 7    | 5/1/2006 14:55 | 275          | 207      | 278 |   |   |   |
| 36   | 159    | 7    | 5/1/2006 14:56 | 266          | 207      | 278 |   |   |   |
| 37   | 159    | 7    | 5/1/2006 14:56 | 260          | 207      | 258 |   |   |   |
| 38   | 159    | 7    | 5/1/2006 9:49  | 289          | 207      | 289 |   |   |   |
| 39   | 159    | 7    | 5/1/2006 9:49  | 285          | 207      | 289 |   |   |   |
| 40   | 159    | 7    | 5/1/2006 1:20  | 271          | 207      | 270 |   |   |   |
| 41   | 159    | 7    | 5/1/2006 1:20  | 272          | 207      | 270 |   |   |   |
| 42   | 159    | 7    | 5/1/2006 2:04  | 277          | 207      | 270 |   |   |   |

Multi-dimensional data

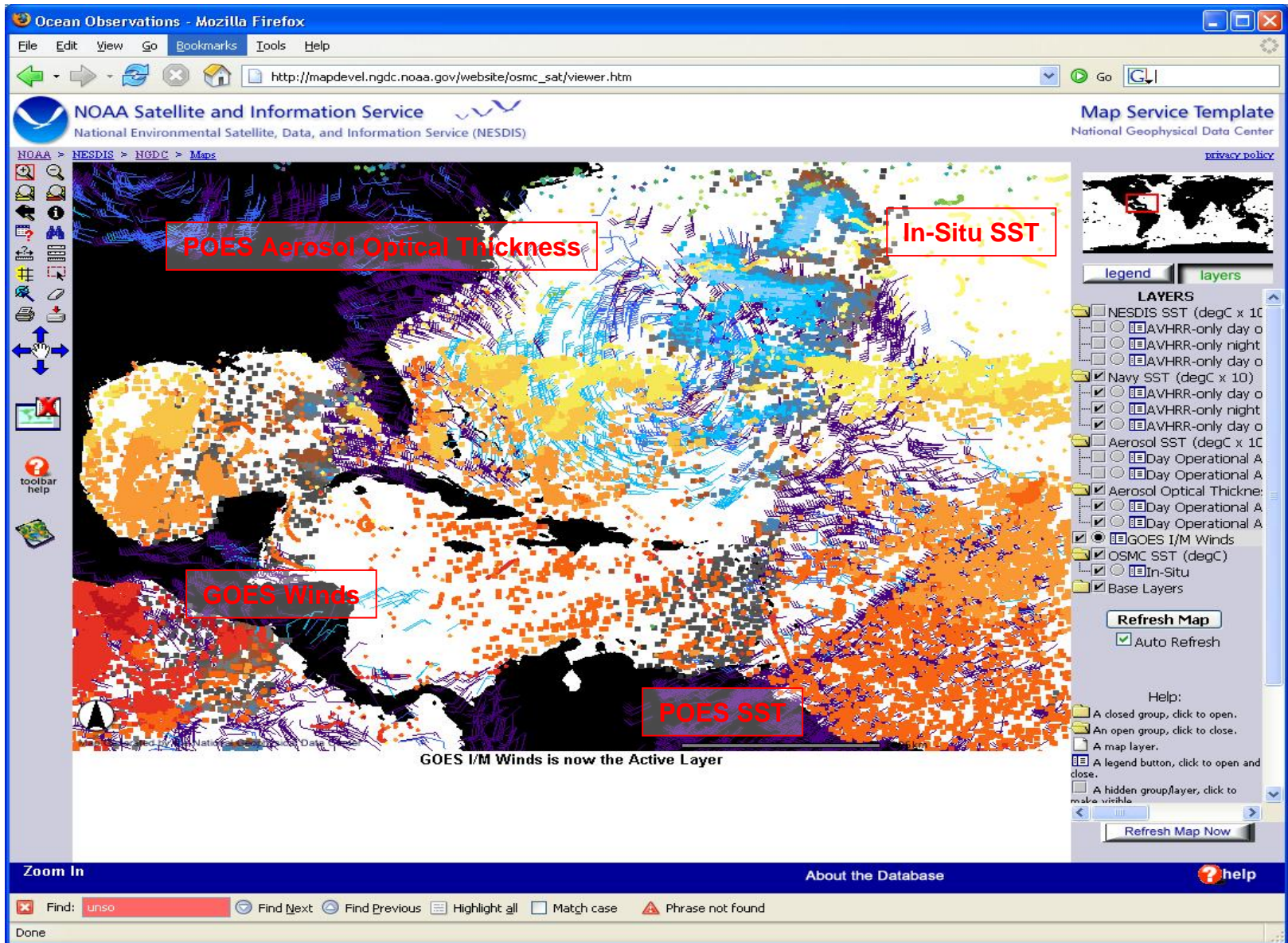


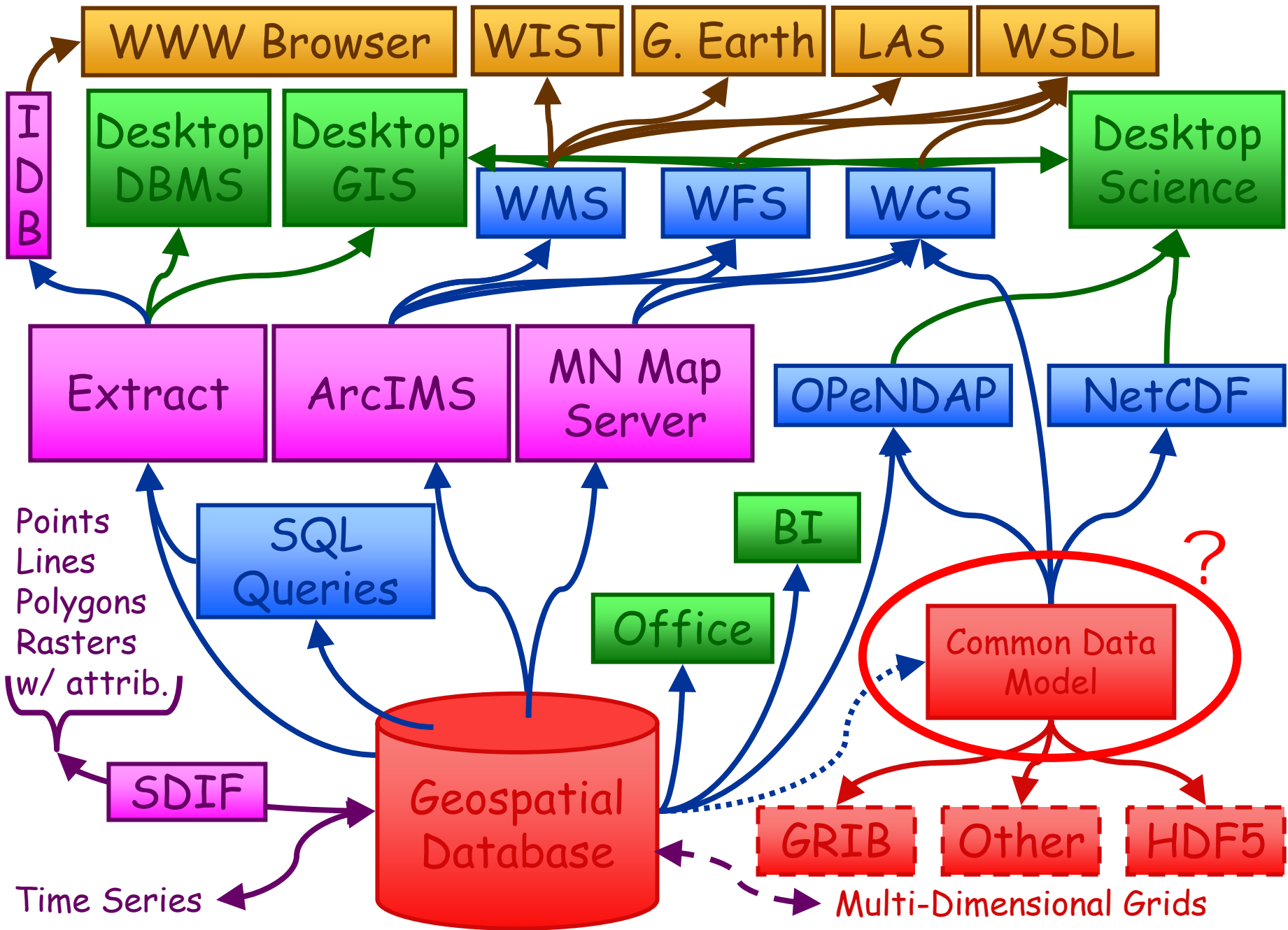






# Integrated Visualization (GIS)







# Partnership?

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

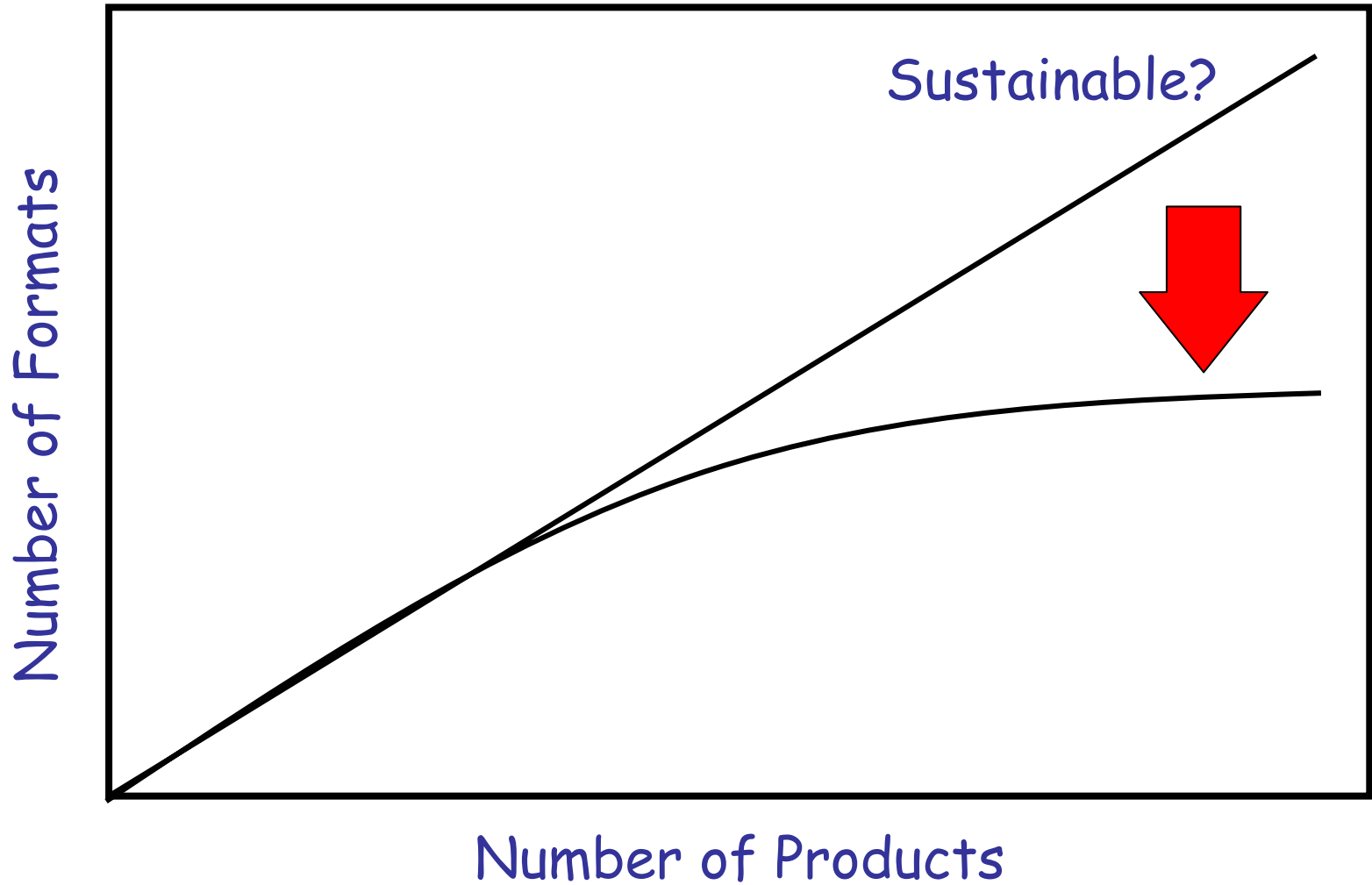
NOAA Data Management Integration Team (DMIT) voted "Support for Common Data Model" as the #1 recommendation to IOOS for work that is consistent with the NOAA GEO-Integrated Data Environment Plan.

10 NOAA people attended Unidata training.

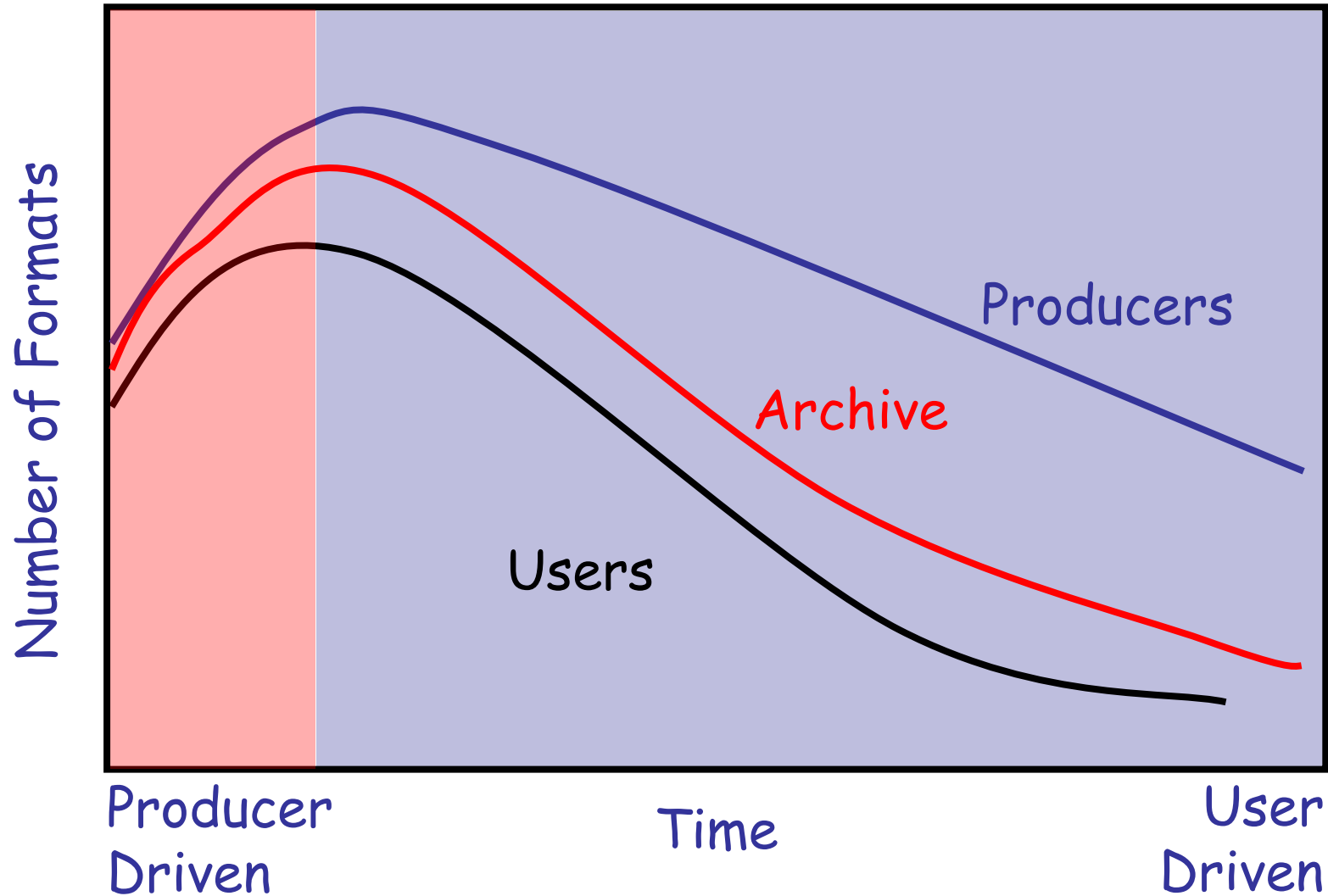


8 CLASS developers and others attending HDF Conference.

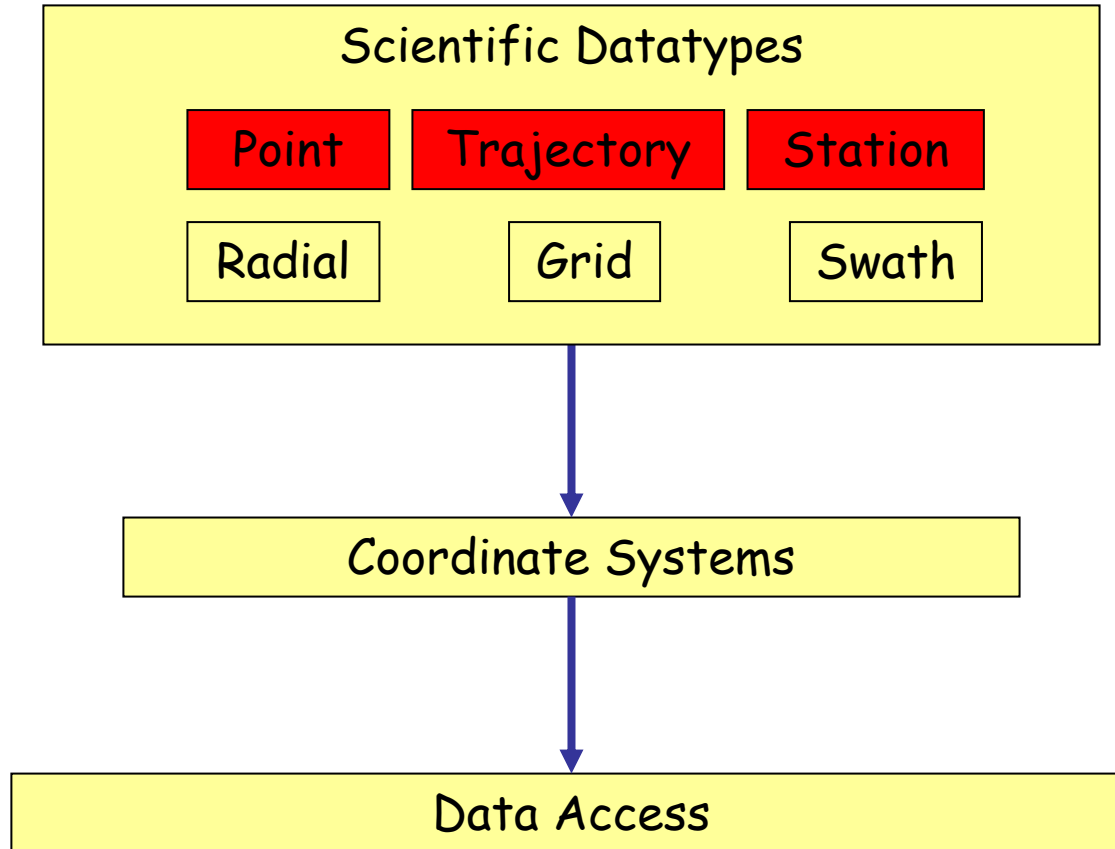
# Formats and Products



# Format Evolution



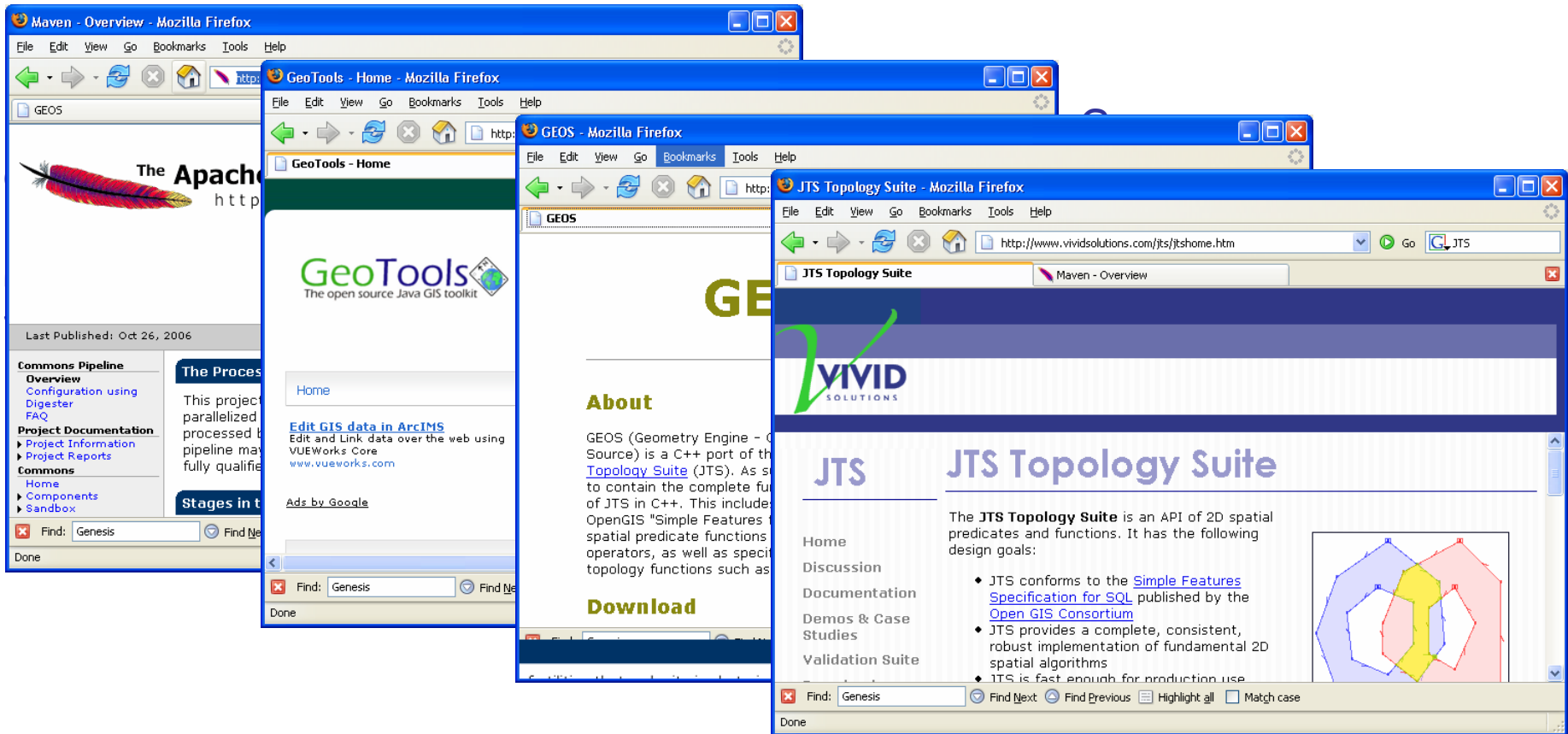
# Common Data Model



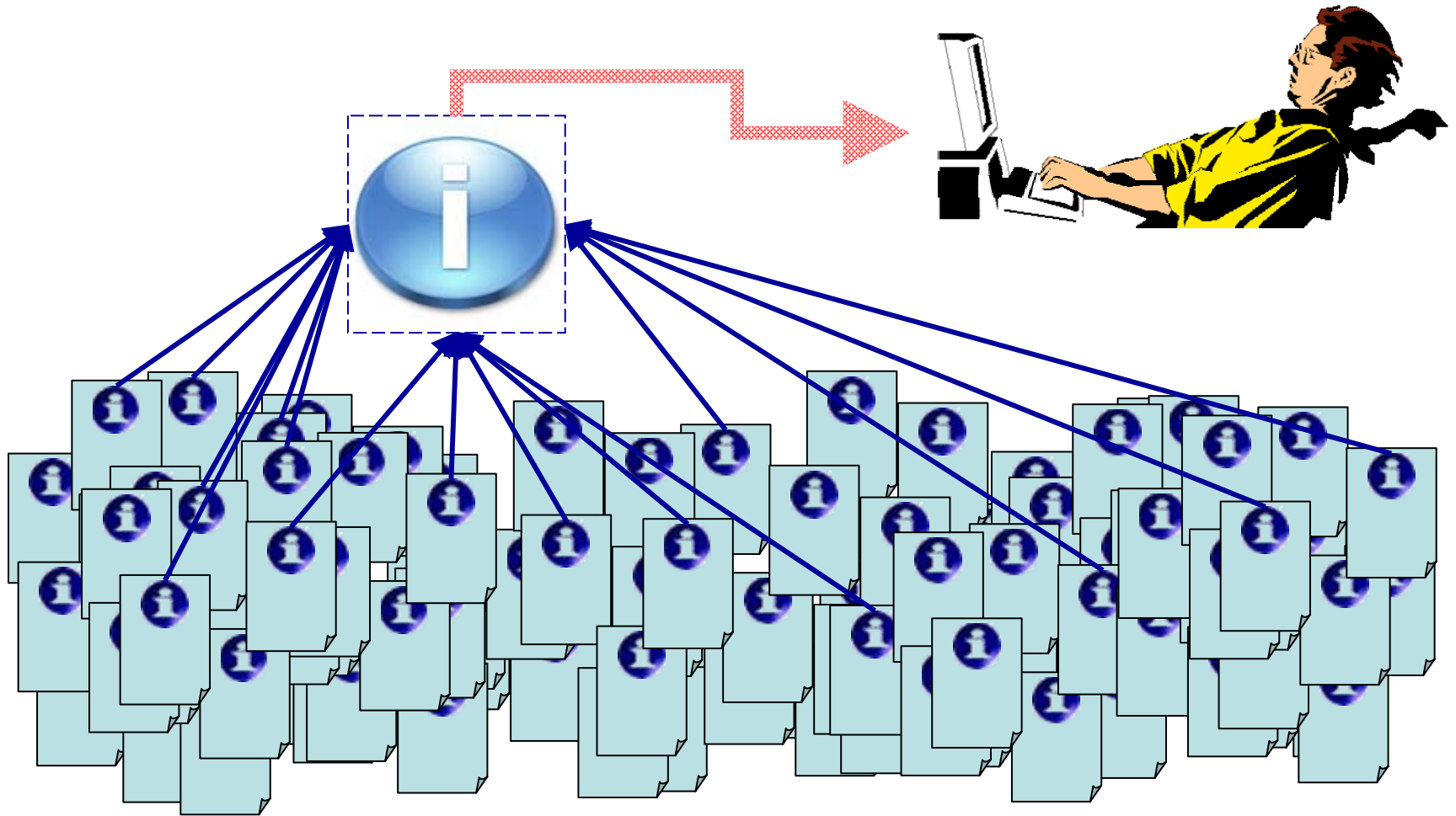
Open Geospatial Consortium Simple Features

# 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.

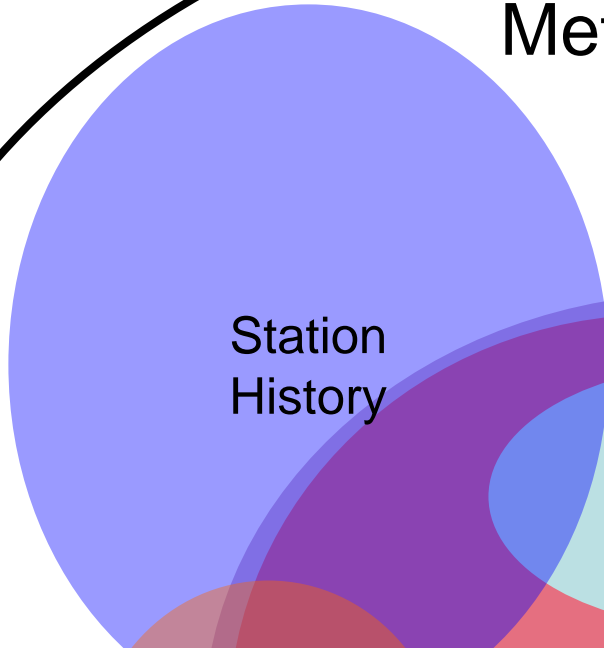


# The Rich Inventory Concept

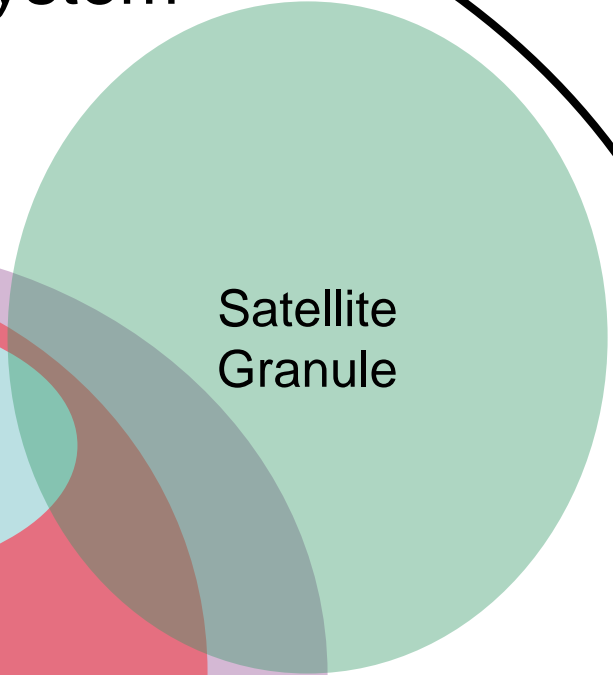


Very similar to "file content metadata" at NCAR

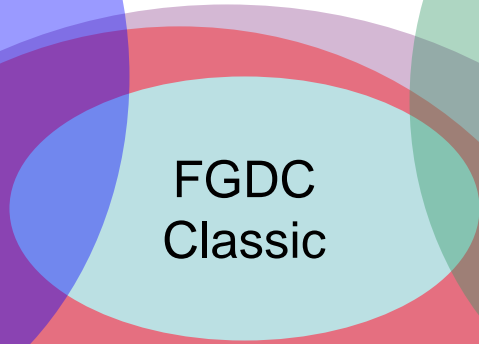
# Integrated NOAA Metadata System



Station  
History



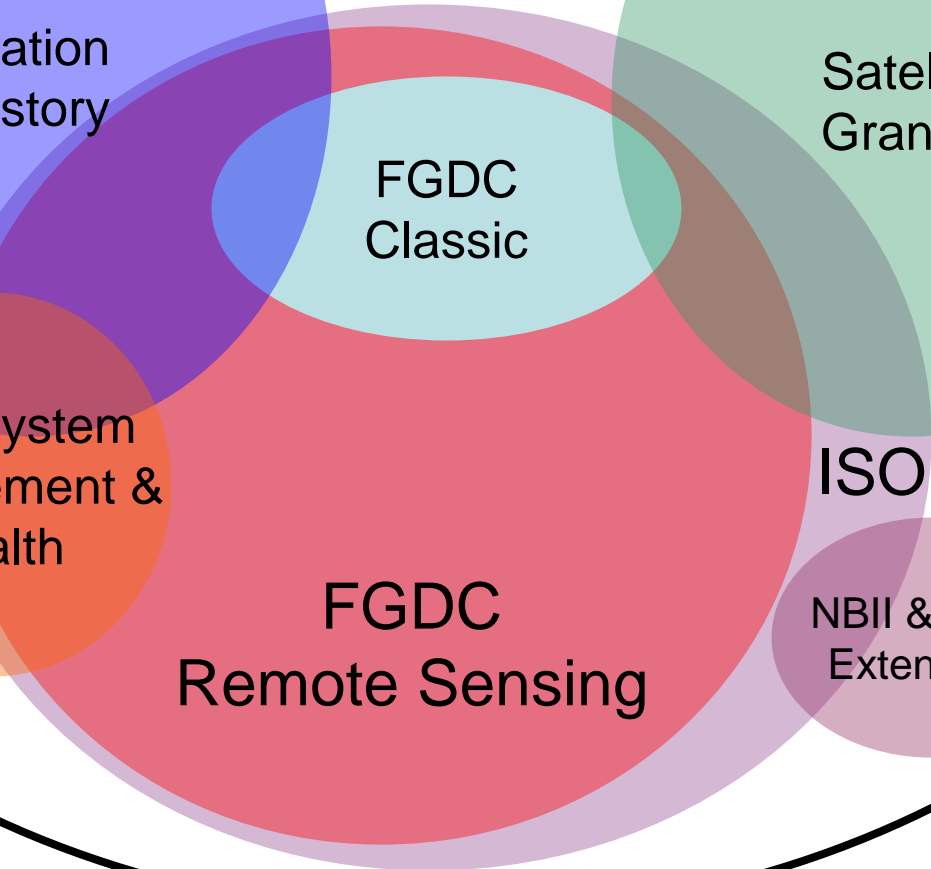
Satellite  
Granule



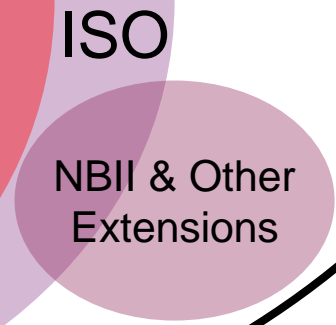
FGDC  
Classic



Obs. System  
Management &  
Health



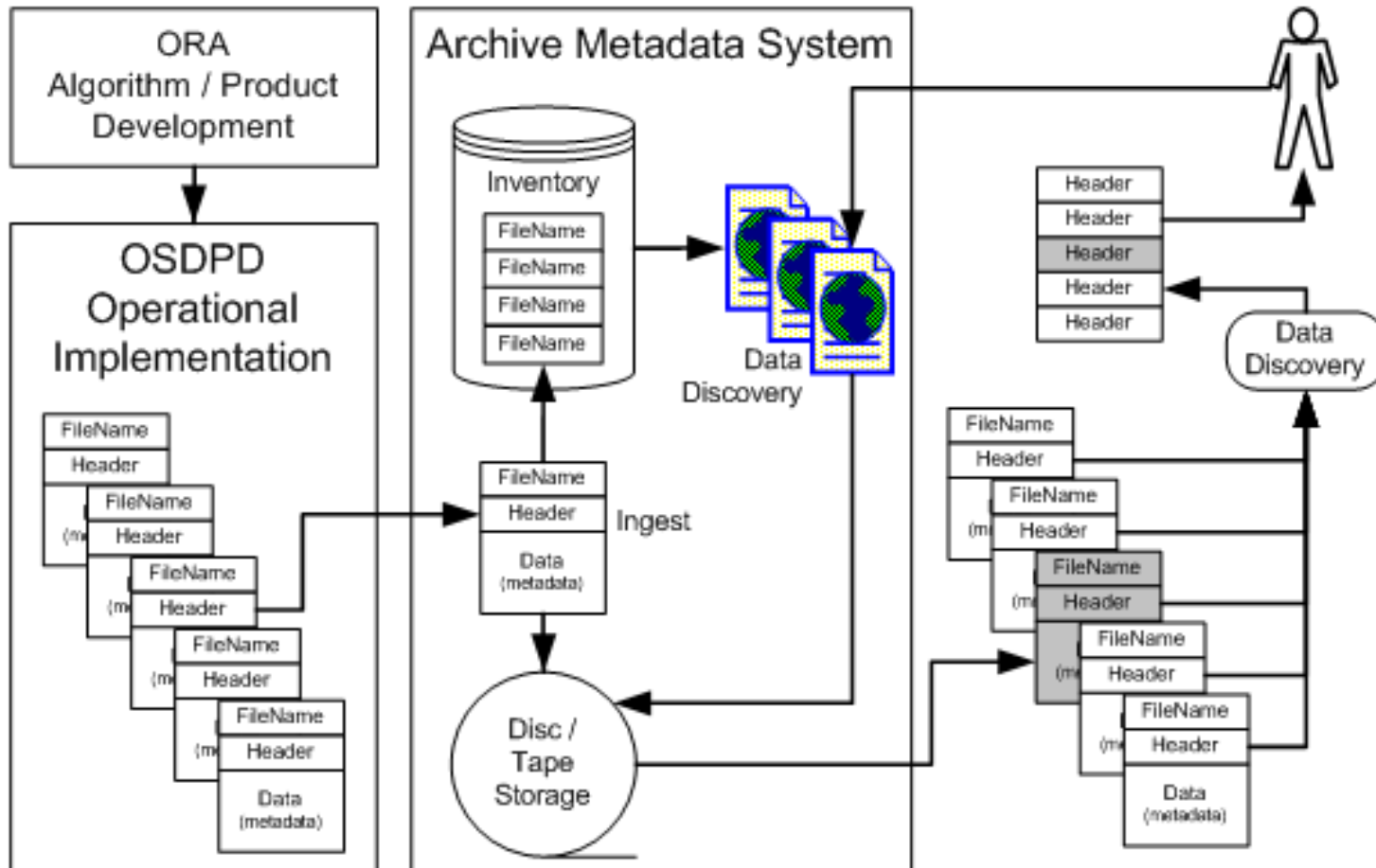
FGDC  
Remote Sensing



ISO

NBII & Other  
Extensions

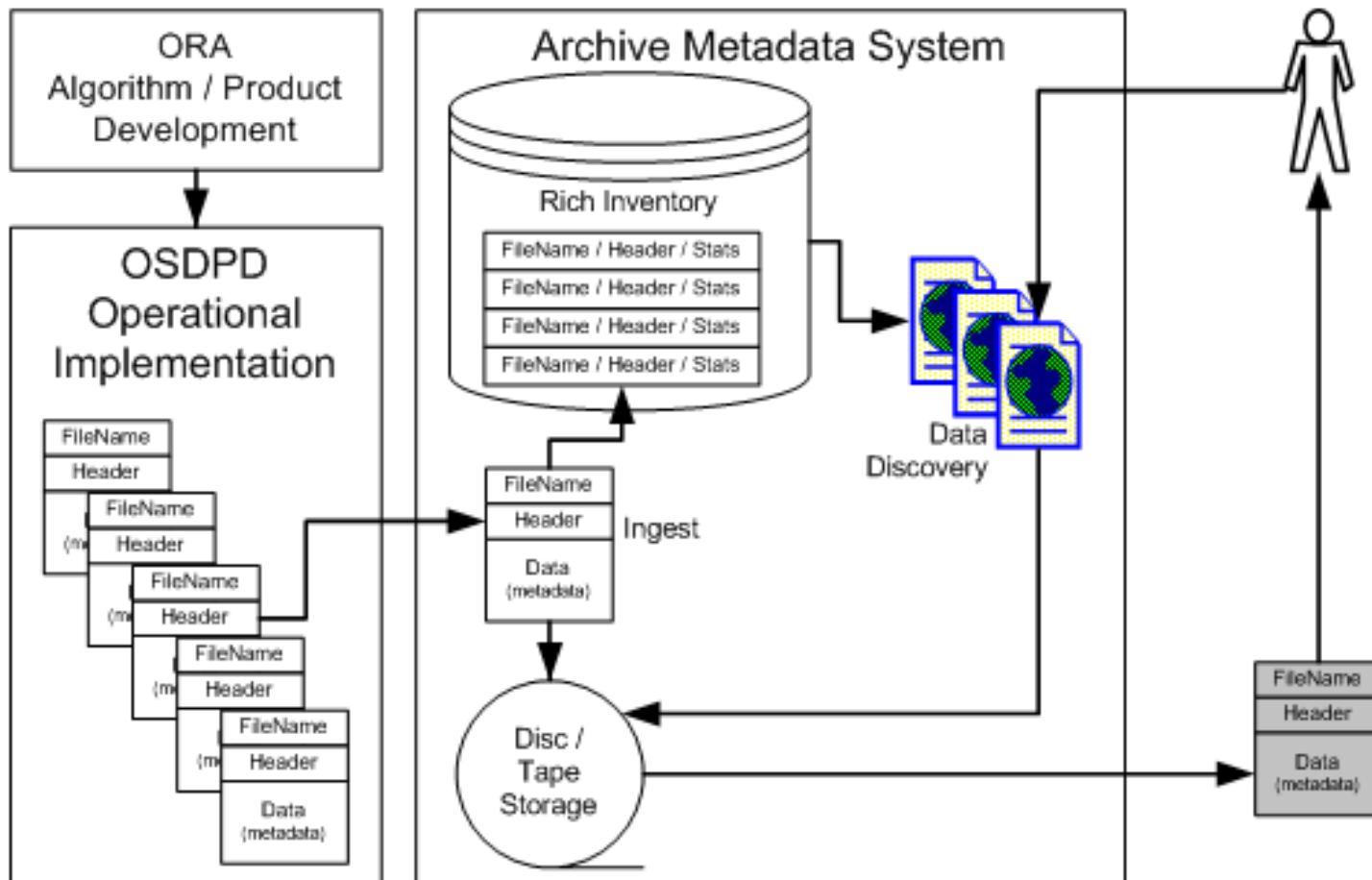
# Granule Metadata Management - Present



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.

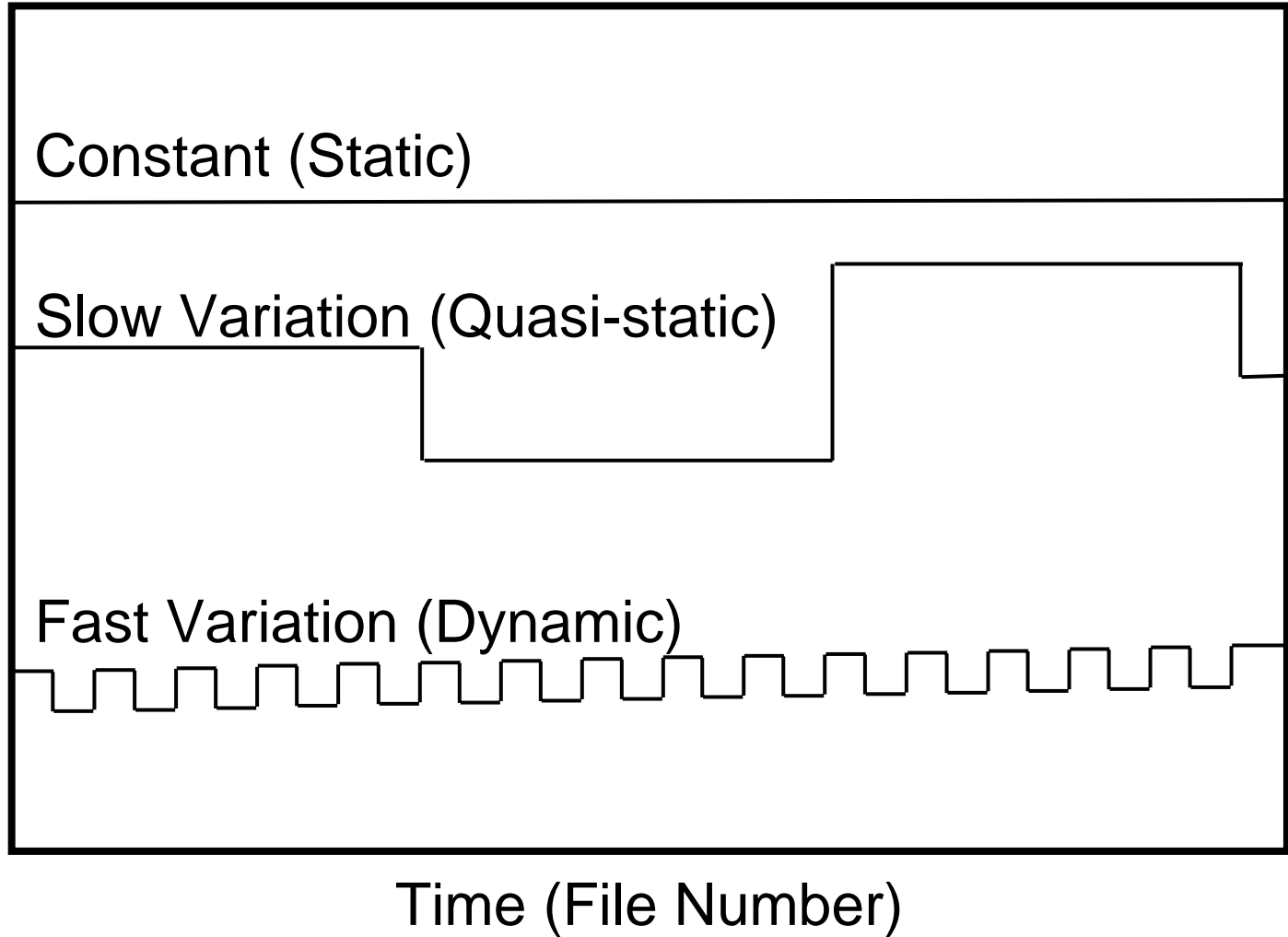


# Granule Metadata Management - Rich Inventory

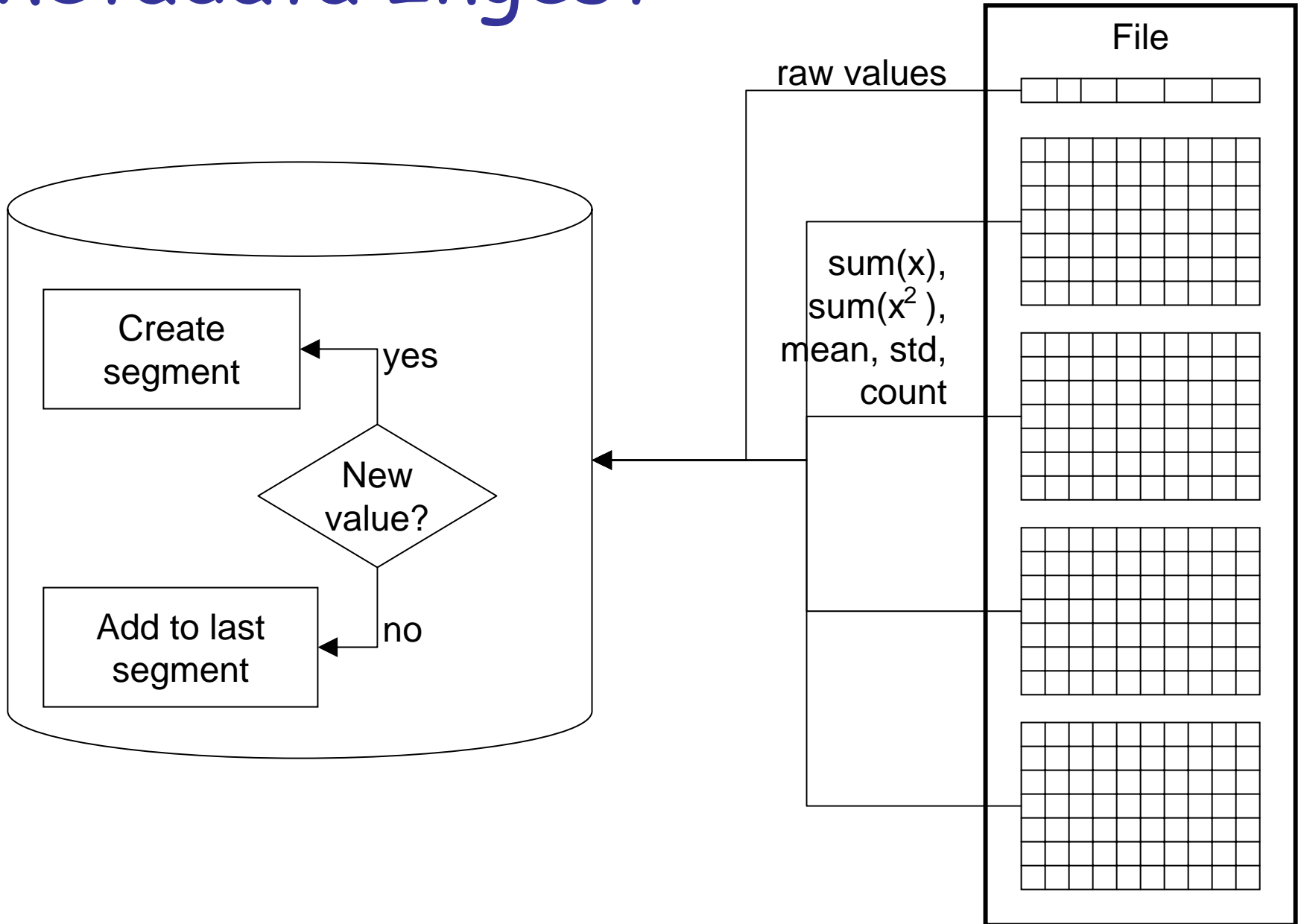


1. Files come to CLASS
2. Filename and fileheader metadata are added to inventory.
3. Descriptive Statistics are calculated and added to inventory.
4. All metadata is available to the data discovery system and users get the data they need without secondary data discovery.

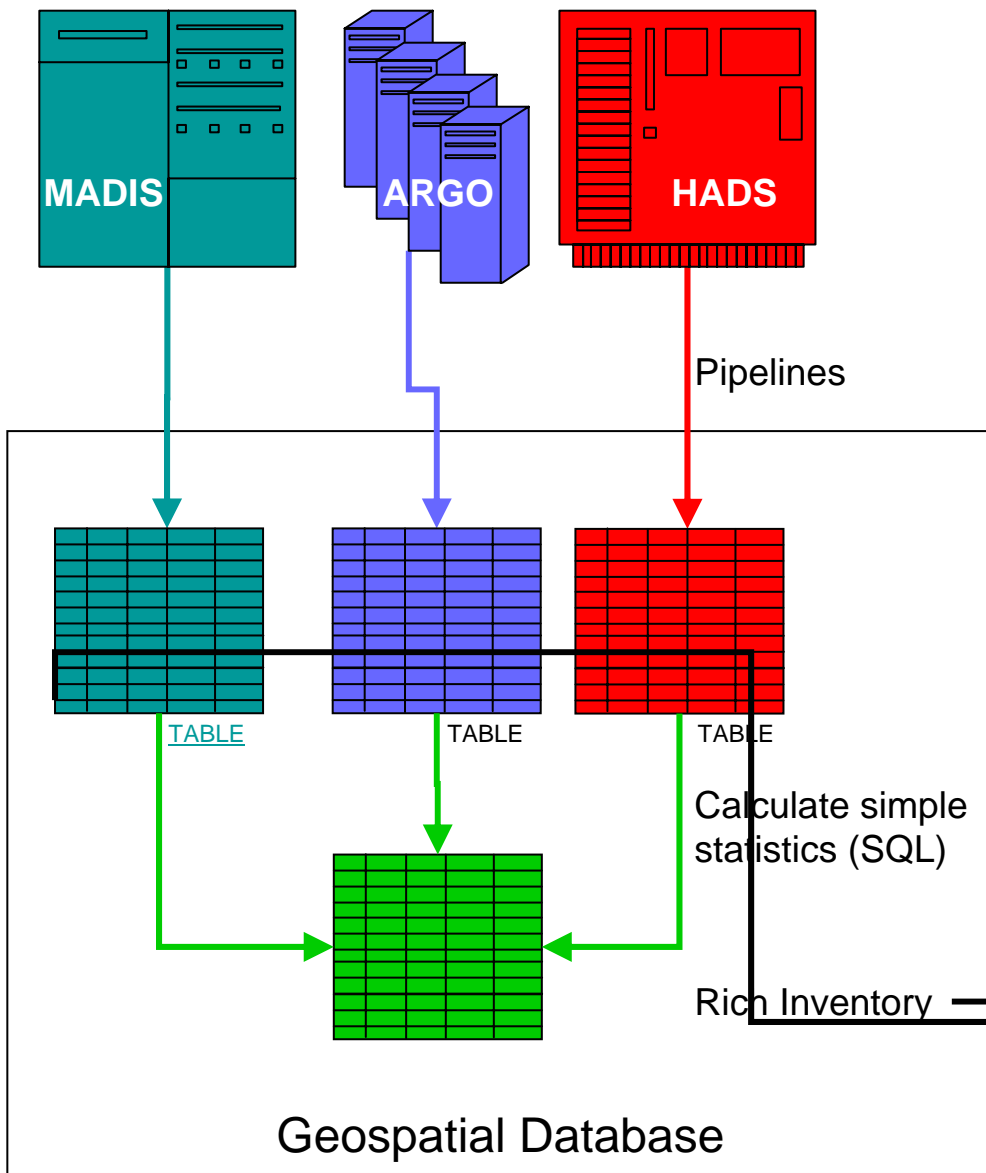
# Segment Model



# Metadata Ingest



# Automated Observing System Ingest



NOSA Automatic Loads - Mozilla Firefox

Apache Tomcat/5.5.15 - Error report

NOSA Automatic Loads

| Table Name               | Row Count | Rows Altered | Last Update Finish    | Repeat Interval | Update History   |
|--------------------------|-----------|--------------|-----------------------|-----------------|--|
| NOSA_10DEGREE_COUNT_TSQA | 648       |              | 2006-11-27 01:14:07.0 | daily           | <a href="#">Recent Updates</a> <a href="#">History</a> |
| GOESWINDS_TSQP           | 36212     |              | 2006-11-27 00:33:24.0 | monday          | <a href="#">Recent Updates</a> <a href="#">History</a> |
| ACARS_TSQL               | 46407     | 48767        | 2006-11-27 00:22:14.0 | daily           | <a href="#">Recent Updates</a> <a href="#">History</a> |
| ARGO_TSQL                | 1867      | 1937         | 2006-11-27 00:25:55.0 | monday          | <a href="#">Recent Updates</a> <a href="#">History</a> |
| MESONET_TSQP             | 19191     | 20685        | 2006-11-23 00:08:19.0 | thursday        | <a href="#">Recent Updates</a> <a href="#">History</a> |
| NASA_AERONET_TSQP        | 505       | 545          | 2006-11-27 00:05:55.0 | monday          | <a href="#">Recent Updates</a> <a href="#">History</a> |
| NWS_HADS_TSQP            | 12031     | 12397        | 2006-11-27 00:07:17.0 | daily           | <a href="#">Recent Updates</a> <a href="#">History</a> |
| NWS_VOS_TSQP             | 1297      | 1522         | 2006-11-27 00:06:20.0 | monday          | <a href="#">Recent Updates</a> <a href="#">History</a> |
| OAR_CWOP_TSQP            | 4445      | 5126         | 2006-11-22 00:05:56.0 | wednesday       | <a href="#">Recent Updates</a> <a href="#">History</a> |
| OAR_ENSO_XBT_TSQP        | 32851     | 43396        | 2006-11-27 00:12:40.0 | monday          | <a href="#">Recent Updates</a> <a href="#">History</a> |
| OAR_GPSMET_TSQP          | 419       | 423          | 2006-11-27 00:05:42.0 | monday          | <a href="#">Recent Updates</a> <a href="#">History</a> |
| RADIOSONDES_TSQL         | 222       | 14           | 2006-11-23 00:05:16.0 | daily           | <a href="#">Recent Updates</a> <a href="#">History</a> |

WISTs: [ACARS](#) | [ARGO](#) | [CWOP](#) | [ENSO](#) | [GOESWinds](#) | [GPSMET](#) | [HADS](#) | [Mesonet](#) | [Radiosondes](#) | [Volunteer Ships](#) | [Counts](#)

Find: Genesis  Find Next  Find Previous  Highlight all  Match case

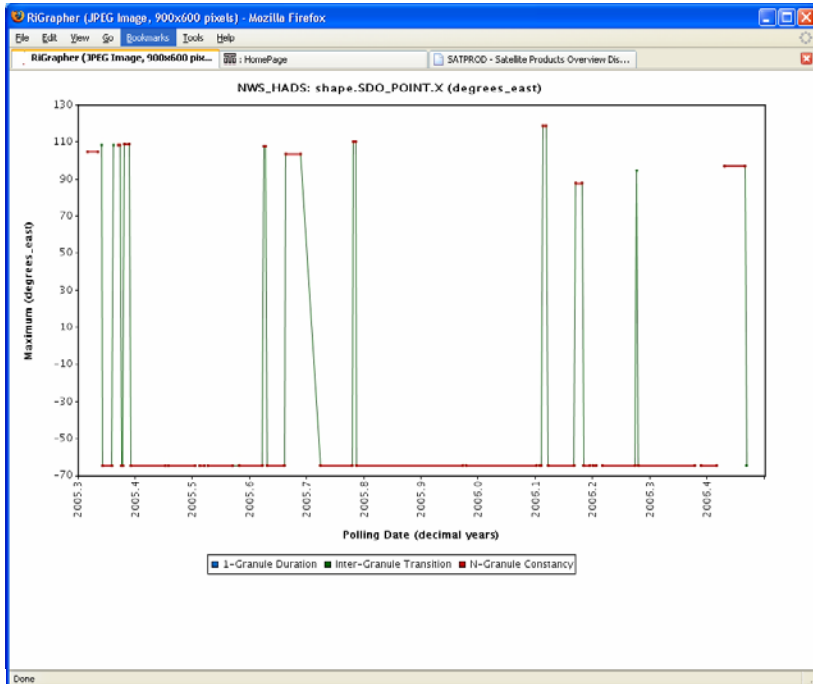
Done

**WIST: Rich Inventory Counts** - Shows a grid of data points and several line graphs representing time-series data for various parameters.

**Automated NWS Update Statistics** - Displays a table with columns for table name, parameters, full loop date, count, and resource, providing a summary of update activities.

**Integrated Surface Observing System** - Features a map of the United States with a color-coded overlay representing different observing stations or data points. A sidebar on the right lists various layers and data sources.

# HADS Network Monitoring



NOAA Satellite and Information Service  
National Environmental Satellite, Data, and Information Service (NESDIS)

Integrated Surface Observing System (ISOS)

Layers:

- Observing Station Location
- NOAA
- Non-NOAA
- RAWS
- MADIS-Mesone
- MADIS
- HADS
- Station Density Layers
- One-Degree Grid
- Forecast Zones
- HUC Regions
- HUC Subregions
- Base Layers

Refresh Map

Auto Refresh

Layers Help:

- A closed group, click to open.
- An open group, click to close.
- A legend button, click to open and close.
- A hidden group(layer, click to make visible.
- A visible group(layer, click to hide.
- A visible layer, but not at this scale.
- A partially visible group, click to

| Owner | State | Service Area | Latitude | Longitude | Daily Trans Time | Trans Interval | Location | Decode Mode |
|-------|-------|--------------|----------|-----------|------------------|----------------|----------|-------------|
| F1    | TX    | FWD          | 31.918   | 96.886    | 43               | 0              | MALONE   | S           |

About the Database

help

DCP Info - Mozilla Firefox

Meta Data

NESDIS ID: CE2764EE NWS Location ID: MLET2

Location: MALONE

Latitude: N 31°55'04" Longitude: E 96°53'46" HSA: FWD State: TX

Owner: CESWF1 Channel: 4 Initial Transmit Time (HHMM): 0043 Transmission Interval (min): 240

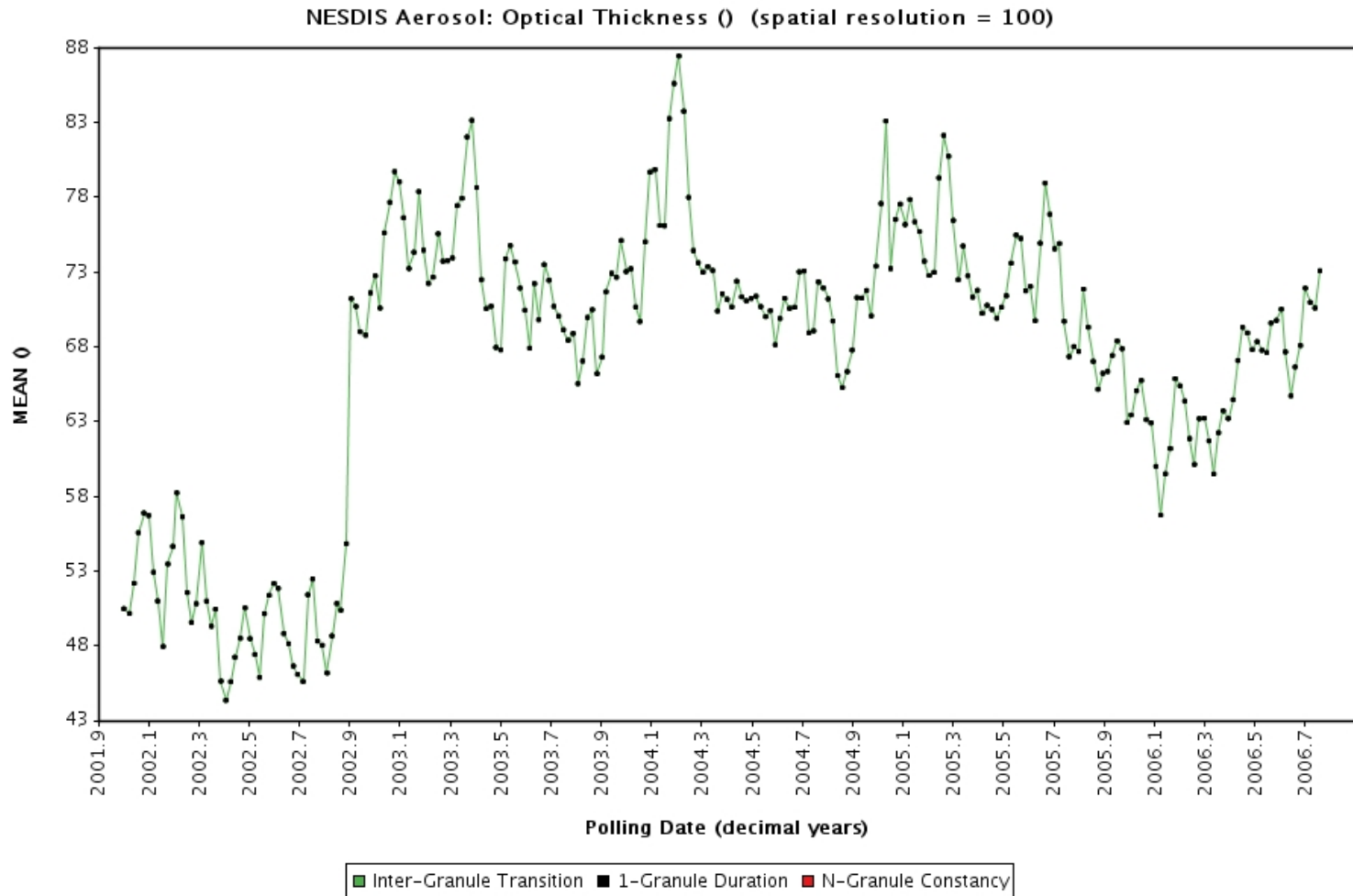
Transmission Times: GMT 00:43 04:43 08:43 12:43 16:43 20:43

Decode Info

| NWSLI | Data Interval(min) Self-time | Data Interval(min) Random | SHEF Code | Time Offset(min) | Base Elevation(ft) | Coefficient Self-time | Coefficient Random | Constant |
|-------|------------------------------|---------------------------|-----------|------------------|--------------------|-----------------------|--------------------|----------|
| MLET2 | 15                           |                           | PCIRG     | 13               | 0                  | 0.01                  |                    | 0        |
| MLET2 | 15                           |                           | FPIRG     | 13               | 0                  | 0.01                  |                    | 0        |

Today Decoded Data

# 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:

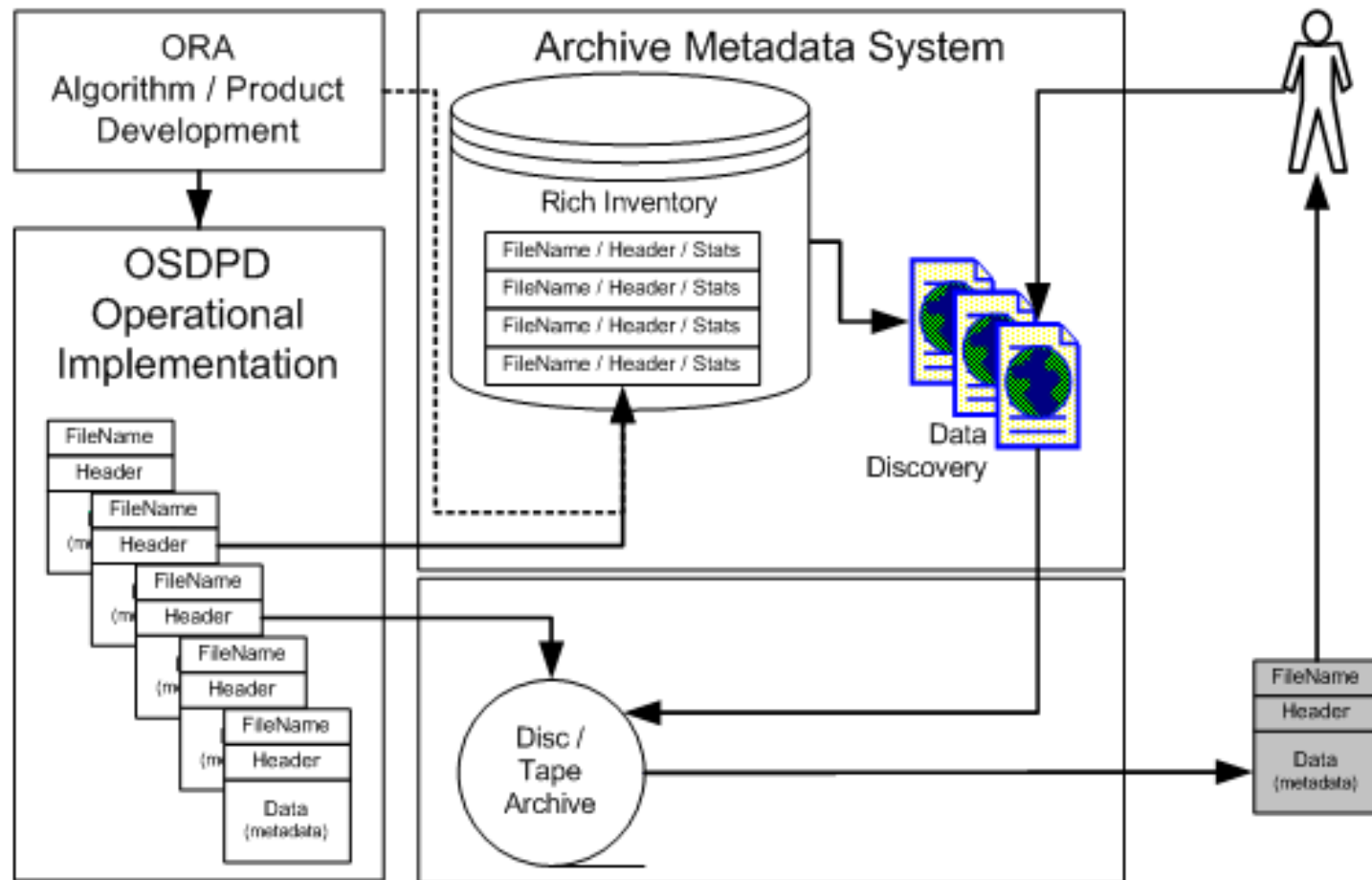
it appears that the current 3rd generation aerosol algorithm was implemented into operations around Oct-Nov 2002 time frame. cannot say more precisely, as all email correspondence i am looking at, talks about this indirectly. (maybe it's what Steve refers to as the Phase II aerosol-SST algorithm.) At 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 ( $\sim 0.08$ ) look more realistic than the previous ones ( $\sim 0.05$  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](mailto:ted.habermann@noaa.gov)

